An Examination of Moral Judgment in College: Integrating Developmental and College Impact Frameworks

by

Matthew A. Holsapple

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Higher Education)
in the University of Michigan
2014

Doctoral Committee:

Professor Patricia M. King, Chair
Associate Professor Daniel Eisenberg
Professor Lisa R. Lattuca
Associate Professor Robert D. Reason, Iowa State University

Dedication

This dissertation is dedicated to the memory of Dr. Eric Dey.

Eric, though you were gone before I wrote the first word, your influence is felt on every page.

Acknowledgements

This dissertation is the culmination of 30(!) years of education that would not have been possible with countless people believing in me, supporting me, and challenging me. Though I can never thank them all, I owe a debt of gratitude to all of who have impacted me in big way, and small ways, and in ways they will never know. My cousin Jamie Lewis teaching Sunday school and offering me a soccer ball if I memorized all of the books of the Bible, teaching me that a seemingly impossible tasks can be accomplished with determination and the right kind of incentive. Sandra Grindle, the best teacher a student could ask for, in a wheelchair and with a very broken leg cheering for me at the Indiana Geography Bee, showing me that what I worked toward mattered, even if it was weird. The teachers and staff at Montezuma Elementary School, recognizing my potential and taking a chance on a different way of challenging me. All of you and so many more are a part of this accomplishment, and I thank you.

But some people need some specific acknowledgement here. To Pat King, who advised me for seven years at Michigan and chaired this dissertation: Thank you for your guidance, you support, and your rigor. You never let me get away with imprecision, unsupported claims, or ideas that were, at best, half-baked. This continual push for improvement frustrated me, but it has also made me a better researcher, writer, and thinker. To the rest of my committee: Thank you for offering your expertise and being my thought-partners on this journey. You have all expanded my thinking in different directions, and both this dissertation and my future work are stronger for it. Other faculty helped me in ways they may not have even realized, both in and out of the classroom. Thank you especially to Jan Lawrence, Cindy Finelli, Daniel Little, Steve

Desjardins, Valerie Lee, and Susan Dynarski, all of whom (mostly unknowingly) have left important fingerprints on this dissertation. Classmates and other students were also essential, both in contributing to my growth as a thinker and a scholar, and also in being supportive and helping me mentally and emotionally get through this process. I am especially grateful to Kerri Wakefield, Julie Posselt, Rosie Perez, Nathan Harris, Eunjeong Ra, Inger Bergom, Brian Burt, Cassie Barnhardt, and Molly Ott. You have all helped me more than you can know.

Thank you also to Deborah Taub, who introduced me to higher education as a field of study and who helped me to understand the importance of this research; without that guidance, my professional life would have taken a very different path. Finally, thank you to Melissa Roderick, without whose support and generosity, I never would have been able to start pursuing new exciting work while I finished this dissertation. Both the new and the old work benefitted from one another in ways that were unexpected and that would not ordinarily have been possible.

And to my family, none of what I have accomplished would be possible without your love, support, and encouragement. My grandparents, Malcolm and Betty Goodwine and Ruth and Leslie Holsapple, all valued education, and they raised families around them that did the same. Because of them, I lived in an environment where learning and accomplishment were valued, recognized, and expected. It never occurred to me not to work hard in school, not to go to college, not to expect the best from myself, and I owe that to growing up surrounded by this wonderful family. You also taught me the value of integrity and hard work, and though I rarely live us to your standard, it gives me something to reach toward. My parents, Carl and Tricia, sacrificed for me, believed in me, and cultivated in me both the interests that brought me here as

well as the mindset to never question what I could accomplish. If I wanted to audition for a play that required nightly hour-long drives to rehearsal and back (which I did!), they didn't discourage me. Instead, they applauded me and never let me hear them complain about the late nights, long drives, and boring waits. I never considered whether something was appropriate to strive for, and for that I am grateful.

Finally, to Christopher: There simply aren't enough words.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Research examining student outcomes in college most often employs either a college impact framework (focusing on the organization and make-up of a college and participation in and experiences with various components of that college) or a developmental framework (focusing on students engagement with cognitive processes that are consistent with theoretically supported mechanisms of development). This study examines the value of integrating these two frameworks in the examination of the developmental of moral judgment during college.

Drawing on data from the Wabash National Study of Liberal Arts Education, this study examines the development of moral judgment in two samples of students at the end of their first and fourth years of college. Hierarchical linear models were first estimated using the developmental and college impact frameworks individually, and then with a framework that integrates the two. Changes in adjusted pseudo-R^2 and coefficient effect sizes were then compared across the models to assess the relative explanatory power of the three frameworks and the threat of omitted variable bias in the restrictive frameworks.

Variables consistent with both developmental and college impact frameworks were significant predictors of moral judgment in both samples; this remained true even after the models were integrated. However, few college experiences or institutional conditions were significant predictors of change in moral development. The variables with the largest estimated effects were students' precollege characteristics (such as academic ability and gender). Additionally, models estimated with an integrated framework provided more explanatory power, both before and after the large effects of precollege variables were taken into account. Integrated

models also yielded different effect sizes for most independent variables than the separate frameworks, indicating the presence of omitted variable bias.

This study demonstrates that using an integrated framework provides a broader and more detailed picture of the development of moral judgment in college. Researchers and educators alike are encouraged to conceptualize this development using frameworks that do not rely solely on participation in institutional programs or the cognitive demands of those programs, but, rather, to design and implement programs that take both approaches into account.

CHAPTER I: INTRODUCTION

The first years of the current century have been marked with a plethora of high profile moral failings in the business, political, and sports worlds that reverberated through the country. Corporate fraud and other unethical practices at organizations like Enron, Tyco, Arthur Anderson, and WorldCom shed light on the importance of preparing new professionals and future business leaders to consider moral and ethical issues in complex ways and behave ethically in difficult situations. More recently, the economic recession of the late 2000s and early 2010s has been partially traced to unethical business practices at financial institutions including Lehman Brothers and Bank of America, with devastating effects to American homeowners. Further, high-profile individuals like Martha Stewart, Bernie Madoff, Rod Blagojevich, Marion Jones, and Joe Paterno put very public faces on public and personal prices of unethical actions. In a 2011 Gallup Poll, 45% of Americans rated the state of moral values in the United States as poor; only 15% rated it as excellent or good (Saad, 2012).

College students are not simply watching ethical scandals and unethical behavior unfold in these high-profile and public cases. Engineering students also report witnessing their professors behave unethically and encourage and endorse unethical student behavior (Holsapple, Carpenter, Sutkus, Finelli, & Harding, 2012). Many studies have found that students are also engaging in unethical behavior themselves in large numbers in the form of academic dishonesty. McCabe and Trevino (1993) found that 84% of students reported cheating on written coursework at some point during their time in college, with students in business and engineering majors the

most likely to cheat. McCabe (1992) also found that almost 70% of students reported cheating an exam or major assignment. In a more recent survey of engineering undergraduate students by Finelli et al. (2012), almost 80% of students acknowledged that they had participated in some form of academic dishonesty. Cano and Sams (2011) conducted a study illuminates the scope of the problem: using a sample of business students, one group was exposed to a sensitizing experience of frequent messages about the importance of ethical behavior; this was designed to cause them to think more explicitly about the moral implications of their actions. A second (control) group was not exposed to the messages, and the two groups were both given a test in which they had the opportunity to cheat. Of the students who had not received the sensitizing messages about unethical behavior, 100% cheated on the test, as did almost 90% students from the sensitized group, suggesting that the high levels of cheating are resistant to the effects of at least some interventions. These high levels of academic dishonesty among students are particularly concerning since research has shown that cheating in college is highly predictive of unethical business practices after college (Dupont & Craig, 1996; Harding, Carpenter, Finelli, & Passow, 2004; Nonis & Swift, 2001; Ogilby, 1995; Sims, 1993).

Understanding the moral development of college students is of vital importance because those students will likely have a considerable impact on their world in the years after college. Boyer pointed to the importance of moral education for college students, stating, "College graduates often take leadership positions in both their employment settings and their communities -- positions in which they make decisions affecting the lives of others" (1987, p 3). More recently Shepherd, Macatangay, Colby, and Sullivan (2009) pointed to the moral and ethical imperative inherent in the ever-advancing technology of today. They argue that as technology grows more complex and its effects on the world become harder to predict, the

ethical issues faced by those who develop and use that technology have also grown in complexity and uncertainty. As higher education is the primary gatekeeper to the types of professional roles in which today's students will find themselves making decisions with the power to affect – and potentially harm – the rest of the society, higher education must play a strong role in preparing those students to consider the moral implications of those decisions.

As a result, colleges and universities have come under pressure from a variety of directions to address students' moral development; for example, 20 years ago the Wingspread Group on Higher Education (1993) issued the following call for action:

> Every institution of higher education should ask itself — *now* — what it proposes to do to assure that next year's entering students will graduate as individuals of character more sensitive to the needs of community, more competent in their ability to contribute to society, and more civil in their habits of thought, speech, and action (p. 9).

Other education advocacy groups have followed suit with similar calls (Association of American Colleges and Universities, 2002; National Association of State Universities and Land Grant Colleges, 1997), and politicians, policy makers, and industry leaders have echoed these demands. Even Congress emphasized the importance of integrating moral education into the college experience when the 1998 Amendment of the Higher Education Act of 1965 included language stating that character development should be one of the primary goals of American higher education.

Within the academy, administrators and educators have also positioned moral development of students as central to higher education (e.g., Casteen, Gibson, & Lampkin, 2007; Evans & Reason, 2001; Hersh & Schneider, 2005; Humphreys & Daveport, 2005; Young, 2003). In a survey of faculty, administrators, and student affairs professionals at 23 colleges and

universities, 86.9% agreed that helping students develop academic and personal integrity should be a focus of their institution and 70.8% said that same of ethical and moral reasoning (Dey, Antonaros, Ott, Barnhardt, & Holsapple, 2010).

Moral Development in College

Recognizing that moral development is an important outcome of a college education, colleges and universities have devoted considerable energy and resources to evaluating and reinvigorating their character development efforts (e.g., Colby, Ehrlich, Beaumont, & Stephens, 2003; Ehrlich, 2000). Institutions have implemented a plethora of programs on character education, and, in particular, curricular interventions aimed at cultivating students' character development. For example, courses and programs with the goal to educate students to be positive contributors to their local communities and the larger society have seen exponential growth, including practices such as service-learning, problem-based learning, and community-based learning (Campus Compact, 2009; Colby, et al., 2003; Sax, 2004).

There is evidence that colleges and universities are having a positive impact on the moral development of their students. In both of their extensive reviews of literature on the impact of college on students, Pascarella and Terenzini (1991, 2005) reported consistent evidence that students show increases in moral development during college, that those increases are larger than for those who do not attend college, and that development can be encouraged by specific experiences during college. Similarly, in their meta-analyses of research on the development of moral judgment (the most commonly studied aspect of moral development) during college, King and Mayhew (2002, 2004) reported that a long line of studies has consistently found a relationship between increased levels of schooling and the development of moral judgment. In an earlier test of college effects, Rest (1979a, 1979b), Rest and Thoma (1985), and Wison, Rest,

Boldizar, and Deemer (1992) found that students who attend college experience increases in moral judgment not experienced by their same-aged peers who do not attend college.

However, it is not enough to know that on average, attending college has a positive effect on moral development. If higher education is to take seriously the calls for greater attention to character development made by professional associations over the past two decades, it is of the utmost importance for educators and institutions to be able to identify the most effective ways to encourage that development in students. Particularly in an age of reduced education funding and scarce resources, it is imperative to understand what types of experiences are most effective in promoting moral development and how these might be different for different students and students attending different types of institutions.

Study Framework

In light of recent ethical scandals, the overabundance of unethical behavior on college campuses, and the role of higher education in educating moral citizens and leaders, this study focuses on better understanding college experiences that encourage students' moral development. More specifically, it considers ways that different conceptualizations of student change in the collegiate setting color our understanding of students' development of moral judgment during college. According to Pascarella and Terenzini (2005), college student outcomes – including moral outcomes – are most commonly studied using theories and models that fit into one of two frameworks: *developmental* and *college impact*. Developmental frameworks focus on the "nature, structure, and processes of individual human growth" (p. 17). Conversely, college impact frameworks focus on "change associated with the characteristics of the institutions students attend and the experiences students have while enrolled" (p. 18). Developmental frameworks typically place the individual at the center of the research question,

while college impact frameworks typically place the institution or a specific institutionally-sponsored program at the center.

These two distinct frameworks are rarely integrated in student outcomes research, and research on moral judgment is no exception. Most prior research on the outcome has used a college impact framework, and has examined narrow sets of variables and experiences. In their reviews of research moral judgment in college, King and Mayhew (2002, 2004) and Pascarella and Terenzini (2005) reported that researchers have examined the impact of aspects of the college experience, including academic discipline, curricular experiences, and cocurricular participation. However, in many of these studies they reviewed, researchers focus on one isolated characteristic or experience (e.g., participation in social fraternities and sororities, exposure to a curricular intervention like an ethics module, informal interactions with faculty), comparing moral judgment before and after the experience or comparing the moral judgment of students who had the experience with students who did not. In many studies, researchers did not consider the experience of interest within the larger context of students' experiences in their institution (a *college impact* framework) or the components of the experience that would theoretically be expected to encourage development (a *developmental* framework). Consequently, research on moral development has largely neglected ways in which those experiences that promote students' development interact across the students' experiences within their institution. Here, I argue that integrating these frameworks will allow researchers to draw a more complete picture of the breadth of students' experiences and the complex combinations of factors affecting moral development.

Research Questions

The purpose of this study is threefold: 1) to investigate developmental and college impact factors that affect students' development of moral judgment during college; 2) to compare the explanatory power of both of those frameworks in explaining moral judgment; and 3) to propose and test a framework for studying moral judgment that integrates the two frameworks. With these purposes in mind, I address the following overarching research question: *How does the integration of developmental and college impact frameworks explain changes in students' moral judgment during college?* More specifically, I address the following sub-questions:

1. What developmental factors predict the development of moral judgment during college?
2. What college impact factors predict the development of moral judgment during the college?
3. When integrated in to a single model, what developmental and college impact factors predict the development of moral judgment during college?
4. How does the explanatory power of the estimated models change when the developmental and college impact models are integrated?
5. How do the estimated effects of independent variables change when the development and college impact frameworks are integrated?

Definitions

It is important to note that the word "development" is used to refer to multiple aspects of college experiences and outcomes. In comparing developmental and college impact frameworks for researching student outcomes, Pascarella and Terenzini (1991, 2005) use the term to refer to

a framework for examining outcomes (which, as described above define as "nature, structure, and processes of individual human growth" (2005, p. 17) and contrast this with the institutionally-focused college impact framework. It is also used to describe processes of development (such as the evolution of moral thinking), as well as the desired outcomes of college that follow those processes (such as being able to make discerning moral judgments). For this dissertation, I use the term "developmental framework" to refer to the framework of researching student outcomes described by Pascarella and Terenzini (1991, 2005) which focuses on the experiences and internal changes of the student as central and identifies and analyzes predictors of change based on existing developmental theories. I use the term "development" to refer to the process through which a person organizes his or her meaning making in increasingly complex ways. Consequently, "moral development" refers to the increasing complexity through which individuals (here, students) organize their thinking and meaning making around moral issues. Finally, in regard to outcomes associated with moral development, I will refer to the specific outcomes themselves, such as moral judgment, moral sensitivity, or moral behavior, which are components of a person's morality (Rest, 1979). This study focuses primarily on the component of moral judgment, which is discussed in more detail in the next section.

Contributions of the Study

This study has the potential to make three main contributions to the larger literature on the development of moral judgment in college in addition to contributing to the literature on college outcomes more generally. First, by using a developmental framework to study the moral judgment, I am able to point to students' personal experiences within the college setting that encourage the development in ways that are supported by cognitive developmental theory in general and Rest's (1979a; Rest, Narvaez, Bebeau & Thoma, 1999b) model in particular.

Although other studies have examined some of these effects before, this study uses a broader and more comprehensive set of theoretically supported independent variables than previous studies have used. For educators, it provides guidance for developing educational programs that harness developmental theory in ways that can effectively encourage the development of moral judgment.

Second, this study examines the college impact factors that affect the development of moral judgment during college. I will introduce Terenzini and Reason's (2005) Comprehensive Model of Influences on Student Learning and Persistence to the study of moral judgment, which provides for a broader and more comprehensive set of independent variables than have been used in previous studies. The results of these comprehensive models will provide educators with evidence about the types of educational efforts that can help students develop moral judgment and help to guide individual educators and institutions to better address these issues. It will also provide support for the expansion of the study of moral judgment beyond its relationship students' individual experiences to a wider consideration of the institutional and organizational characteristics with the potential for affecting it.

Third, I integrate these more comprehensive sets of developmental and college impact variables into an integrated framework. Virtually no existing literature on the development of moral judgment uses this type of integrated framework in the organization or design of research. Doing so will provide researchers with a tool to better understand how the totality of a students' experiences at his or her institution, and suggest avenues for studying the interaction of college impact and developmental experiences in ways that are as yet unexamined.

Organization of Study

This dissertation is organized into five chapters. In this introductory chapter, I have presented a statement of the problem and the study's research questions. In the second chapter, I present a review of relevant literature, including discussions of college impact models of student change, developmental theories of moral development, and a review of the frameworks used in research on the development of moral judgment in college. In the third chapter, I present the methods that will be used to address the research questions, including a discussion of the Wabash National Study of Liberal Arts Education and the data used. In Chapter 4, I offer a thorough presentation of the results using the analyses proposed in Chapter 3, and in Chapter 5, I discuss implications of those results, giving particular attention to implications for educational practice and future research.

CHAPTER II: LITERATURE REVIEW

In this study, I propose to examine the development of moral judgment in college using three frameworks: college impact, developmental, and integrated. In this chapter, I offer the rationale for this approach by reviewing literature relevant to the investigation of college students' moral judgment and the addressing of the research questions outlined in Chapter I. First, I discuss the traditions of college impact research by outlining the early attempts by Feldman and Newcomb (1969) to organize and improve the research and then presenting several subsequent college impact models of student change (Astin, 1970a, 1970b, 1977, 1993; Pascarella, 1985; Weidman, 1989; Berger & Milem, 2000; Terenzini & Reason, 2005). Next, I discuss traditions of developmental research on student change, first discussing the tenets of cognitive-structural development and then outlining two related models of moral development often applied to college students (Kohlberg, 1976; Rest, 1979b; Rest, Narvaez, Bebeau, & Thoma, 1999b). Finally, I present the results of 20 years of research on the effect of college experiences on students' development of moral judgment, noting in particular the use of college impact, developmental, or integrated frameworks in the research and considering limitations of the existing body of literature.

College Impact Frameworks

The first of the two major frameworks for studying student change during college is the college impact framework. When distinguishing between college impact and developmental frameworks, Pascarella & Terenzini (2005) described college impact frameworks as those focus

on student change "associated with the characteristics of the institutions students attend and the experiences students have while enrolled" (p. 18). Thus, this framework considers the catalyst for student change to be the organization and make-up of the students college and the students participation in and experiences with various components of that college. I introduce this section of the literature review by presenting information about Feldman and Newcomb's (1969) review of literature that signaled the beginning of contemporary college impact research. I then present five college impact models – Astin's (1970a; 1970b, 1977, 1993) Inputs-Environments-Outputs (I-E-O) model and four others that were developed to extend Astin's work and include more detail and suggest causal mechanisms for ways that the college environment and student experiences encourage or inhibit student change. Feldman and Newcomb and the College Impact Foundation

The beginning of modern college impact research can be traced primarily to Feldman and Newcomb's (1969) book, *The Impact of College on Students*, a review and synthesis of studies on the impact of college and college experiences from the previous four decades. They reviewed more than 1000 studies of the impact of college attendance and college experiences on student outcomes. By compiling, reviewing, and analyzing these studies, they drew conclusions about the impact of college on students, described methodological limitations of the body of college impact research, and established the many of the future directions of research that are still being followed today. Their review laid the foundation that college impact research has built upon in the more than four decades since its publication.

Through their analysis of existing literature, Feldman and Newcomb (1969) offered nine conclusions on the impact of college on students:

- There are freshman-to-senior changes in a variety of outcomes, and these changes occur with "considerable uniformity in most American colleges and universities" (p. 326).
- There are patterns of differences in the incoming characteristics of students based on types of institution, and those differences affect the impact of college on student outcomes.
- Within institutions, different majors may have a differential effect on outcomes beyond the institutional effect.
- An institution can affect a student by helping him or her to maintain existing values and attitudes that may have otherwise changed.
- The impact of a student's interactions with individual faculty members may be high, but the impact of faculty on an institutional level on outcomes is limited.
- Small, residential, four-year colleges have the largest impact on student outcomes, a condition credited to the "relative homogeneity of both faculty and student body together with the opportunity for conditioning interaction, not exclusively formal, among students and between students and faculty" (p. 331).
- The impact of college on student outcomes is moderated by students' individual backgrounds and personalities.
- Changes in attitudes and values during college tend to persist after college. This is especially true if graduates live in environments that continue to support these changed attitudes and values, and they may even continue to change along the trajectory begun during college.

- A student's characteristics that influence his or her choice of institution and choice of academic and social subgroup membership while in college are most likely to be the same characteristics that those environments reinforce.

Each of these conclusions has been reinforced in subsequent college impact research. In addition to these conclusions, however, Feldman and Newcomb also pointed to several methodological limitations in the research they had reviewed. These limitations called into question the conclusions reached by individual studies; addressing these limitations led to the conceptual and methodological improvements that would become the foundation of the college impact models that followed. These limitations referred to many components of the research questions, designs, and interpretations; the five listed here are directly addressed in the college impact models discussed later in the chapter for the development of college impact models of student change that followed.

- *Overly simplistic hypotheses*: The largest conceptual limitation that Feldman and Newcomb (1969) found was that existing studies too frequently used hypotheses and research questions that were too simple to provide insightful or generalizable conclusions. Common hypotheses included claims that students were changed in certain ways by attending college or were changed in different ways by attending two different institutions or types of institutions. Missing in these hypotheses were fine-tuned differences between students, institutions, and institutional sub-cultures, which limited what could be learned from these studies. Feldman and Newcomb said of these simple hypotheses, "As the very least, it seems imperative to build in a consideration of the *specifics* of the backgrounds of the particular students as well as the *specifics* of the particular college environment" (p. 283).

- *Single-institution samples*: A large number of the studies sampled students from only one institution. These studies often then assumed that changes seen between the freshman and senior year (in either cross-sectional or longitudinal designs) occurred *because* the student attended that institution. This type of design ignores several other possibilities. First, students in the institutional sample may not change in ways that are different from students who do not attend college, yet studies rarely included a non-college control sample, a control sample that would need to be like the college sample in observable and unobservable ways. Even within the body of students who attended college, these studies often were used to demonstrate the effect of attending one specific institution or type of institution. Like the non-college sample problem, this ignores the possibility that students see the same effects in all institutions, not just the one that was included in the study. The design of the single-institution studies did not allow researchers to determine the true effect of attending an institution and what aspects of the institution might lead to that effect.
- *Lack of pre-college controls*: Many studies included in the review attempted to compare effects of different types of institutions without accounting for students' pre-college characteristics. For example, studies compared institution rates of attending graduate school without accounting for academic or aspirational differences of students beginning college which could be systematically different at different types of institutions.
- *Cross-sectional rather than longitudinal data*: One of the common research designs that Feldman and Newcomb identify in their review is a cross-sectional design, in which researchers collected data from freshman and seniors at an institution at the same

time to determine change during time at that institution, which was conceptualized as the impact of attendance. This was problematic for two reasons. The first is that it does not account for the sample mortality during the years between the freshman and senior years. If students who leave school before reaching their senior year are systematically different from those who do not, then the senior sample would exhibit differences from the first-year students simply because of the absence of those students. Second, more directly applicable to later models, this cross-sectional design does not account for potential pre-college differences between the students in the freshman and senior classes. If these students were systematically different before entering college – for example, if one year saw noticeably different admissions criteria or current events that caused differing social attitudes – those differences could be present in the cross-sectional data even if no change occurred during college.

- *Large groups that mask subgroup differences*: In the reviewed studies, it was uncommon for differential effects among student sub-groups to be compared or for student characteristics or most student experiences to be considered as moderating effects. Because of this, effects were most commonly seen – or not seen – based on entire samples. This can lead to the masking of differences among student subpopulations, and, in extreme cases, can lead to the conclusion that effects do not exist when they may be quite large for some students. It can also lead to the conclusion that effects exist for the overall student population, when they only exist for some subpopulations.

In light of their nine conclusions and the methodological limitations outlined above, Feldman and Newcomb (1969) offered a framing question improve future research and

encourage researchers to use more sophisticated methods and study designs. Rather than ask whether students change as a result of going to college – or attending a specific college – they suggest that the question should be more multi-faceted and nuanced, considering "… what kinds of students change in what kinds of ways, following what kinds of experiences, mediated by what kinds of institutional arrangements" (p. 5). This consideration of student change is made up of the four components – student characteristics, student experiences, institutional context, and outcomes – that form the basis of all major college impact models that have followed.

Astin's Inputs-Environments-Outputs Model

Since Feldman and Newcomb's review of studies addressing the impact of college on students and their subsequent recommendations for such research, the dominant framework for college impact has been Alexander Astin's I-E-O Model (1970a, 1970b, 1977, 1993). Astin had presented this model prior to the publication of Feldman and Newcomb's (1969) review (e.g., Astin, 1965; Astin & Panos, 1966); however, in his two 1970 articles in *Sociology of Education*, he related this model directly to Feldman and Newcomb's work and positioned it as a way to ameliorate the methodological limitations present in the studies they reviewed. These methodological articles related to the model signaled the beginning of the influence of the I-E-O model in the college impact literature.

Acknowledging many of the methodological limitations Feldman and Newcomb had outlined, such as a reliance on single-institution and cross-sectional data, a lack of consideration of student pre-college differences, and differential effects of college on different types of students, Astin suggested a model with the capacity to address all of those limitations. The I-E-O model provided a way of conceptualizing the complex personal and institutional characteristics that could affect the impact of the college experience on students in ways that had been largely

missing from the research. It does so by considering the effects of student characteristics, institutional environments, and student experiences (as well as interactions among all three) on outcomes.

The model (Figure 2.1) comprises three distinct components: Inputs, Environments, and Outputs. In the model, Outputs derive jointly from Inputs and Environments.

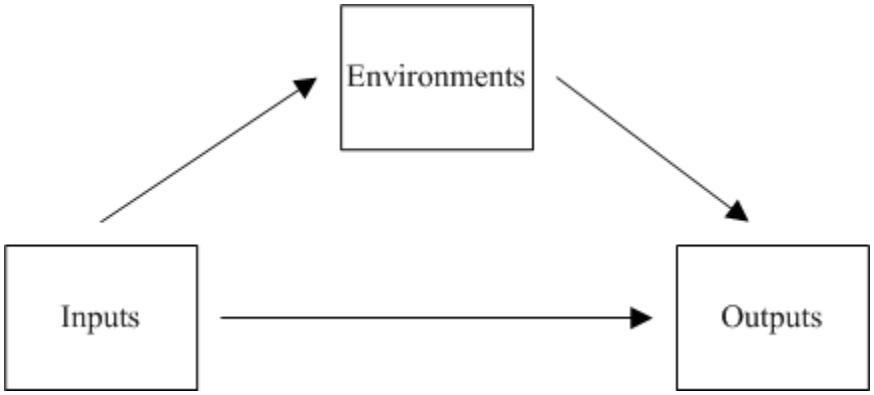

Figure 2.1. Inputs-Environments-Outputs Model (Astin, 1970a, 1970b, 1977, 1993)

Inputs are students' pre-college characteristics, "the raw materials with which the institution has to deal" (Astin, 1970a, p. 265). These inputs include usually stable characteristics, such as gender, race, or other sociodemographic characteristics, which allow researchers to understand student subpopulations and the differential effects of educational treatments on them. Inputs also include attitudes and values, aspirations, academic abilities, skills and talents, development, and other changeable factors. Accounting for these types of inputs allows researchers to consider change in them during college, and attribute that change to college

attendance or specific experiences during college. In the model, inputs are shown as affecting outputs both indirectly (by directly affecting environments) and directly.

Environments refers to the "aspects of the higher education institution that are capable of affecting the student" (Astin, 1970a, p. 225). He divided these environments into two types: characteristics of the total institution and characteristics of experiences and environments within the institution. The former includes factors such as size, student body make-up, organizational structures, academic policies and curricula, physical facilities, and other institutional characteristics. These institution-wide environment factors affect all students of the institution and help researchers to examine between-college effects, that is, the effect of attending one institution or type of institution compared to others. Environments that refer to differences within the institution include such features as student subcultures, pedagogical techniques of individual departments or classes, housing options, presence of student organizations and other co-curricular activities, and a host of other environmental factors and experiences that could be experienced by some students and not by others. These individual-specific environmental components allow for researchers to examine the effects of different student experiences and explain students' differential changes during college, even within the same institution. In the model, environments directly affect outcomes and are affected by inputs, with pre-college characteristics affecting both the institution that the student attends (between-college effects) and the experiences the student has at that institution (within-college effects).

Outputs refers to the student outcome that is being studied. Of note, the output is not simply the measure of the outcome (which can be a range of outcomes, including learning and development) but is also the amount the outcome has changed while the student has been in college. The outcome can be conceptualized in the model as the effect of the institutional

environment and student experiences on the outcome, after controlling for the students' pre-college characteristics.

Based on this model, Astin (1977) outlines nine effects that can be studied for any outcome:

- The direct effects on inputs on outputs;
- Effects of inputs on the college environment to which a student is exposed;
- Direct effects of within-college environments;
- Interaction effects of within-college environments;
- Interactions between inputs and within-college environments;
- Effects of between-college environments;
- Interactions among between-college environments;
- Interactions between inputs and within-college environments;
- Interactions between within-college environments and between college environments.

Research that considers these nine effects within the framework of the model can address the limitations of previous college impact research as discussed by Feldman and Newcomb (1969) and Astin (1970a; 1970b). It allows researchers to address the overarching question that Feldman and Newcomb presented as the future direction of such research by isolating the effects of institutions from the characteristics students bring with them to college, understanding how different types of institutions and different experiences within those institutions, estimating differential effects of the college experiences for different types of students, and understanding how different institutional factors and in-college experiences work together to encourage or discourage outcomes. More than 40 years later, his I-E-O model continues to form the basis for most contemporary college impact research.

The model provides a framework that can be used to consider the effect of any type of student and institutional characteristic and student experience; however, it does not provide detail about the types of institutional characteristics or students experiences that would be expected to affect different types of outcomes, nor does it identify causal mechanisms to explain how those characteristics and experiences cause the changes associated with them. This yields a model that is flexible so that researchers can adapt it to fit virtually any research question involving inputs, environments, and outputs; however, the model does not provide researchers with direction when considering which variables affect outcomes and why. To address this limitation of the model, several researchers have presented adaptations of the I-E-O model that provide more detail about the environments. In the rest of this section, I describe four of these adaptations.

Pascarella's General Causal Model

Pascarella (1985) conducted a review of studies that had considered the impact of college attendance and experiences in college on students' academic achievement and cognitive development. Like Feldman and Newcomb (1969) and Astin (1970a, 1970b), Pascarella pointed to significant methodological and conceptual limitations with that research; the model he presents in the paper is designed to enable researchers to limit those limitations. The model also explicitly points to new directions of research by laying out pathways for effects that had been understudied in the literature he reviewed.

For example, Pascarella (1985) acknowledges that there was limited evidence of differential impacts of education on different types of students. He lays out a variety of student characteristics that should be considered for potential differential effects: ethnicity, gender, age, socioeconomic status, aptitude, learning styles, academic preparation for college, personality

traits, and educational and occupational aspirations. In addition to differential effects based on these student characteristics, he also urged researchers to consider the differential effects of institutions and institutional environments, as well as interactional effects for different types of students within different types of environments. He said, "Determining what kinds of institutional environments maximize learning and cognitive development outcomes for specific types of students is a research issue that sorely needs attention" (p. 47).

Pascarella (1985) also calls for causal modeling in college impact research. Although in today's research the term "causal modeling" suggests approaches such as regression discontinuity, natural experiments, and other quasi-experimental research designs, Pascarella was instead calling for considering not just whether change occurred, by why and how it occurred. Rather than simply establishing that there was change during college or after exposure to certain types of college experiences, this type of research would "portray the system of direct and indirect influences in a causal system. Thus it is an attempt to understand the pattern of causal influences leading to a particular criterion, rather than simply attempting to predict that criterion" (p. 47). According to Pascarella, this approach also leads to model parsimony and fewer concerns of multicollinearity in regression models.

Thus, the model Pascarella (1985) presents (Figure 2.2), provides researchers with a "general causal model for assessing the effects of differential college environments on student learning and cognitive development" (p. 50). In addition to providing a more complex causal model, the model also accounts for the differential effects based on student characteristics and instructional environments and interactional effects between the two that he had identified as missing in previous literature. The model draws both on the literature review presented by Pascarella (1985), as well as his own previous work Pascarella (1980) and work by Feldman

(1971), Lacy (1978), Pace (1980), Walberg (1982), Walberg, et al., (1982), and Weidman (1984).

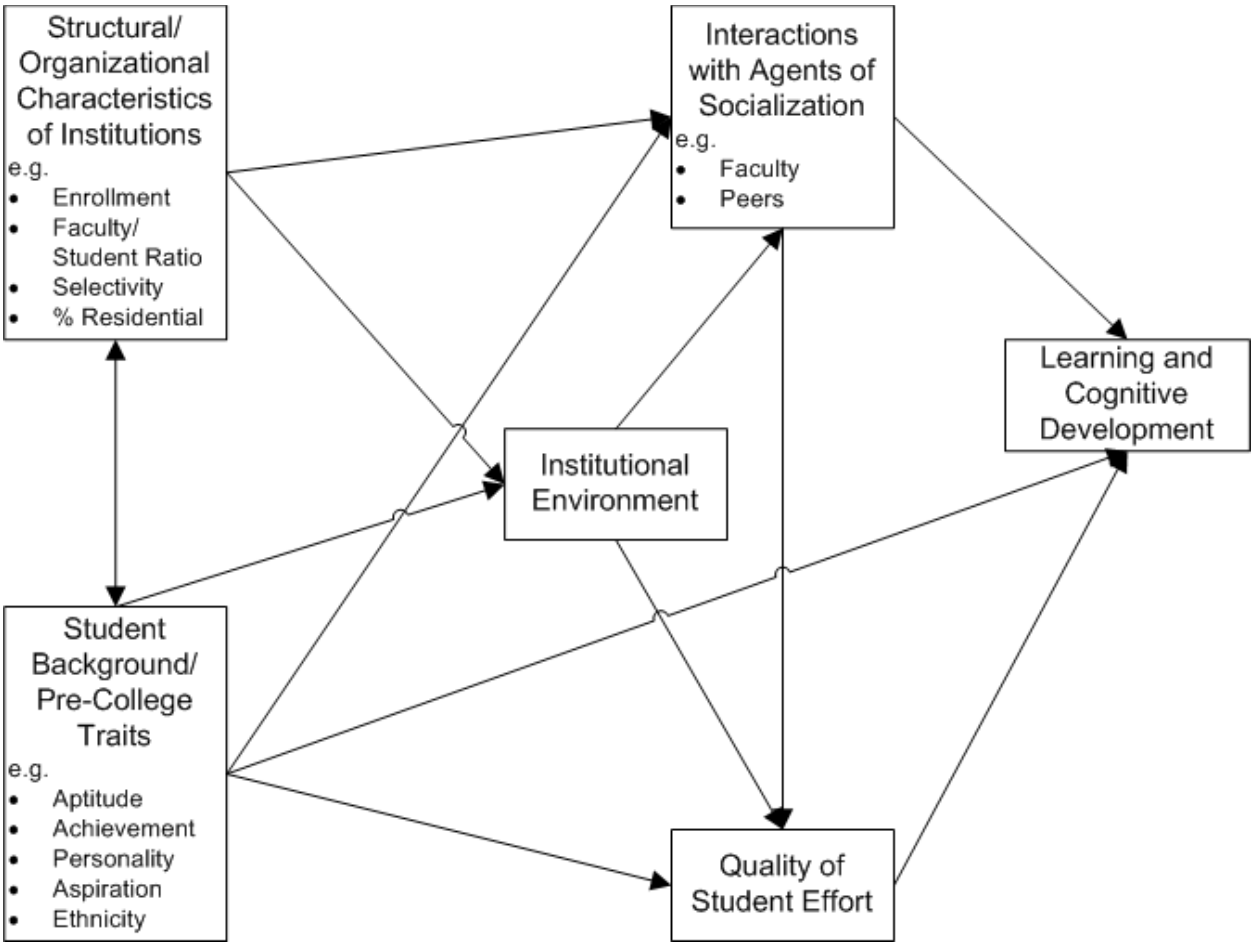

Figure 2.2. General Causal Model for Assessing the Effects of Differential College Environments on Student Learning and Cognitive Development (Pascarella, 1985)

The model (Figure 2.2) conceptualizes student outcomes as arriving from direct and indirect effects from five types of variables: student background/precollege traits, structural/organizational characteristics of institutions, interactions with agents of socialization, institutional environment, and quality of student effort. The first of these equates to Astin's inputs, and the final four are components Astin environments. The model depicts outcomes as

the direct effect of student characteristics, quality of student effort and interactions with agents of socialization, but also depicts a series of indirect effects that also affect the outcome through those three direct effects. Student characteristics and structural/organizational characteristics have both a direct effect on one another and direct effects on the institutional environment. In turn, all three of those variables have direct effects on a student's interactions with agents of socialization, and all dimensions except structural/organizational characteristics have direct effects on the quality of student effort. The direct effects on the outcome are from those dimensions relating to the students and their actions, but characteristics of the institution and its environment have an important role in effecting those actions.

Pascarella (1985) did not present this model as a complete, prescriptive explanation of the direct and indirect effects on student outcomes. Rather, he presented it as an example of a causal model that researchers could use a starting point for more complex empirical work, stating "its estimation should be expected to lead to more refined and accurate alternative models which better explain the causal structure in different contexts" (p. 49). For example, the model does not include students' course-taking patterns or cocurricular activities, which could be further explored in the additional research he encourages.

Weidman's Conceptual Model of Undergraduate Socialization

Like the researchers who had developed previous models in of college impact, Weidman did so to address what he saw as limitations in the previous research. Previous studies, he argued, were focused primarily on describing outcomes in detail rather than on "the development of comprehensive theoretical explanations for their occurrence or the building of conceptual frameworks" (1989, p. 289). Weidman turned his attention to explaining not *which* outcomes

college attendance promoted, but rather the processes that explained *how* college attendance led to those outcomes.

Weidman used the process of undergraduate socialization to explain how those outcomes developed, identifying career choices, lifestyle preferences, aspirations, and values as arising from that socialization. Drawing on theoretical foundations of adult socialization (Brim, 1966; Mortimer & Simmons, 1978) as well as his own prior research examining these socialization processes on college campuses (Weidman, 1984; Weidman, 1989; Weidman & Friedmann, 1984; Weidman & White, 1985), Weidman provided an expansion of Astin's I-E-O Model (1970a, 1970b, 1977). Like Pascarella (1985), he did so primarily by including more detail in the Environments portion of the model, but, unlike Pascarella, Weidman (1989) also brought extra-institutional factors and experiences to the conceptualization of student outcomes.

The Conceptual Model of Undergraduate Socialization (Weidman, 1989; Figure 2.3) conceptualizes outcomes arising from the interaction of four main domains: student background characteristics, parental socialization, non-college reference groups, and the collegiate experience. The student background characteristics and college experience mirror Astin's (1970a, 1970b, 1977, 1993) Inputs and Environments, respectively. However, Weidman expands the college experience to explicitly include the normative contexts and socialization processes to which students are exposed. He divides those normative contexts into formal and informal academic and social contexts. In these contexts, students engage in socialization processes through interpersonal interaction, intrapersonal processes, and social and academic integration. It is through socialization processes in these normative contexts that students are influenced by the normative order, expectations, and attitudes of members of their institution.

Weidman also includes in his model both direct and indirect effects arising from extra-institutional socialization. Both parental socialization and non-college reference groups, such as extra-institutional peers, employers, and community groups, directly affect outcomes, but they also have an indirect effect on student outcomes by affecting the student's collegiate experiences.

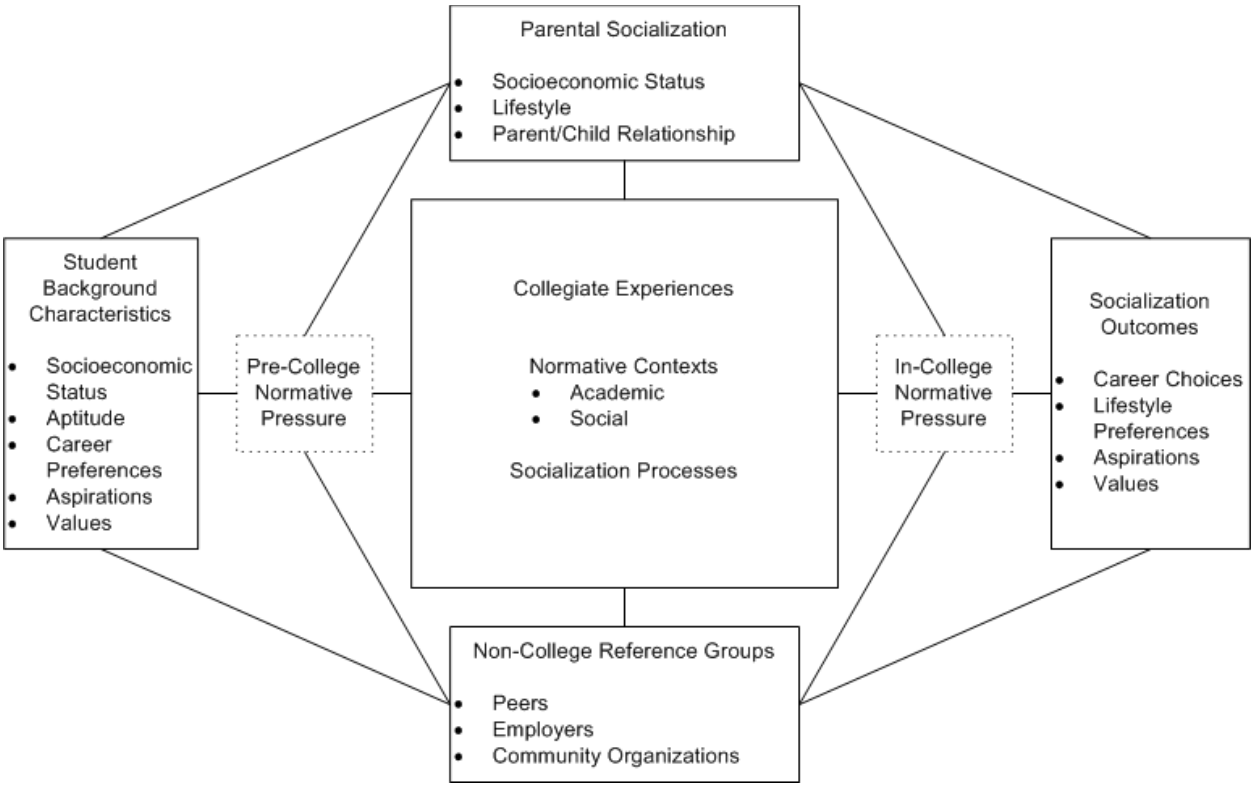

Figure 2.3. Conceptual Model of Undergraduate Socialization (Weidman, 1989).

This model is important to the study of the student outcomes for two main reasons. First, it explicitly brought family influences and other extra-institutional influences to the conceptualization of student change during college, acknowledging that being a member of a campus community does not bar students from being members of other communities. Second, it

made an attempt to explain the processes through which the change occurred within the institution. This model places membership in the campus community as the most important factor affecting how students change while they are in college, and it goes on to hypothesize the mechanisms – socialization processes and normative contexts – through which belonging to that community caused that change.

Berger and Milem's Conceptual Model for Researching Organizational Impact on Student Outcomes

Like Pascarella (1985) and Weidman (1989), Berger and Milem (2000) introduced a college impact model to address what they saw as important limitations in the student outcomes literature at the time. In their case, this limitation was relative lack of consideration of organizational characteristics in the research. They stated, "While organizational studies in higher education largely ignore the student, research on college impact generally ignores the influence of organizational factors on student outcomes" (Berger & Milem, 2000, p. 268).

Berger and Milem's Conceptual Model for Researching Organizational Impact on Student Outcomes (2000) starts from Tinto's (1993) assertion that "colleges and universities are organizations and organizational behavior does affect students" (Berger & Milem, 2000, p. 273). Tinto (1993) criticized previous attempts to incorporate organizational factors into the study of student outcomes for two main reasons: 1) that organizational models typically ignore student subcultures and student experiences as moderating variables in the effects of organizational variables on students; and 2) that organizational models assume that students interact with organizations in the same way rather than incorporating student-level differences. As Berger and Milem (2000) summarized Tinto, "these applications of organizational theory assume colleges

and universities to be organizationally monolithic in form, function, and in the way they affect students" (p. 273).

Other theorists have made similar critiques of the way that organizational theory is applied in a way that reifies organizations rather than focusing on the process of organizing (e.g., Hannan, et al., 1976; Pfeffer, 1982; Scott, 1977; Weick, 1969). From these critiques, Berger and Milem (2000) conclude that organizational models of student change should eschew that tendency for reification and instead focus on the behaviors of actors within the organization (i.e., faculty, staff, students) rather than ascribe behavior to the organization itself. For example, within a college or university, individual faculty and staff members provide instruction and teach individual students; the college itself does not provide any instruction or do any teaching. So while discussing the impact of organizational factors on student outcomes, it is appropriate to speak of organizational behavior by administrators and faculty *at colleges*; alluding to the organizational behavior *of colleges* inaccurately attributes agency to the college as an animate social actor.

In addition to this problem of reification, Berger and Milem (2000) also noted in the body of previous research a reliance on structural-demographic characteristics – such as institution size, selectivity, and student body demographics – to the exclusion of other types of organizational characteristics. They argued that in order to more fully understand the impact of institutions on student outcomes, researchers need to move beyond these types of factors toward and understanding organizational behaviors and culture.

With evidence that organizational behavior can and does impact student outcomes, and having criticized previous college impact research and models for largely ignoring the organizational dimension, Berger and Milem (2000) propose a college impact model that

explicitly includes measures of organizational behavior, structural-demographic features of the institutions, and student peer group characteristics as affecting student outcomes. Like most other expansions of Astin's (1970a, 1970b, 1977, 1993) I-E-O model, Berger and Milem's Conceptual Model for Researching Organizational Impacts of Student Outcomes expands upon the Environments component, making more explicit the ways that different components of the Environment (including organizational factors) interact and directly and indirectly affect student outcomes.

Berger and Milem's (2000) conceptual model for researching the organizational impact on college students comprises five conceptual dimensions: student entry characteristics, organizational characteristics, peer group characteristics, student experience, and student outcomes (See Figure 2.4). In the model, three organizational dimensions have either direct or indirect effects on student outcomes. Organizational behaviors and structural demographic features affect the types of students who attend the institution (directly affecting peer group characteristics) and student experience once students arrive on campus. The two also have a reciprocal relationship with one another. The third organizational dimension of the model – peer group characteristics – is also directly affected by the student entry characteristics. These individual student characteristics, peer group characteristics, and the student experience in turn have direct effects on student outcomes, while organizational behavior and structural demographic features exert an indirect effect on those outcomes.

The strength of this model is the inclusion of explicit and well-defined organizational characteristics within collegiate environments. It turns attention to the effects of environmental variables and goes further than its predecessors by outlining different dimensions of the institution itself that can affect those outcomes; accordingly, the authors suggest that researchers

collect and model extensive data about institution culture and behavior as part of studying student outcomes. Berger and Milem avoid institutional reification by focusing on actions taken by actors and the perceptions of individuals rather than claiming that entire organizations can be considers as "acting" in some specific way; instead, the organizational components of their model should be thought of as measures of the ways individual administrators, faculty, students, or other actors act within the organization.

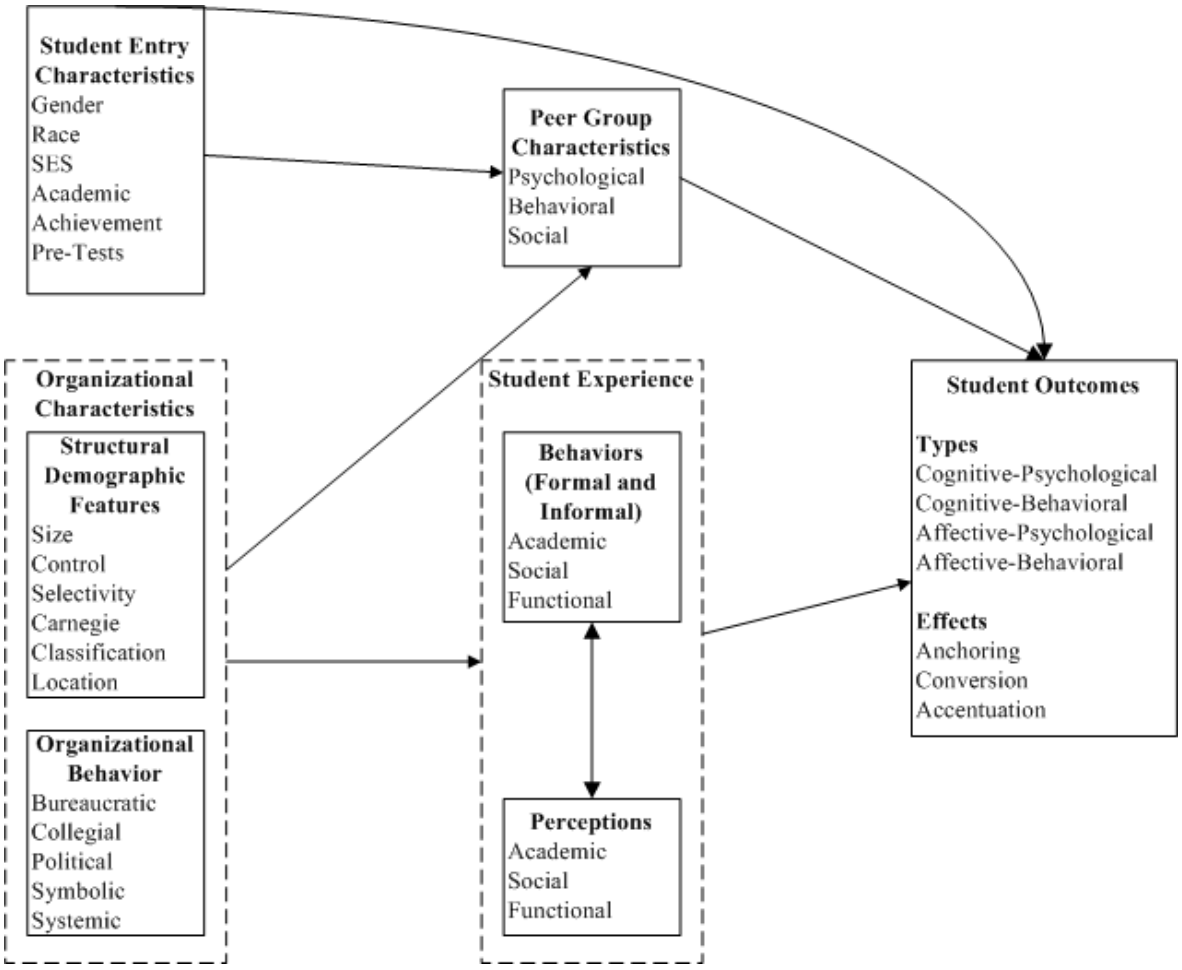

Figure 2.4. Conceptual Model for Researching Organizational Impacts of Student Outcomes (Berger & Milem, 2000)

Terenzini and Reason's Comprehensive Model

Terenzini and Reason (2005) proposed a comprehensive model of influences on student learning and persistence that consolidated aspects of previous models and added new, understudied dimensions (described below). The model was designed to move college impact research beyond what they saw as an atomistic, virtually atheoretical approach underlying much of the existing research. Too much of this research, they argue, relies on "an overly narrow conceptual, empirical, and practical view of the college experience and a myopic focus on discrete conditions, interventions, and reforms relating to student learning and degree completion" (Terenzini & Reason, 2010, p. 1). Their model (Figure 2.5) instead recognizes a much broader set of factors that they hypothesize as influencing student outcomes, including a more fine-grained consideration of the academic and classroom experiences of students and a broader consideration of the institutional environment and context.

The Terenzini and Reason (2005) model uses the basic framework of Astin's I-E-O model, but provides a more defined conceptualization of what comprises the environments component: the college experience. This component represents the organizational context (comprising internal structures, policies, and practices; academic and co-curricular programs, policies, and practices; faculty culture); the peer environment ("the ethos of the student body (which) is assumed to be distinct from any institutional ethos" (p. 11), and individual student experiences (including classroom experiences, out-of-class experiences, and curricular experiences). These components of the college experience are represented as affecting outcomes both directly and indirectly by influencing other experiences and moderating their effect on the outcomes.

There are several benefits to using this model to investigate college outcomes rather than less comprehensive models. First, the model explicitly incorporates a wide range of organizational factors, including faculty and peer environments, administrative policies and priorities, and academic practices as affecting student outcomes. This moves beyond the structural-demographic factors most commonly used in college impact research such as size, sector, selectivity, addressing, which many studies (e.g., Dey, Hurtado, Rhee, Inkelas, Wimsatt, & Guan, 1997; Milem & Berger, 2000; Pascarella & Terenzini, 2005) have found to have little predictive power on student outcomes.

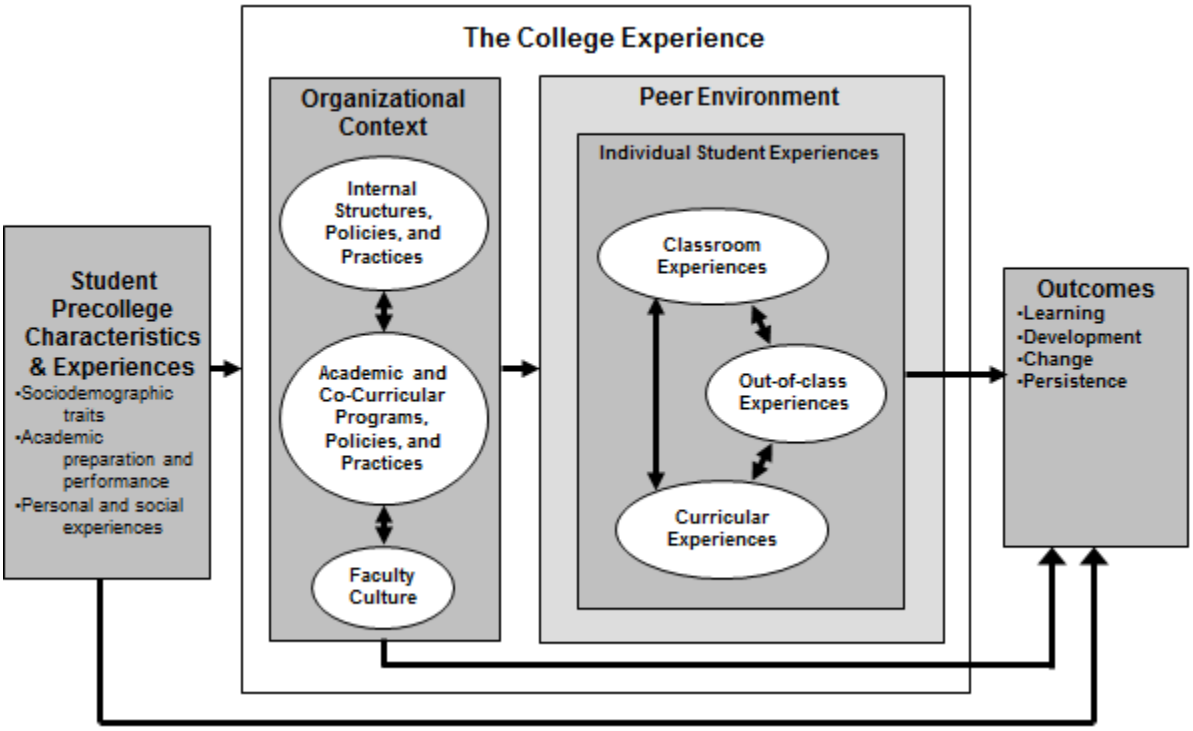

Figure 2.5. Comprehensive Model of Influences on Student Learning and Persistence (Terenzini & Reason, 2005)

Second, student experiences are explicitly divided into in- and out-of-class experiences, recognizing differences between the two and proposing an independent impact of each on student outcomes. These experiences are also presented as affecting one another, acknowledging that they may have both a direct and an indirect impact on outcomes.

Third, the peer environment holds a central focus of the model and is given comparable attention as that given to the organizational context. This component subsumes individual student experiences, acknowledging the presence of the environment created by the aggregate of student body characteristics (Holland, 1997; Strange, 1994; Strange & Banning, 2000). Astin (1993) found that interactions with peers were the most important factor affecting student change in college. Other empirical and theoretical literature has established the importance of students' microsystems (Bronfenbrenner, 1993) and that students often concurrently occupy more than one peer environment (Renn & Arnold, 2003). By contrast, Terenzini and Reason (2005) conceive of the peer environment as encapsulating the entire student body, and representing not only aggregated student characteristics, but also the dominant values, attitudes, beliefs, and behaviors of the student body.

Fourth, student characteristics are treated in a complex way. The model acknowledges the direct and indirect effects they can have on student outcomes by affecting the institutional environment of which students are a part and the experiences they have there.

Finally, the model can accommodate a wide range of disciplinary lenses and perspectives to better explain student change. In a follow-up to their introduction of the model, Terenzini and Reason (2010) state that because this model incorporates a broad array of influences on student outcomes, it can be adapted to fit sociological, psychological, economic, organizational, and other disciplinary lenses, which then can be incorporated into single studies.

In the years since they introduced the model, Terenzini, Reason and colleagues have demonstrated the utility of the model for addressing outcomes including academic competence (Reason, Terenzini, & Domingo, 2006), social and personal competence (Reason, Terenzini, & Domingo, 2007), interaction with difference (Reason, Cox, Quaye, & Terenzini, 2010) and persistence (Reason, 2009). Other researchers have also begun to use the framework to study a range of outcomes and contexts. Murphy (2010) used the model to examine the effect of living in residence halls on a range of student outcomes, including cultural competence and engaging in civic and campus life, and Lincoln (2009) used it to examine theological development in seminary students. Researchers have also used the model to examine a range of outcomes among first year students, including in studies spiritual development (Lovik 2011; Lovik & Volkwein, 2010), sense of community and retention in first year STEM-students (Falls, 2009), and the effects of part-time faculty on first-year community college students' retention (Jaeger & Hinz, 2008) and transfer to a four-year institution (Eagan & Jaeger, 2008). Holsapple, Carpenter, Sutkus, Finelli, and Harding (2012) applied the model to the study of ethics education in undergraduate engineering curricula. They found that aspects of the faculty culture, the peer environment, and academic policies and emphases all interfere with student learning and moral development in ethics education.

College impact frameworks for student change as described above seek to understand the complex relationships between students and institutional characteristics and student experiences within those institutions that lead to student change. I have shown that the roots of contemporary college impact research are grounded in Feldman and Newcomb's (1969) review of more of than 1,000 studies of student change and Astin's I-E-O model, which he positioned as a response to the methodological limitations found in their review (1970a, 1970b). Although Astin's model has

been the dominant model in college impact research, other writers, including Pascarella (1985), Weidman (1989), and Berger and Milem (2000) have presented extensions of that model that provide more detail about the causal mechanisms in the college environments that lead to student outcomes. More recently, Terenzini and Reason (2005) presented a model that incorporates dimensions of these previous models in an attempt to more comprehensively represent those dimensions of the environment. These models present one framework for examining the development of moral judgment during college; later in this chapter, I discuss the body of studies that have used them for this purpose.

Cognitive Developmental Models of Student Change

In categorizing the research on student change during college, Pascarella and Terenzini (2005) described the developmental framework as those that focus on the "nature, structure, and processes of individual human growth" (p. 18); this framework places the individual rather than the institution at the center of the research question. Using Pascarella and Terenzini's definitions, a developmental framework asks how students change *while* they are in college and participating in the college environment rather than *because* of that participation, which is the focus of a college impact framework. For example, researchers might use a college impact framework to examine the effects of a taking a service-learning class on moral judgment, whereas researchers would use a developmental framework to examine how the different types of experiences and cognitive activities within that service-learning class affect moral judgment.

In the developmental framework, the examined experiences within the institutional programs (such as the service-learning class) are guided by student development theories. Torres (2011) described developmental theories as those that describe new and more complexly organized ways of making meaning of knowledge and experiences, which develop in orderly,

sequential ways. Overton (2010, as cited by King and Kitchener, in press) presented five criteria that distinguish development from other types of change:

1. The change transforms an open, self-organizing, and self-regulating structure or system; this occurs through interaction with physical, biological, and sociocultural environments;
2. The change is orderly and sequential;
3. The change is in the direction of less to more complexity;
4. The change brings to that complexity new properties that were not previously seen in the system;
5. The change is relatively irreversible, making development relatively permanent.

Cognitive-structural developmental theories that are commonly used in higher education program development and research fit those criteria, and developmental frameworks for studying student outcomes are marked with the same focus. In this section, I discuss cognitive-structural development theories, describing the definition of development in this family of theories and the mechanisms they suggest trigger development. I then discuss Kohlberg's theory of moral development and Rest's concept of moral judgment in depth.

Cognitive-structural theories have been used to describe late adolescent and adult cognitive development (e.g., Baxter Magolda, 1992, 2001; Belenky, Clinchy, Golderberger, & Tarule, 1986; Kegan, 1994; King & Kitchener, 1994; Perry, 1970; Piaget, 1970). This paradigm has also been used to describe moral development (Kohlberg, 1976; Rest 1979b; Rest, Narvaez, Bebeau, & Thoma, 1999b).

A feature of cognitive-structural theories is the existence of what Piaget (1970) referred to as *structures d'ensemble* (and which Overton referred to as systems), or the mental

frameworks people use to make meaning of information and experiences. It is these meaning-making structures that change during development, consistently moving from simplistic ways of making meaning, to more complex, nuanced frameworks. The focus of these theories is on these changing structures of meaning-making, rather than the specific meaning that is made (i.e., the content); that is, cognitive-structural perspectives focus on the basis of a person's rationale and underlying assumptions that inform their ideas and decisions, not the content of the ideas and decisions. Two students using similar frameworks could arrive at different decisions, and two students using different frameworks could arrive at the same decision. King (2009) provides an illustration of this distinction by considering how students determine the candidate for whom they will vote in an election. Two students may decide to vote for the same candidate, yet one may do so after careful consideration of her own beliefs and values and determining the candidate that best represents what is important to her, while the other believes that her parents know best and follows their suggestions in voting. So although the *content* of their decisions (the candidate for whom they voted) is the same, the meaning-making *structure* employed (basing one's decisions on an examination of one's own values versus basing them on the absolute word of trusted authorities) belie different levels of development. This focus on structure rather than content fits with the educational perspective that colleges and universities should help students develop the "habits of mind" to grapple with new and complex ideas and dilemmas rather than simply lead students to desired points of view (Mezirow, 1997).

Through a cognitive-structural framework, development can be seen as a continual process of assimilation, disequilibrium, and accommodation (Piaget, 1973). Assimilation is the process under which individuals encounter new information, ideas, and perspectives and integrating those into their existing meaning-making structures, "rounding them out and

contributing to their expansion" (Evans, 2003, p. 187). However, when students encounter new information that they cannot fit into their existing structures, they enter a state of disequilibrium, the distress caused by encountering information that does not conform to current ways of making meaning. Disequilibrium sometimes triggers accommodation, the creation of new meaning making structures that fit the new information, and thus help the student reestablish a state of equilibrium. This continual process of disequilibrium and accommodation leads to development of individuals' abilities to making meaning of their experiences in more complex ways. This process is not easy for college students; rather it is slow, hard work (King, 2009) and students often find working through disequilibrium and accommodation to be a very emotional experience (King, Baxter Magolda, Barber, Kendall Brown, and Lindsay, 2009; Perez, Shim, King, & Baxter Magolda, 2011; Pizzoloto, Chaudhari, Murrell, Podobnik, & Schaeffer, 2008).

This process of disequilibrium, assimilation, and accommodation is represented in different ways in different developmental theories, but it follows a similar trajectory across theories. Development in all cognitive-structural theories is presented as sequential (but not linear) process, meaning that as individuals develop, they do so in a manner consistent with progressively complex levels of development demonstrating identifiable characteristics in meaning making. Theorists refer to these patterns of development in different ways, such as stages (King & Kitchener, 1994), ways of knowing (Belenky, Clinchy, Golderberger, & Tarule, 1986), schema (Rest et al., 1999b) and positions (Perry, 1970), but in each, individuals develop from simpler to more complex ways of making meaning over time. These increases in complexity happen as individuals mature; they are age-related but they are not age-dependent (King, 2009). In all of these cognitive-structural models of development, individuals progress through the levels as they age, with their own experiences and contexts affecting the speed at

which they develop. In all of these theories, there are identifiable characteristics that are present in meaning making at one level that are different from the characteristics of meaning making at other levels.

Although these theories describing development as progression through increasingly complex ways of meaning-making, that progression should not be thought of as linear and direct within any individual; rather, individuals' development trace the general patterns of the theory, but may experience development in a much more complex way, drawing on different levels depending on the situation and progressing through levels in a less linear path. More than two decades before Perry's (1970) developmental theory ushered in the study of college student development, Piaget (1948), offered a caveat for considerations of developmental trajectories:

> These stages must of course be taken for what they are worth. It is convenient for the purposes of exposition to divide the children in age-classes or stages, but the facts present themselves as a continuum which cannot be cut up into sections. This continuum, moreover, is not linear in character, and its general direction can only be observed by schematizing the material and ignoring the minor oscillations which render it infinitely complicated in detail (p. 27-28).

Developmental theories describe a process of development that is sequential and directional, but it is important to acknowledge that development is not a simple lock-step, straight-line process. Instead, development often happens through what Rest (1979a) described as complex stages; rather than engaging in meaning-making consistent with one level of development 100% of the time, people utilize different types of meaning-making in different contexts and in the same context at different times. So while individuals typically display a dominant level of meaning-making consistent with one developmental level and that dominant level demonstrates a change toward more complexity over time, the way an individual makes meaning at any given time or in any given context may not be consistent with that dominant level. Since Perry introduced his theory of cognitive development in 1970, researchers have

presented several models of the cognitive development of students during college (Baxter Magolda, 1992, 2001; Belenky, Clinchy, Golderberger, & Tarule, 1986; Kegan, 1994; King & Kitchener, 1994; Perry, 1970; Piaget, 1970). This paradigm has also been used to describe moral development (Kohlberg, 1976; Rest 1979b; Rest, Narvaez, Bebeau, & Thoma, 1999b). Rooted in a Piagetian perspective, these theories share several commonalities, including development of more complex ways of making-meaning of information and experiences and the importance of encountering disequilibrium in triggering changes toward that increased complexity. Those cognitive developmental concepts have also been applied to moral development by Kohlberg and Rest, and researchers and educators have extensively applied these theories to college students.

Kohlberg's Theory of Moral Development

It was Kohlberg's belief that moral education should focus on developing students' moral reasoning abilities (the structure of their moral thinking), rather than teaching students to hold specific moral beliefs (the content of their decisions): "…indoctrination is neither a way to teach morality nor a moral way of teaching…true morality involves *making thoughtful decisions* about values which may be in conflict…" (Power & Kohlberg, 1986, p.16). In line with this view of the place of moral reasoning at the forefront of moral development, Kohlberg (1976) relied on Piaget's cognitive-structural approach to development when developing his theory of moral judgment development. In doing so, he separated moral reasoning (the structure of moral reasoning) from moral behavior, establishing moral reasoning as a process worth studying in its own right rather than simply as a means of influencing behavior. This distinction acknowledged the difference between content and structure in moral judgment, with a focus on "general organizing principles or patterns of thought rather than specific moral beliefs or opinions" (Colby & Kohlberg, 1987, p. 2). Consistent with Piaget's claims, Kohlberg observed that development

of moral reasoning took place in response to disequilibrium that challenged current ways of thinking (Kohlberg, 1972) about moral issues.

Kohlberg's (1976) theory of moral development posits that development occurs sequentially through six distinct stages; this development is unidirectional and individuals move simplistic to more complex moral reasoning. Kohlberg grouped these stages into three levels – preconventional, conventional, and postconventional moral reasoning – which in turn each comprised two stages of moral development.

Preconventional moral reasoning (Stages 1 and 2) is marked by an emphasis on self-preservation (acting morally to avoid punishment), which is directed toward maintaining positive relationships with the people important to him or her (such as parents), rather than an emphasis on organizing a larger society. In Stage 1 (*Obedience and Punishment*), to be moral is to follow rules and expectations to avoid punishment and other negative consequences for oneself. This stage is marked by a lack of consideration for others and their points of view in moral decisions. Stage 2 (*Naively Egoistic Orientation*) is marked by individuals making moral decisions in ways that will serve their own best interests. Morality is seen as an exchange process, with individuals entering into agreements where one may serve the needs of others, but also as a way of serving one's own needs.

Conventional moral reasoning is marked by a focus on the rules, standards, and expectations of the larger community (acting morally to sustain the established social order and avoid social chaos). In Stage 3 (*Good Boy/Nice Girl Orientation*), individuals recognize the existence of shared social expectations and agreements that can take precedence over individual desires so as not to disappoint those who are important to the individual. Moral decisions are made to fit within those social expectations. In Stage 4 (*Authority and Social Order Maintaining*

Orientation), individuals continue to see morality as fitting into the needs and expectations of the larger society, and moral decisions are made to uphold the existing social order. Moral decisions are made based on the importance of following society's rules and expectations, without regard to the justice of those rules and expectations.

Postconventional moral reasoning is marked by the use of moral principles as the basis for making moral decisions. These moral principles transcend the needs of both the individual and the social order by placing a clear focus on questions of moral right and wrong and fundamental human rights. In Stage 5 (*Contractual and Legalistic Orientation*), individuals recognize that the basic rights and values of society may conflict with rules and laws. In this stage, moral decisions are made by considering "their long-term consequences for the welfare of each person or group in society" (Colby & Kohlberg, 1987, p. 30). Stage 6 (*Conscience or Principle Orientation*) sees the individual following engaging in reasoning that is grounded in moral principles, even when those may conflict with both society's laws and rules and higher moral values. This stage is marked by a belief that these self-chosen values and moral principles are rational and just and that they are built on universal principles of morality and justice.

Kohlberg's theory of moral development forms the foundation of the most commonly used conception of students' moral development during the college years (Evans, Forney, Guido, Patton, & Renn, 2010). In presenting his theory, Kohlberg presented a picture of moral development in which people adopt understandings of morality that become more complicated and allow for more nuance as they develop, shifting from placing utmost importance on personal benefits to the maintenance of societal law and order to the value of societal justice and overarching moral principles. This view of moral development underlies many of the educational interventions colleges and universities enact to encourage moral development in students.

The most prominent critique of Kohlberg's model of moral development came from his Harvard University colleague Carol Gilligan (1982). Kohlberg had developed his model on research using only men, with no women in his longitudinal sample. Gilligan argued that as a consequence, his model was an inherently male-biased approach to morality, privileging what she called a justice-based approach over a care-based approach to morality favored by many women. Gilligan's perspective on morality "pushed connection to others, not universalism and individualism, into the forefront of moral reasoning" (Evans, Forney, Guido, Patton, & Renn, 2010), with a focus on relationships, love, and caring for others as the hallmarks of moral development. Some researchers found support that women (including college students) were more likely than their male peers to engage in moral meaning making with this care orientation (Mennuti & Creamer, 1991; Lyons, 1987; Stiller & Forrest), although other research suggested that it was not gender that was the distinguishing factor, but rather the types of moral dilemmas the subjects most commonly encounter in their daily lives (Clopton & Sorell, 1993; Hare-Mustin & Marcek, 1988; Mednick, 1989).

Despite a Gilligan's reasonable critique of Kohlberg's sampling, his model has not been demonstrated to be gender-biased against women. In fact, women consistently score slightly higher on assessments of moral development using Kohlberg's framework (Thoma, 1986). King and Mayhew (2006), in their review research on college students' moral judgment, found that this finding repeatedly replicated, with female students most commonly scoring higher or no differently than their male counterparts. This suggests that, while Gilligan's framework is an important way to consider morality, Kohlberg's model does not suffer from the gender differences that she originally argued.

Rest's Contributions to the Study of Moral Development

A former student of Lawrence Kohlberg, James Rest expounded upon Kohlberg's theory of moral development and presented the Four Component Model of moral processes that helps explain the connection between moral reasoning (or moral judgment, as he termed it) and moral behavior. In addition, a second thrust of Rest's research and theorizing has focused on the development of moral judgment, and Rest and colleagues subsequently framed moral judgment in the context of moral schema.

Four Component Model. The Four Component Model of Morality (Rest, 1979b, 1982, 1983, 1984, 1986) represents four distinct psychological processes that must occur for an individual to enact moral behavior in a given situation: moral sensitivity, moral judgment, moral motivation, and moral character. *Moral sensitivity* refers to a person's ability to recognize a situation as a moral one (i.e., that the situation affects others) and awareness of potential ways of responding to that moral dilemma. *Moral judgment* refers to the process of weighing different options and determining which course of action is morally right depending on one's definition of fairness and structure for meaning making. *Moral motivation* refers to the process of weighing one's moral values against non-moral values that are at play in the dilemma and potential decision. This process allows a person to determine whether moral action is compatible with social pressures of other personal concerns. *Moral character* refers to the process by which a person decides how to enact the selected moral action, creating a plan of action and developing the fortitude not to deviate from that path.

In the Four Component Model, moral judgment is the component that derives from Piaget's cognitive-structural tradition and Kohlberg's theory of moral development. Rest (1979a) replaced Kohlberg's (1976) strict stage model with what he referred to as a complex stage model.

Kohlberg's theory conceptualizes development as linear through the six stages, meaning that an individual's level of development can be thought of as either characteristic of one stage or as being in transition from one stage to the next; Rest characterized this as a simple stage approach. In this type of movement, once an individual moves to from one stage from the next, he or she uses that the new type of moral reasoning exclusively, leaving behind the reasoning of the previous stage. Rest's complex stage approach to moral judgment development relaxes the rigidity of Kohlberg's theory; it conceptualizes an individual's movement through the developmental stages as changing the predominant way that he or she engages in moral judgment. The complexity of that judgment increases consistently and uni-directionally, as in Kohlberg's model; however, an individual may reach back and use the moral judgment predominant at previous stages while in any more advanced stage. In this complex stage model, how an individual reasons through one moral dilemma does not necessarily demonstrate how he or she would reason through other dilemmas.

Moral schema theory. In more recent work, Rest and colleagues moved beyond conceptualizing moral judgment in Kohlberg's terms, and instead have adopted what they refer to as a neo-Kohlbergian approach (Rest, Narvaez, Bebeau, & Thoma, 1999a; Rest, et al., 1999b; Rest, Narvaez, Thomas, & Bebeau, 2000). Central to this current approach to moral judgment is the role of cognitive schema, meaning making structures that reflect individuals' "conceptions of the institutions and role-systems in society" (Rest, et al., 2000, p. 385). They describe moral judgment as meaning-making around issues of macro-morality (moral issues that affect larger societal issues) rather than micro-morality (moral issues that affect individual actions and interactions).

A cognitive schema refers to an existing mental structure of a person's knowledge that is used to interpret a new experience or stimulus. When exposed to a new experience, a person will try to match that new experience to an existing cognitive schema, using the organization of prior knowledge to make sense of the new experience (Wenger & Wheatley, 1999). Cognitive psychological research suggests that the majority of a person's understanding of experiences comes from these sorts of stored schema rather than by conscious thought (Narvaez & Bock, 2002; Reber, 1993). The schema are mental representations of general cases of phenomena, and by representing what is similar about related instances of a phenomenon, they allow individuals to simplify how they make meaning of the reality they encounter, allowing them to assimilate new stimuli into existing schema and fill in missing information.

Schema are arranged in in three basic levels of increasing complexity (e.g., Bock & Narvaez, 2002; Derry, 1996; DiSessa, 1993):

- *Memory objects*, specific relationships or experiences, such as a trip to the dentist;

- *Cognitive fields*, an interrelated set of memory objects, such as a schema for dentists as a group of people; and
- *Mental models*, wide-reaching meaning making models that combine applicable cognitive fields to make sense of the different parts of an experience, such as the importance of dental health and dentists' role in this undertaking.

Together, these three types of schema provide the building blocks to draw on previous experiences and knowledge to understand new situations and moral dilemmas. Those who make meaning of moral problems in more developmentally advanced ways are able to access more memory objects and cognitive fields and arrange them in more complex ways (Narvaez & Bock, 2002): "In terms of mental architecture, the expert has castles of knowledge, while the novice may have a bare foundation" (p. 300). Rest, et al. (1999a) suggest that what had previously been thought of as movement through developmental stages can be re-imagined as development of more complex ways of activating moral schemas. When making sense of moral problems from a schema perspective, individuals construct mental models that organize different memory objects and cognitive fields that together allow him or her to address the new problem.

Applying schema theory to the study of morality, Rest and colleagues (Rest, et al., 1999a; 1999b; Rest, et al., 2000) identify three moral schemas that people use to consider macromorality:

- *Personal Interest Schema*: Those using this schema make sense of moral dilemmas based on the personal stakes and consequences involved for themselves; it corresponds to Kohlberg's second and third stages of moral development. The primary moral schema for pre-adolescents, the personal interest schema places personal benefit over societal benefit and

"induces prudential concerns and concerns for those with whom one has an affectionate relationship" (Rest, et al., 2000, p. 387).

- *Maintaining Norms Schema*: Those using this schema make sense of moral dilemmas based on existing rules and authority figures; it corresponds to Kohlberg's fourth stage. Rest et al. (2000) suggest five priorities used in moral judgment by those who rely on the Maintaining Norms schema: 1) the need for people to following generally accepted social norms; 2) the application of those norms to all members of a society without exception; 3) norms that are clear and codified (e.g., into rules and laws); 4) an expectation of partial reciprocity – that the rules are applied to all and that all will follow them; and 5) a societal hierarchy and established authority that will enforce those rules. Activating this moral schema, then, leads to making meaning of moral questions in a way that privileges the upholding and uniform enforcement of societal rules and expectations beyond all else as a way of organizing an orderly society.
- *Postconventional Schema*: Those using this schema to consider a moral problem place an emphasis on "shared ideals, which are reciprocal and are open to debate and tests of logical consistency, and on the experience of the community" (Rest, et al., 1999a, p. 307). This schema corresponds to Kohlberg's postconventional moral thinking – his fifth and sixth stages of moral development. Instead of emphasizing established societal norms and rules, this schema acknowledges that those rules and norms may be flawed, and that morality should look beyond those rules to broader values. They suggest four priorities for the Postconventional schema: 1) a recognition that just because societal rules are set up in a certain way does not mean that they *should* be set up in that way; 2) an appeal to restructuring societal rules in some new and idealized way; 3) an emphasis on sharable

ideals, those that could be jointly held by a range of people; and 4) an expectation of full reciprocity – that all will follow the rules and that the rules themselves are fair and unbiased.

Although the application of schema theory to moral development moves farther away from Kohlberg's original strict stage conceptualization, Rest and colleague's neo-Kohlbergian conceptualization shows its cognitive-structural roots and shares important features with this early model. As in more traditional cognitive-structural models, schema become more complex – or more advanced – through Piaget's (1973) processes of assimilation, disequilibrium, and accommodation. As individuals encounter new stimuli that fit into existing schema, they use those stimuli to expand the schema and fill in missing information. However, when an individual encounters new stimuli that do not readily fit into existing schema, he or she may be thrown into a state of disequilibrium, which can result in the expanding of or making more complex existing schema. It is this process that leads to the development of more advanced moral judgment schema.

Educational experiences do not, however, always lead to the development of more complex schema; students can meet the disequilibrium by forcing new information into their pre-existing mental frameworks. Although she did not explain it in terms of cognitive schema, Boyle-Baise (1998; Boyle-Baise & Sleeter, 1998) described this effect in studies of pre-service teachers experiences with service-learning. When middle class college students spent time working with impoverished children, some responded to learning about and experiences first-hand the students' conditions by developing more complex understandings of poverty and were less likely to blame the students or their parents for their positions. However, for other students, the experience reinforced their pre-conceived ideas that the children had poor, uncaring parents and were disadvantaged because of their parents rather than their socio-economic conditions. An

educator can help students to work through this disequilibrium between their existing schema and the new information to expand their poverty schema to account for new potential reasons for poverty, leading to increasing schematic complexity.

Schema represent the structure of moral meaning-making rather than the content of the moral decisions that students make. That is, it is possible for two people who use the same moral schema to come to different decisions when encountering moral dilemmas, as is it possible for two people to come to the same decision by using different moral schema. For an example in the collegiate context, consider the case of NCAA student-athletes considering the morality of accepting benefits (such as cash payments from athletic boosters) that are not allowed under NCAA rules. A student-athlete who predominantly relies upon a Maintaining Norms moral schema could recognize that accepting the benefits violates the regulations that organize intercollegiate athletics and view accepting the payment as not moral, while another student-athlete using the same schema could determine that it is moral because accepting the extra benefit is a common occurrence and fits within the accepted norms of intercollegiate athletics.

Moral judgment refers to the way that individuals make meaning of the moral dilemmas they encounter and determine what they consider to be the morally ideal response to those dilemmas. Rest and his colleagues rooted their work in the concepts first developed by Piaget and Kohlberg, but adapted those concepts to present the development of moral judgment as the development of more complex schema for understanding moral issues.

Assessing moral judgment with the Defining Issues Test. Rest (Rest 1979b) also developed the Defining Issues Test (DIT), which has become the most common tool for assessing moral judgment in higher education research (King & Mayhew, 2002; King & Mayhew, 2004; Pascarella & Terenzini, 2005; Rogers, 2002). The DIT (since revised as the DIT-

2; Rest, Narvaez, Thoma, & Bebeau, 1999b) presents respondents with a series of five moral scenarios that are designed to activate the respondents' moral schemas. For example, in one scenario, respondents are presented with the case of a sick woman, Mrs. Bennett, who wants to die, and a doctor who must decide whether to give her enough medication to commit suicide. Respondents are then asked to judge and rank the importance of certain considerations when determining how the doctor should handle the dilemma, including whether Mrs. Bennett's family approves of her decision to die, and whether the doctor could be held legally liable for her death.

The DIT[1] is an assessment instrument rooted in the theoretical underpinnings of cognitive-structural development. It places an emphasis not on the content of a decision (here, what the doctor should do), but rather on its structure. This is seen through the factors they deem as important in making a decision, which reflect the schema they use to make a decision about the dilemma. The instrument is used to determine the extent to which individuals use personal interest, maintaining norms, and postconventional moral schemas when responding to these scenarios. Respondents may apply different schema to different considerations within the same dilemma and to the different types of dilemmas they encounter on the assessment tool.

The DIT provides a range of information about how students make meaning of the moral scenarios and dilemmas presented in the assessment instrument. Most commonly, the P-score and N2 score have provided an assessment of respondents' use of higher levels moral judgment. The P-score is a measure of how frequently respondents employed principled moral reasoning when evaluating the DIT's scenarios (Rest, 1979b). In the 1999 revision of the DIT (the DIT-2), that score was supplemented with the N2 score, which in addition to assessing respondents' use

[1] The Defining Issues Test and the Defining Issues Test-2 are similar instruments to assess moral judgment, and they are used in the same way in the literature on the development of moral judgment in college. Except when the distinction between the two versions of the instrument is essential, I used the term Defining Issues Test (and the abbreviation DIT) to refer to both versions of the instrument.

of higher level moral reasoning, also accounts for the degree to which they eschew lower level reasoning (Rest, Thoma, Narvaez, & Bebeau, 1997). Rest and colleagues determined the N2 score to perform as well or better than the P-score on each of the following comparisons: internal reliability, longitudinal change, correlations with education, correlations with moral comprehension, correlations with life experiences, gains from moral education interventions, and correlations with attitudinal and behavioral assessments. Since its development, this N2 score has become "the index of choice" (Thoma, 2002, p. 239), and it has been used in the majority of studies examining the influence of college experiences on students' development of moral judgment.

An additional score, the D-score, was also developed to assess moral judgment using more complicated scaling techniques; however, this measure has been shown to be a weaker measure than both the P-score and the N2 score (Evens, 1995; Rest, Thoma, Narvaez, & Bebeau. 1997). The DIT also allows for the assessment of aspects of moral meaning-making beyond moral judgment, including religious orthodoxy, decisiveness, and humanitarian/liberal perspectives, utilization of concepts of justice. These measures help assess issues of construct validity, but rarely have been used in research on the impact of college on moral development.

The N2 score and its P-score predecessor have provided researchers with an instrument to assess the impact of college and college experiences on students, and researchers have made it a commonly used tool. King and Mayhew (2002, 2006) provide extensive reviews of this literature, and I provide my own review of a portion of it later in this chapter. Beyond this literature on higher education, both versions of the Defining Issues Test have also been used extensively in a broad range of fields. Despite this widespread use, there has been concern that the DIT is not an independent measure moral judgment, but instead is a proxy for constructs such

intellectual ability, verbal ability, or political identity; meta-analyses of these studies using the DIT, however, have shown these to be correlated with but independent from P-score and N2 (see Thoma, 2002, 2006, for a thorough discussion of this evidence).

Among college students, several other characteristics have shown to be correlated to students' assessed moral judgment, even before the influence of college experiences are taken into account. In their reviews of the literature on the development of moral reasoning among college students, King and Mayhew (2002, 2006) summarize the results of several studies that examine the effects of students' precollege characteristics on assessments of moral judgment. The studies they reviewed pointed to a strong relationship between students' precollege academic abilities and moral judgment: those students who scored higher on the SAT, PSAT, written assignments, and Terman's (1973) Content Mastery Test, as well as those with higher high school class rank, earned higher moral judgment scores. Of note, high school GPA was the one measure of academic ability that did *not* show a significant relationship with moral judgment in the four studies in their review that included it (Green, 1981; Mentkowski & Strait, 1983; Quarry, 1997; Steppe, 2002). Also, their reviewed studies also showed that students who identified as politically or religiously liberal had higher levels of moral judgment.

For other characteristics, the relationships were less clear. Studies that investigated age among traditionally-aged college students or socioeconomic status found no significant relationships between these factors and moral judgment. Several studies included investigations of the impact of race and ethnicity, but King and Mayhew (2002, 2006) could identify no clear pattern in these findings. Finally, despite the concerns discussed earlier that Kohlberg's conceptions of moral development were gender-biased, almost all of studies that King and Mayhew reviewed showed either no differences between male and female students or that female

students had higher levels of moral judgment, a finding that is consistent with most empirical studies of gender differences in moral judgment (these broader gender findings have been discussed earlier in this chapter).

A college impact framework of student change focuses on the institutional characteristics and students' experiences within those institutions. A developmental framework, on the other hand, places the emphasis on the individual considering the experiences and internal processes that lead to change. Drawing on the earlier work or Piaget, a long line of developmental theorists have presented explanations of the cognitive development of students during college. While these theories differ in foci, they support the importance of encountering disequilibrium in the development of more complex ways of understanding new experiences. This is also true of the dominant theories of college student moral development (Kohlberg, 1976; Rest, 1979a; Rest et al., 1999b).

The Impact of College on Moral Judgment

In the previous sections of this chapter, I have discussed the college impact and developmental frameworks used to study student changes during college, paying specific attention to Terenzini and Reason's Comprehensive Model of Influences on Student Learning and Persistence (2005) and the cognitive-structural models of moral development developed by Kohlberg (1976) and Rest (1979a, Rest et al., 1999b). In this section, I turn to the literature on the effect of college experiences and examine how results in those studies suggest a range of experiences and institutional characteristics encourage or inhibit the development of moral judgment in college. I also present the way that college impact, developmental, and integrated frameworks are used in that research.

King and Mayhew (2002, 2004) and Pascarella and Terenzini (2005) have published extensive reviews of research on moral development in college, with King and Mayhew focusing exclusively on research on the development of moral judgment. These reviews present a wide breadth of research on the outcome and suggest that a variety of college experiences have the potential to encourage or inhibit the development of moral judgment. I do not attempt to reconstruct the reviews conducted by these authors; to do so would be both redundant and not consistent with the purpose of this chapter. Instead, I rely on these earlier reviews to help provide guidance in identifying and organizing relevant research. King and Mayhew (2002, 2004) drew on Astin's I-E-O (1970a, 1970b, 1977, 1993) (discussed earlier in this chapter) to present a conceptual model for the research they reviewed. This conceptualization presents literature as investigating the relationship of moral judgment with three other types of variables: student characteristics, collegiate contexts, and other collegiate outcomes (cognitive, identity, and social). Pascarella and Terenzini (1991), on the other hand, draw on their own framework, first presented in the first edition of *How College Affects Students*. They organized studies as those that examine the net effects of college, between-college effects, and within-college effects. As this study is focused specifically on the effects of students' college experience, I rely on the components of the authors' models that describe that experience; these are King and Mayhew's collegiate contexts and Pascarella and Terenzini's between-college and within-college effects, which all encompass the components of the College Experience in Terenzini and Reason's (2005) model discussed earlier in this chapter. Each set of authors identified categories of research within the research corresponding to the college experience. King and Mayhew presented research examining the relationship between moral judgment and institutional type, academic discipline, curricular experiences, and co-curricular experiences; Pascarella and

Terenzini are more fine-grained in their presentation, categorizing research institutional type, moral development interventions, service-learning, major field of study, extracurricular-peer involvement, intercollegiate athletic involvement, off-campus learning experiences, and interaction with faculty,

Based on the categories of research identified in these earlier reviews, I searched for published research that examined the relationship between moral judgment and each of those types of experience. Both King and Mayhew (2002, 2004) and Pascarella and Terenzini (2005) included unpublished studies (including conference proceedings and dissertations). I, however, have limited my review to only those studies that were published in peer-reviewed journals. After conducting my own literature search, I also consulted the reviews by King and Mayhew (2002, 2004) and Pascarella and Terenzini (2005) to confirm that applicable studies in their reviews were also included here. In addition, for six of the studies that were part of the Wabash National Study of Liberal Arts Education (WNS) (Mayhew, Seifert, & Pascarella, 2012; Mayhew, Seifert, Pascarella, Nelson Laird, & Blaich, 2012; Padgett, Johnson, & Pascarella, 2012; Mayhew, 2012; Martin, Hevel, Asel, & Pascarella, 2011; Mayhew, Seifert, Pascarella, 2010), I consulted supporting documentation about the scales and mega-scales used (Pascarella & Colleagues, 2008).

This search yielded 55 studies using the DIT and DIT-2 to examine the impact of the college experience on students' moral judgment. Due to the wide range of discipline- and field-based journals, I do not claim that this search is exhaustive; however, analysis of the articles in this review reached theoretical saturation and information redundancy (Strauss & Corbin, 1990), suggesting that the sample of reviewed papers was sufficient for analysis.

After identifying studies for the review, I sorted those studies based on the independent

variables relating to the college experiences that the studies used to predict the development of moral judgment. The college impact, developmental, and integrated frameworks refer to the overall design of the study, but that framework is determined by the independent variables used by the researcher. Thus, I classified each independent variable as consistent with a college impact, developmental, or integrated framework, and then identified the framework employed by the classification of those independent variables. Based on the discussions of the frameworks earlier in this chapter, the criterion I used to determine consistency with a *developmental framework* if it examined experiences in which students engage in the cognitive processes that were consistent with theoretically supported mechanisms of developmental change (i.e. increasing complexity over time). The most frequent examples of this were encountering disequilibrium and considering class material or other information or experiences in more complex and nuanced ways. The criterion used to determine consistency with *college impact framework* was if the variable described students' experiences with college characteristics or participation in programs offered by the institution. The criterion I used to determine consistency with an *integrated framework* was if the independent variables drew on both developmental and college impact frameworks; in the studies I reviewed, the only use of variables consistent with an integrated framework were those in which researchers constructed factors that included both developmental and college impact variables. However, I also considered a study to employ an integrated framework if it included at least one independent variable consistent with each of the developmental and college impact frameworks. The way these three frameworks are used in these reviewed studies are discussed in more detail in the remainder of this chapter.

College Impact Frameworks

A large majority of the studies that have investigated the development of moral judgment during college have done so utilizing variables consistent with college impact frameworks. Of the 55 studies included in this review, 51 included at least one variable consistent with a college impact framework, and 44 of those utilized only such variables (and none consistent developmental or integrated frameworks). In these studies, researchers used variables applying to almost all dimensions of the college experience included in Terenzini and Reason's (2005) Comprehensive Model of Influences on Student Learning and Persistence: organizational context, faculty culture, peer environment, out-of-class experiences, classroom experiences, and curricular experiences. Although the results sometimes conflict, as reported below, these studies provide evidence of the potential of aspects of the college experience to positively influence the development of moral judgment and support the use of college impact frameworks for studying that development.

Organizational context. As discussed earlier in this chapter, several authors, including Milem and Berger (2000) and Terenzini and Reason (2005), have stated that most of the research on the effect of organizational factors on student outcomes has focused on structural-demographic features of institutions, such as size, selectivity, and institution type. Terenzini and Reason's comprehensive college impact model expands organizational impacts beyond these structural-demographic features to place a focus on what institutions and institutional actors *do* rather than what the institutions *are*, specifying internal structures, policies, and practices; curricular and co-curricular programs, policies, and practices; and faculty culture as affecting students' outcomes. Despite this acknowledgment of the important organizational factors beyond structural-demographic features, almost all of the research on the impact of organizational

context on the development of students' moral judgment has examined the impact of these features. The most commonly studied has been the effect of attending schools with different Carnegie classifications or religious affiliations. An additional study considered the effect of one aspect of faculty culture on students' development. I now discuss the results of these studies.

Institution type. Results of the studies of these institutional effects present conflicting results. Most studies of these institutional effects have found no significant differences in the development of students based on the type of institution they attend. For example, using data from the Wabash National Study of Liberal Arts Education (WNS), Padgett, Johnson, and Pascarella (2012), Mayhew (2012), and Martin, Hevel, Asel, and Pascarella (2011), found no significant differences in the first-year development among students attending research universities, liberal arts colleges, or regional universities. This lack of significance, however, may be the result of the small number of institutions and a lack of statistical power; the first two studies examined 19 institutions and the third examined only 11, and the numbers of institutions were as low as three per type. Using a larger sample in a secondary analysis of data collected at 65 institutions, Maeda, Thoma, and Bebeau (2009), found that institutions with a stronger research focus and higher degrees offered (for example, bachelor's degrees compared to associate's degrees) did have students with statistically significantly higher levels of moral judgment.

The literature also shows contradictory results about the effect of attending a religiously-affiliated institution. In their analysis, Maeda, Thoma, and Bebeau (2009) found no significant differences between students attending institutions with a religious affiliation, and Traiser and Eighmy (2011) similarly found no differences between students attending seven public institutions compared to six private, religiously-affiliated institutions. By contrast, Elm,

Kennedy, and Lawton (2001) found that students attending a Catholic and Evangelical Christian institution had significantly higher moral judgment scores than those attending public institutions; however, they noted that this difference might be due to the fact that the two religiously-affiliated schools in the study were stronger academically than the two public schools. In the one study that has examined differences among religiously-affiliated schools, Good and Cartwright (1998) found that students attending a Bible college had significantly lower levels of moral judgment than those attending a more mainstream Christian liberal arts college. Both the small numbers of studies examining the effects of institutional type and religious affiliation and the small size of the institutional samples in most of those studies make if difficult to reconcile the conflicting results and make any definitive claims about the different effects these institutional characteristics have on students' development of moral judgment.

Faculty culture. Although I found no studies that considered the effects of internal structures, policies, and practices or curricular and co-curricular programs, policies, and practices, one recent study did examine the effect of faculty culture on students' moral judgment. Among several other aspects of first-year student's experiences on at their institutions, Padgett, Johnson, and Pascarella (2012) included as an independent variable a factor representing students' experiences with good teaching and high quality interactions of faculty. That factor included a subscale that assessed students' perspectives on faculty interest on teaching and student development, which included items such as how students' perceive if faculty have interest in students and helping students grow in ways beyond academics, and if they are excellent teachers, have genuine interest in teaching, and are willing to spend time discussing things of importance to students with them outside of class. Although this larger factor was not significant in the authors' analysis, they assessed only one aspect of faculty culture (and did so as

part of a larger piece of their analysis) to use this one study to assume that there is no link between faculty culture and students' moral judgment.

In Terenzini and Reason's (2005) model, the organizational context of a student's college or university is conceptualized as playing a major role in students' outcomes during college. In the extant research on students' development of moral judgment, however, this context is understudied and the results from the research are inconclusive. The studies discussed here that do examine organizational context almost exclusively focus on the structural-demographic characteristics of institutional type and religious affiliation. Because of the small samples as well as methodological inconsistencies, it is difficult to use these studies – as well as the one study that examines faculty culture – to draw conclusions about the role of the organizational context on moral judgment.

Peer environment. As early as Feldman and Newcomb's (1969) book, students' interactions with their peers have played a central role in students' change during college, and Astin (1993) stated, "the student's peer group is the single most potent source of influence on growth and development" (p. 398). Terenzini and Reason (2005) conceptualized an influence of peers beyond students' individual interactions with their friends, classmates, and other peers. The peer environment is made up of the "dominant and normative attitudes, beliefs, values, and expectations that characterize a campus' student body" (p. 11). Although other studies (discussed later in this section) examine the effects a student's individual interactions with other students, only one study has considered the effect of this larger peer environment. In their secondary analysis of DIT-2 data from 65 institutions, Maeda, Thoma, and Bebeau (2009) found that students attending institutions with a predominately conservative political orientation were

assessed with lower moral judgment (measured by N2 scores, described below) even when controlling for individual- and institution-level variables.

Individual student experiences. By far, the most frequent component of Terenzini and Reason's (2005) model to be examined in the research on the effects of students' experiences on their development of moral judgment is the individual student experiences; this is comprised of classroom experiences, out-of-class experiences, and curricular experiences. Terenzini and Reason suggest that these experiences are the most influential in affecting student outcomes; this is supported by the multi-level analyses of Maeda, Thoma, and Bebeau (2009) and Mayhew (2012), which both suggest that more than 80% of the variance in students' scores on the DIT-2 is at the individual level rather than the institutional.

Out-of-class experiences. Fewer studies have examined the effects of out-of-class experiences on moral judgment. Most commonly, these have examined the effect of participation in specific types of student activities and cocurricular involvement. Traiser and Eighmy (2011) found that the number of cocurricular activities in which a student participated had a positive effect on moral judgment, despite the mostly consistent finding across the studies of the lack of a significant effect of specific activities and involvement on students' moral judgment. Researchers have found non-significant effects on moral judgment from working on- and off-campus (Padgett, Johnson, & Pascarella, 2012; Martin et al., 2011); joining social fraternities and sororities (Padgett, Johnson, Pascarella; Martin et al.); participating in varsity intercollegiate athletics (Padgett, Johnson, & Pascarella; Martin, et al.; Baldizan & Frey, 1995; Traiser & Eighmy, 2011); participating in student government (Brown-Liburd & Porco, 2011); and living in a residence hall (Martin, et al.).

There were, however, exceptions. In a study of almost 400 accounting majors at seven institutions, Brown-Liburd and Porco (2011) found a statistically significant positive effect on moral judgment for students who participated in a national accounting honorary fraternity, an effect the authors suggests may be from the focus on accounting best practices, including ethics, service work, and other prosocial activities. Similarly, they also found significant positive effects of volunteerism for students. These findings, though limited to one study, suggest that involvement in out-of-class volunteering and other prosocial activities can positively influence moral judgment development.

Out-of-class experiences include not only participation in student activities, but also the interactions that students have with one another in formal and informal settings. Four studies (Padgett, Johnson, & Pascarella, 2012; Martin et al., 2011; Mayhew, Seifert, & Pascarella, 2012; and Mayhew, Seifert, & Pascarella, 2010) used data from the Wabash National Study of Liberal Arts Education to assess the impact of interactions with peers on moral judgment. Using the same nine-item factor that included items on students' cocurricular involvement, friendships, and the influence of peers on personal growth, all four studies found no significant effect. Finger, Borduin, and Baumstark (1992) did find that students' informal social interactions have a positive effect on moral judgment. While they did not test the mechanisms that led to these effects, they did hypothesize, "Social activities with friends and other acquaintances may encourage and stimulate role-taking and group problem solving and may ultimately promote more mature judgment" (p. 222).

Classroom experiences. While studies into the effects of out-of-class experiences on students' moral judgment have shown few significant effects, the research on the effects of classroom experiences is more promising. Terenzini and Reason (2005) define students'

classroom experiences as the specific activities, interactions, assignments, and pedagogies they experience as part of their academic work. In recent years, several studies have examined a wide range of these experiences on students' moral judgment. Most commonly, these studies have examined the change in students' moral judgment during the course of one class, although a smaller number of studies have tried to assess the way the totality of students' classroom experiences effect their moral judgment during the first year of college.

Researchers have examined the effects of a range of pedagogical approaches have affected moral judgment. For example, studies have found positive effects of taking courses that involved a service-learning component (Bernacki & Jaeger, 2008; Lies, Bock, Brandenberger, & Trozzolo, 2012; Goodman, Duffy, & Heffernan, 1994) and using films to teach ethics (Self, Baldwin, & Olivarez, 1993; Loui, 2006; Sheppard & Young, 2007). These studies neither investigated the mechanisms through which those pedagogies may lead to development nor used control groups for comparisons, but they do suggest the potential of both strategies in encouraging students' development of moral judgment.

Other studies have tried to compare the effects of different ways of teaching ethics and other moral issues. Cain and Smith (2009), for example, randomly assigned 124 students to one of four different discussion methods for a class on pharmacy law and ethics: face-to-face discussion, online non-anonymous discussion, online anonymous discussion, and no discussion. Over the course of the semester, the three discussion groups used their assigned method to discuss the moral issues covered in class with their classmates. They found that students in the two online groups – both non-anonymous and anonymous – experienced gains in the N2 score from the beginning to the end of the semester; for the anonymous group, the increase was by more than one-third of a standard deviation. Though they did not test this hypothesis, the authors

suggested this may have been because the asynchronous nature of the online discussions, compared to the face-to-face discussions gave students more time to reflect on their thoughts and the thoughts of their classmates while participating in the discussion.

Bunch (2005) took advantage of a natural experiment comparing the moral judgment change of religiously conservative and fundamentalist divinity students in three semesters of a course on Christian ethics. In each of these semesters, the material and instructor were the same, but the amount of formal discussion of moral issues changed from none to seven hours to 30 hours. There was no significant change in moral judgment in the semesters with either no or seven hours of discussion, but in the semester with 30 hours of discussion, students' N2 scores increased by almost one-half of a standard deviation.

Not all of these comparison studies showed significant effects for pedagogical methods. Auvinen, Suominen, Leino-Kilpi, and Halkama (2004) asked 54 nursing students to retrospectively report how often different instructional methods had been used to teach them ethics in their four years of college, and they found no relationship between students' moral judgment and the frequency with which they had been taught using lectures, case studies, discussion, group work, and individual work. This lack of a significant relationship, however, may be due to the small sample in the study and the lack of variation in the reported instructional methods.

Two additional studies used WNS data to consider the ways that the teaching methods used by all instructors during a students' first year of classes, with potentially conflicting results. Both Martin et al. (2011) and Mayhew, Seifert, and Pascarella (2010) examined the effect of a scale representing students' experiences of "good teaching and high quality interactions with faculty" (Pascarella & Colleagues, 2008). This factor included subscales that assessed the extent

to which students were exposed to prompt feedback and clear and organized instruction in their classes (along with two additional subscales assessing non-classroom experiences), and Mayhew, Seifert, and Pascarella found a small, but significant, positive effect on moral judgment. However, while Mayhew, Seifert, and Pascarella used a sample comprised of students from all 19 institutions in first wave of the WNS, Martin et al. restricted their sample to only those students at the 11 institutions with a significant social fraternity and sorority presence. With this restricted sample, they found no significant effects. This discrepancy points to the importance of institutional factors as potential moderating influences on the effects of individual student experiences.

Curricular experiences. In Terenzini and Reason's (2005) model, classroom experience refer to teaching methods to which students are exposed, and curricular experiences refer to the actual classes that students take and the academic opportunities they have, including their majors and their socialization into those academic fields, their course-taking patterns, and their participation in other academic experiences, such as internships and co-operative learning. While classroom experiences refers to the *how* of instruction, curricular experiences refers to the *what* of instruction. More research has been published about the effect of curricular experiences on moral judgment than any other component of the model; this research primarily focused on student majors and types of courses they have taken.

A large number of studies have considered how taking an ethics course – or a course with an ethics module – can affect students' moral judgment. These studies have yielded mixed results, but together provide support for the potential of these courses to positively affect students' development of moral judgment.

In the only study to used propensity score analysis to estimate causal effects of taking different these courses with a focus on ethics and morality, Grunwald and Mayhew (2008) examined changes in moral judgment for students who had taken one of four courses with a moral component to the course material. These classes – in moral psychology, moral philosophy, service-learning, and intergroup dialogue - had in common "the instructors' intentional use of moral content for communicating messages about how to respond to contemporary social issues using an explicit or implied moral frame" (p. 764). They found no significant differences in moral judgment development during the semester for students in the morally-focused classes when compared to students in an introductory sociology class; this lack of effect was present in the non-causal and causal models they present. However, when Mayhew and King (2008) examined the same set of classes using non-causal methods, they did find differences among the effect of the four morally-focused classes, finding positive effects for students taking the moral philosophy, moral psychology, and service-learning classes, but not the intergroup dialogue class. These differences suggest that it might be the idiosyncratic nature of the specific classes studied that are leading to the effects rather than the focus on moral issues.

Other studies have looked simply at whether students experienced significant changes in moral judgment during the semester in which they were exposed to ethics education. Several studies found statistically significant positive effects of semester-long ethics courses (Jagger, 2011; Jagger & Strain, 2007; Abdolmohammadi & Reeves, 2000) and ethics modules within other classes (Wilhelm & Czyzewski, 2012; Latiff, 2000; Loe & Weeks, 2000; Jones, 2009; Traiser & Eighmy, 2011), though others have found no significant effects from ethics courses or modules (Earley & Kelly, 2004; Ponemon, 1993; Armstrong, 1993; Auvinen et al., 2004; Kaplan, 2006; Fleming, Romanus, & Lightner, 2009). Traiser and Eighmy also found no

significant relationship between the number of courses related to ethics that students took and their moral judgment.

These results suggest that while simply including ethics material in a course doesn't guarantee that it will have a positive impact on students' moral judgment, there is potential for these courses to have that impact. Unfortunately, most of the studies that make up this body of research do not provide enough detail about the courses they are studying to determine what aspects of the courses led to – or did not lead to – changes in students' moral judgment. A small number of studies, though, attempt to parse how specific aspects of ethics courses affect students. For example, Krawczyk (1997) compared student outcomes for three common ways of teaching ethics in undergraduate nursing programs: a semester-long ethics class taught by an ethicist, integration of ethics into multiple nursing theory classes, and encounters with ethics education only during students' clinical experiences. Students who were taught ethics within a formal ethics class showed higher gains in moral judgment that those taught by the other two methods. Drake, Griffin, Kirkman, and Swann (2005) and Klimek and Wennell (2011) investigated whether students saw differing effects on their moral judgment from a full ethics course compared to learning ethics from a module in a broader class; the first study found no difference while the latter found that students who took the full ethics course had much higher levels of moral judgment than those who simply had an ethics module included in another class.

Unfortunately, the results of the studies by Krawczyk (1997) and Klimek and Wennell (2011) can shed little light on these different methods of teaching ethics because of the designs of the study. In each case, the students had not just taken different styles of ethics courses; they had attended entirely different colleges! In Krawczyk's study, the students were tested at the beginning of their freshman years and the end of their fourth years, with any differences

attributed to the difference in the approach to ethics teaching. Klimek and Wennell, on the other hand, only assessed moral judgment at the end of the students' ethics class, not taking into account the potential differences in moral judgment of students at the two institutions before they took the classes included in the study. So although it is possible that the differences in students' moral judgment could because of the differences in their ethics education, the methodological limitations make it impossible to draw any conclusions about the effectiveness of these different approaches.

Researchers have also studied the effects of a range of types of courses beyond those focusing on moral or ethical issues on students' moral judgment. Using WNS data, Mayhew, Seifert, Pascarella, Nelson Laird, and Blaich (2012) investigate the effect of several types of classes on moral judgment during the first year of college. They grouped together the number of classes students had taken in the humanities, social sciences, education, and business and the classes they had taken in the natural sciences, technology, engineering, mathematics, or health sciences. They found no effect from the number of classes in the first group, but a small negative effect from the second group. Using different samples from the WNS, two other groups of researchers found conflicting results about the effect of the number of classes from traditional liberal arts fields (fine arts, humanities, mathematics, computer science, statistics, natural sciences, and social sciences) a student took in his or her first year. Martin et al. (2011) found no significant effect, but Padgett, Johnson, and Pascarella (2012) found a very small negative relationship between liberal arts classes and moral judgment. Although these studies are the largest to have considered the effects of different types of courses (both in number of students and institutions), the broad categories of classes make it difficult to interpret these results in ways that would suggest which specific types of classes are driving the effects; in fact, there are so

many disparate subjects in the groups of subjects that positive and negative effects of individual subjects are cancelling each other out within the same variable. Other researchers found no effect of taking accounting (Fleming, Romanus, & Lightner, 2009) and counseling (Brendel, Kolbert, & Foster, 2002), but these similarly looked at all classes in the fields simultaneously rather than consider specific classes.

Other studies have taken a more fine-grained approach to the question, examining specific classes or other curricular experiences. Smith, Strand, and Bunting (2002) and Smith and Bunting (1999) found that participation in an semester physical education class focusing on outdoor ropes courses positive effect on students' moral judgment compared to a control group. The authors hypothesized several aspects of the class that may have led to the effect: "… consequences were real rather than contrived, participants agreed to work toward an atmosphere of mutual respect, situations presented themselves through the natural occurrences of the activities that were analogous to real life, and occasionally reflective discussions were held" (Smith & Bunting, p. 73). Similarly, Brown-Liburd and Porco (2011) found positive effects on moral judgment from the real-world experiences in accounting internships.

Rather than look at broad subjects or individual courses, Hurtado, Mayhew, and Engberg (2012) and Bowman (2009) considered the effects of courses that specifically cover diversity-related topics, to conflicting results. Bowman used data on first-year students from the WNS and found that enrolling in one or more courses covering "'diverse cultures and perspectives (e.g., African American Studies, Latino Studies),' 'women's/gender studies,' and 'equality and/or social justice'" (p. 186) had no significant effects on students' moral judgment, with the exception of a small positive effect for students from lower- and middle-income families who took three or more such classes. Using data from a different study, Hurtado, Mayhew, and

Engberg, on the other hand, found positive effects for students who took a course on either women's studies or social diversity.

Although most studies examined curricular experiences did so by examining the effects of the courses which students took on their moral judgment, a small group looked at the effect of a students' major. For example, several studies examined the moral judgment of business majors; one study found no difference between business majors and other majors (Snodgrass & Behling, 1996), but two others found that business majors have lower levels of moral judgment (McNeel, Abou-Zeid, Essenburg, Smith, Danforth, & Weaver, 1996; Elm, Kennedy, & Lawton, 2001). Even among students majoring in business, studies have found higher levels of moral judgment among accounting (Jeffrey, 1993) and marketing (Herington & Weaven, 2007) compared to general business students. Researchers have also found lower levels of moral judgment among students majoring in education (Cummings, Dyas, Maddux, & Kochman, 2001; McNeel et al.; Livingston, Derryberry, King, & Vendetti, 2006) and social work (Kaplan, 2006), and higher levels for psychology students (Livingstone et al.)

The research using college impact frameworks to study the development of moral judgment during college is extensive and at least some studies consider most of the dimensions of Terenzini and Reason's (2005) model. However, even studies that consider the same dimension often produce conflicting results, making it difficult to determine what makes for developmentally effective experiences and institutional conditions. Further, while there is room for additional study in all of the Terenzini and Reason's dimensions, the Organizational Context particularly understudied. Several studies have considered the effects of attending different types of colleges and universities, but the faculty culture has been little studied and no studies have examined the effects of internal structures, policies, and practices and academic and co-curricular

programs, policies, and practices. Part of the power of their model is the inclusion of these organizational factors, and they present an opportunity to better understand how colleges affect the moral development of their students and to provide suggests how institutions can better organize to encourage that development.

It is also important to note that a review of this type is only able to take into account the descriptions of interventions and student experiences as they are described by the studies' authors. It certainly possible that some of the interventions that I classified as being part of college impact studies were grounded in development and included helping students to confront disequilibrium and consider moral issues from more complex and nuanced ways, but it is not possible to know that based on the descriptions of these interventions. A lack of information about the pedagogy employed and the classroom experiences for students is a consistent limitation of the literature, and one that makes it virtually impossible to build and refine theory based on much of the extant literature or to provide educators with practical guidelines for encouraging the moral development of their students.

Developmental Frameworks

In the literature on the impact of college on students' moral judgment, the use of independent variables consistent with a developmental framework is uncommon. In the studies reviewed in Chapter II, these variables primarily assess the ways and the extent to which students experiences confrontations with disequilibrium and nuanced considerations of moral dilemmas; only eight studies include independent variables consistent with a developmental framework, with only four of those (Dotger, 2010; Endicott, Bock, & Narvaez, 2003; Beller & Stoll, 1992; Boss, 1994) using only such variables.

As discussed earlier in this chapter, cognitive developmental theories, including those that apply cognitive concepts to moral development (e.g., Kohlberg, 1976; Rest, 1979a; Rest et al., 1999b), point to the importance of encountering disequilibrium – feeling unsettled about an experience that one can't explain using current assumptions about moral issues – in creating the conditions in which development is possible. One way students can encounter disequilibrium in educational settings is to be exposed to discrepant information or alternative perspectives that helps them see the inadequacies in their ways of making meaning. Several researchers have studied the impact of students' challenging their own ideas or having them challenged by others; the results of these studies consistently supported the potential of this challenge to encourage the development of moral judgment. For example, Beller and Stoll (1992) studied the impact of using the Socratic Method in a class teaching intercollegiate athletes about moral issues in sport, such as gambling, performance-enhancing drugs, eligibility, and rules violations. In the class, students "were encouraged to analyze their own beliefs as well as the beliefs of others" (p. 47) and the instructor used the class "to first have student-athletes establish what they value and believe and second to encourage them to examine those beliefs and values relative to sport, academic, and moral issues" (p. 47). This process of recognizing one's own beliefs and values and then closely examining those ideas created the opportunity to confront disequilibrium when recognizing when those beliefs are inadequate to address the issues. Beller and Stoll randomly assigned students to this class or to one that covered similar material but without the focus on the student-athletes' own values, and they found that students who took the Socratic Method class saw much larger gains in moral judgment. Other researchers found similar positive results from other course-based activities when students were placed in positions to challenge their own ideas

and preconceived notions (Dotger, 2010; Jagger, 2011; Mayhew, Seifert, & Pascarella, 2010; Boss, 1994).

A small number of other studies have employed developmental frameworks to study the development of moral judgment in other ways. Endicott, Bock, and Narvaez (2003) found that moral judgment was significantly correlated with both the breadth and depth of multicultural experiences. Students who had spent time working and living in other cultures, were friends with diverse peers, and expressed a higher degree of commitment to intercultural growth were more likely to use post-conventional moral schema. The authors credit those relationships and deeper experiences with providing students a more nuanced and complex cultural understanding. This, in turn, allowed the students to access more information and compile that information into more complex schemas to use when confronting new experiences, including moral problems. This finding is consistent with findings by Mayhew and Engberg (2010) who found that the lack of opportunity to develop and experience those nuanced and complex cultural understandings – in their case, because of hostile, tense, and hurtful interactions with diversity and a feeling of being silenced in intercultural interactions – was significantly and negatively related to moral judgment. Even when not related to intercultural experiences, a deeper consideration of complex ideas encourages students' development. Mayhew et al. (2012) show that engaging in three types of deep learning in their courses – experiences with higher order learning, reflective learning, and integrative learning – was positively related to increased moral judgment the first year of college.

One other study (Mayhew, Seifert, & Pascarella, 2012) used a third concept from cognitive developmental theories: the progress through periods of transition and consolidation during the developmental process. First presented by Snider and Feldman (1984) and later

applied to moral development (e.g. Derryberry & Thoma, 2005; Rest et al., 1999b; Thoma & Rest, 1999), transition and consolidation describe a person's movement between and through levels of development. According to Derryberry and Thoma, "low stage mixture is suggestive of preference for and influence of a specific stage (i.e., consolidation) while high stage mixture denotes a lack of a preference for and inconsistent usage of stages (i.e., transition)" (p.90). During a period of transition, a person engages in meaning making consistent with two levels of meaning making, and during consolidation, she engages in meaning making consistent with one level. For example, a student may begin college in a period of consolidation, engaging in meaning making consistent with a maintaining norms moral schema. She may then begin using a post-conventional moral schema, while still primarily relying on the maintaining norms schema. In this transition phase, she will rely more and more on the post-conventional schema, until once again entering a consolidation phase and relying on moral reasoning consistent with this post-conventional schema. A respondent's degrees of transition and consolidation are also assessed with the DIT and DIT-2 (Thoma & Rest, 1999).

Mayhew, Seifert, and Pascarella (2012) examined whether students were differently affected by several college experiences depending on whether they were in a phase of transition and consolidation. They found that several variables (e.g., taking diversity-related courses, having influential interactions with peers, and connecting course material with historical events) had positive effects on moral development for students in a consolidation phase; however, no comparable effects were found for students in a transition phase.

The previous sections have examined the research on moral development and college students using the two major frameworks described by Pascarella and Terenzini (1991, 2005): college impact and developmental. The first, college impact frameworks, have been used

extensively to study the experiences and institutional conditions that encourage the development of moral judgment. Although these often produce conflicting results, they demonstrate the potential of campus characteristics and institutional programs to encourage that development. Studies using the developmental framework are much smaller in number. However, these studies consistently support the potential of educational interventions and experiences in which students learn new information at deep, nuanced, and complex levels and have the opportunity to confront disequilibrium by challenging their own ideas to positively affect the development of moral judgment.

Integrated Frameworks

I define integrated frameworks as those that draw on elements of both developmental and college impact frameworks. They are rare in the literature on students' development of moral judgment during college; only seven of the reviewed studies used integrated frameworks and all but two of those use data from the WNS. Most commonly these studies utilized an integrated framework by including both variables that were consistent with a developmental framework and those consistent with a college impact framework. In three of the studies, however, the authors present independent variables (constructed scales) that incorporate elements consistent with both of developmental and college impact frameworks.

Two studies (Mayhew, Seifert, & Pascarella, 2012; Padgett, Johnson & Pascarella, 2012) used a scale based on WNS data called Diversity Experiences, that was made up of subscales consistent with both developmental and college impact factors. The developmental subscale assessed the extent to which students had meaningful interactions with diverse peers, and it comprised items that assessed experiences such as encountering diverse perspectives outside of the classroom; engaging in serious conversations with students who are very different from

oneself in terms of religious beliefs, political opinions, and personal values; and engaging in serious discussions with others students about different lifestyles and customs. These items assess the extent to which students engaged in experiences with diverse others that led them to see new perspectives and encounter disequilibrium with their own held ideas; this approach to diversity is consistent with a developmental framework. By contrast, the other subscale assessed the frequency with which students participated in college-sponsored diversity programs, such as intergroup dialogues and diversity-related workshops; this approach to diversity is consistent with a college impact framework. By combining these two subscales, the larger scale defined students' diversity experiences in a way that considered both their encounters with disequilibrium (*developmental*) and their participation in campus activities (*college impact*). However, neither study found the Diversity Experiences scale to be a significant predictor of moral judgment.

Martin et al. (2011) and Padgett, Johnson, and Pascarella (2012) found a significant positive impact on moral judgment from a scale that assessed the level of academic and challenge and high expectations students experienced in their coursework. Similar to the diversity experiences scale described above, this scale also was constructed using subscales consistent with both developmental and college impact frameworks. The subscale consistent with a developmental framework assessed how often students were challenged in class and exposed to high faculty expectations; this scale included items relating to different ways students could encounter disequilibrium in classes, such as having their ideas challenged by faculty and being asked to argue for or against particular points of view. The subscales consistent with college impact factors assess the extent to which students engage in higher order assignments, meet high expectations and exerted effort in their courses, and integrated ideas and experiences across

course material. Like the diversity experiences scale discussed above, these two studies use the academic challenge and high expectations scale to define the construct in a way that integrates factors consistent with both the developmental and college impact frameworks.

Other researchers have used an integrated framework not by utilizing both in the same variable but rather by concurrently using individual variables consistent with developmental or college impact in their analyses. When employing an integrated framework in concurrent method, studies have found significant effects on moral judgment from variables consistent with a developmental framework (Mayhew & Engberg, 2010) a college impact framework (Hurtado, Mayhew, & Engberg, 2012), and both frameworks (Mayhew, Seifert, & Pascarella, 2012; Mayhew et al., 2012; Mayhew, Seifert, & Pascarella, 2010).

Although Terenzini and Reason (2005) did not develop their model for this purpose, it is well suited for integrating the developmental and college impact frameworks to study student outcomes. Specifically, the model incorporates a wide range of potential independent variables at a range of levels; outcomes are modeled by taking into account individual characteristics and experiences and a wider range of institutional factors that other college impact models incorporate. This allows for developmentally-focused variables to be conceptualized in each of the components of their model. As they are mostly commonly used in student outcomes research, developmental variables describe individual student experiences (such as encountering disequilibrium in a specific class or interaction with another student). Using Terenzini and Reason's (2005) model, each dimension can be examined by applying a development framework. For example, faculty culture can include assessments of how likely faculty are to utilize developmentally effective pedagogical tools, and the Peer Environment can include assessments of how different the values and perspectives are of the student body from the

individual student. By incorporating these types of developmentally-focused variables at each level of the college impact model, the two frameworks can be truly integrated rather than simply superimposed on one another and treating developmental factors as just one more block of variables at the individual level.

The literature presented in this chapter lays the groundwork for this study's investigation of the way that students' experiences in college contribute to the development of moral judgment. In the next chapter, I will present the methodology I will use to address the five research questions presented in Chapter I.

CHAPTER III: METHODOLOGY

The purpose of this study, as discussed in Chapter I, is to examine students' development of moral judgment in college using college impact and developmental frameworks, as well as an integrated framework with components from each. This chapter details the methodology I used to examine the factors of each framework that affect moral judgment, how those effects change when the two frameworks are integrated, and how explanatory power of the integrated framework compares to each individual framework.

In this study, I utilize data collected as part of the Wabash National Student of Liberal Arts Education (WNS), as well as data from Integrated Postsecondary Education Data System (IPEDS). The WNS is a longitudinal, multi-institutional, concurrent mixed methods study; it uses student surveys and personal interviews to examine the institutional practices and conditions that encourage seven outcomes of liberal arts education: effective reasoning and problem-solving, inclination to inquire and lifelong learning, integration of learning, intercultural effectiveness, leadership, moral character, and well-being (King, Brown, Lindsay, & Vanhecke, 2007). I focus on the moral character outcome, drawing on quantitative data collected at participating colleges and universities. WNS data are appropriate for use in this study for multiple reasons: 1) The 49 participating institutions provides a sample large enough to examine not only student-level differences but also institutional differences; 2) the data are longitudinal, allowing for the examination of change in the outcome over several years; 3) the administered survey included the Defining Issues Test-2 (DIT-2) was administered to assess moral judgment;

it can be used to assess development and is more valid and reliable than student self-reports; and 4) the data include an extensive array of variables related to students' experiences, activities, and perceptions while in college, which allow the data to be applied to the dimensions of all three models to be tested in this study. This chapter is organized into five sections outlining the methods used to address the research questions. First, I discuss the data collection methods and sample for the study. Then I outline the variables used in the study, describing the measure of moral judgment used and detailing the independent variables, included in the college impact, developmental, and integrated frameworks. Third, I detail the statistical analysis I use in the study (three-stage hierarchical linear models for each of the three frameworks). Finally, I discuss several limitations of the study.

Samples

The samples of this paper comprise students from 44 colleges and universities. WNS researchers sampled institutions from a group of colleges and universities that applied to participate in the study; selection was based on institutions' vision of liberal arts education and the implemented related practices. The selection also took into account a desire to reflect a variety of institutional characteristics, including institutional type, control, size, geography, and student residence; liberal arts colleges are intentionally over-represented in the sample (Pascarella & Colleagues, 2007) in order to yield a sufficient sample among typically small institutions. The institutional sample (those institutions that began the study in three consecutive years between 2006 and 2008) includes 29 liberal arts colleges, six research universities, and 11 regional universities. Two institutions participated in two years of data collection, and an additional institution participated in all three years, resulting in a total of 48 institutional cohorts in the sample (Table 3.1). (The WNS also included community colleges in its sample; however,

because of this study's consideration of the development of moral judgment during both the first year and first four years of college, I have excluded these institutions from this study.)

Table 3.1. Student and Institution Samples for First and Fourth Years of WNS Data Collection

	Total	2006 Cohort	2007 Cohort	2008 Cohort
Number of Students (aggregated across institutions)				
Beginning of First Year	6893	2043	1262	3588
End of First Year	3277	1435	460	1382
End of Fourth Year[a]	2667	1067	449	1151
Number of Institutions[b]				
Research Universities	6	3	1	3
Liberal Arts Colleges	28	11	4	16
Regional Universities	10	3	4	3
Total Institutions	44	17	9	22

Note. [a]Not all students who completed survey at the end of the first year did so at the end of the fourth year, and vice versa. The fourth-year sample is not a subsample of the first-year sample. [b]Three institutions (one research university and two liberal arts colleges) participated in multiple cohorts. Those institutional cohorts are treated as multiple institutions in this table.

For this study, I use two samples: students who completed the survey at the end of the first year and students who complete the survey at the end of the fourth year. I do this for multiple reasons. First, sample attrition, discussed later in this section, is more significant for the fourth-year sample. Because students leave their institutions between the first and fourth year, some of this sample attrition occurs because students leave college; therefore the first-year sample is more consistent with the students who enter college and the fourth-year sample is more consistent with those who complete. Both of these groups are of interest. Second, both of these time periods are of interest. Researchers and educators recognize the importance of the first year of college in providing a successful transition for students and setting a trajectory for success. Four years, though, allow for institutional characteristics and programs to be experienced by

students, providing a better picture of the ways an institution affects students. Third, in this study, several survey items in the fourth year ask students only about their experiences in the past school year. This means that if students complete the survey in the first and fourth years, there is only data collected for some questions about the first and fourth years of college, with the second and third years an unknowable black box. Using both samples provides one where the data covers the entire experience of the students (the first-year sample) and one where student has more time to be affected by his or her experiences in their college or university (fourth-year sample).

Within each institution, the population for the study comprised all first-time, first-year, traditional-aged students. The WNS research team sampled students from these populations in three different ways. At the largest university in the sample, a random sample of students from only the College of Arts and Sciences were invited to participate; at all other universities, students were sampled at random from the entire first-year cohort. At liberal arts colleges, all eligible students in the first-year class were invited to participate (Seifert, Goodman, King, & Baxter Magolda 2010). Of those, 6,893 completed the DIT-2 (see next section), of those 3,277 (47.5%) completed the follow-up survey at the end of their first year and 2,667 (38.7%) did so at the end of their fourth year. These comprise the two analytic samples for this study.

The size of the samples varies widely among the 44 institutions in the study. At the beginning of the first year, when students completed the baseline assessment of moral judgment which was required for being part of the analytic sample, the median size of the institutional samples was 133.5 students, which ranged in size from 562 to 26. Similarly, sample attrition also varied widely among institutions. The median proportion of the students who completed the baseline survey and then also completed the survey at the end of the first-year was 0.500, ranged

from 0.881 to 0.010. For the fourth-year sample, the median proportion of the original sample was 0.412, ranging from 0.735 to 0.058. All institutional samples at all three waves of data collection are reported in Appendix A.

As with any longitudinal study, sample mortality can bias study results. Both the first-year and fourth-year samples exhibit considerable sample mortality over the course of the data collection. Sample mortality can bias results if participants do not leave the study randomly and can lead to systematic differences in the sample at each point of data collection (Fitzgerald, Gottschalk, & Moffitt, 1998). Table 3.2 shows statistical differences in characteristics of students in first- and fourth-year samples compared to those students who completed the DIT-2 at the beginning of their first year of college. The results of *z*-tests of proportion (see Table 3.2) show that the sample of students who completed the DIT-2 at the beginning of their first years of college is significantly different than those who completed the instrument at their end of their fourth year. The proportion of students who are male, African American, and attend regional universities is lower for the both samples than in the original data collection. Further, *t*-tests show that students who are in the first- and fourth-year samples have high baseline measures of moral judgment and high school academic ability. These tests suggest that sample mortality has resulted in two analytic samples that are different in measurable ways than the original sample for this study.

Table 3.2. *T*-test and *Z*-test of Proportion Results for Demographic Differences between Students Completing the Defining Issues Test-2 at the Beginning of the First Year and End of the Fourth Year for the 2006 and 2007 Cohorts

	Beginning of First Year	End of First Year	End of Fourth Year
Number of Students	6893	3277	2667
		Mean	
N2 Baseline Measure	33.89	35.45***	37.17***
High School Academic Ability	25.60	26.18***	27.13***
		Percent[a]	
Attending Liberal Arts Colleges	54.29	57.55**	63.25***
Attending Research University	22.16	23.01	22.38
Attending Regional Universities	23.45	19.44***	14.36***
Male	40.06	36.77**	37.83*
Female	59.93	63.23**	62.17*
White	76.41	78.67*	80.28***
African American	8.36	6.65**	4.84**
Asian/Pacific Islander	5.58	5.86	6.00
Latino/a	4.91	4.76	4.87
Native American	0.35	0.37	0.07
Race Missing	4.39	3.69	3.94
Domestic Student	94.45	94.84	94.30
International Student	5.53	5.16	5.70

Note. *p<.05; **p<.01; ***p<.001. [a] Due to missing data for some variables, not all percentages sum to 100.

To account for sample attrition, the WNS research team developed an algorithm that weighted respondents based on their sex, race, and score on ACT or other standardized entrance test. These weights were designed to match the analytic sample for each institution with the same-cohort population within that institution (Seifert, et al. 2010). All analyses employ these weights in order to create a sample more similar to the population from which it was drawn;

however, sample weighting cannot account for unobservable non-response bias (Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau, 2004).

Data Collection

This study relies primarily on survey data collected as part of the WNS. These data were collected from students at three points during their first four years of college: at the beginning of their first-year, the end of their first year, and the end of their fourth years. Within the first weeks of their first semester, students completed a one-hour survey that provided pre-college information, such as demographics, high school experiences, and family backgrounds, and attitudes and values. In addition, baseline measures of six of the seven liberal arts outcomes (one outcome, integration of learning, was only assessed qualitatively). The instruments used for each outcome are reported in Table 3.3. Because of the length of the Collegiate Assessment of Academic Proficiency (CAAP; ACT, 1991) and the Defining Issues Test–2 (Rest, et al., 1999), approximately half of the students at each institution completed each of these measures; students were assigned one or the other assessment at random (Pascarella & Colleagues, 2007).

In the spring term of their first year, students completed a second two-hour survey. This survey included a second assessment of the six liberal arts outcomes measured in the first survey, as well as the National Student of Student Engagement (Kuh, 2001) and the WNS Student Experiences Survey (Pascarella & Colleagues, 2007). These surveys included students' reports of their experiences during the year, including classroom activities, interactions with faculty and peers, course-taking patterns, and co-curricular engagement; and perceptions of their institutions' student body, faculty, and general climate and culture. In the fourth year, students completed these instruments again, providing information on the entirety of the first four years of college

and changes in the six assessed liberal arts outcomes over that time. Respondents were paid a $50 incentive for each survey they completed.

Table 3.3. Quantitative Assessment Tools for Liberal Arts Outcomes in the WNS Survey

Outcome	Assessment Instrument
Effective Reasoning and Problem-Solving	Collegiate Assessment of Academic Proficiency (ACT, 1991)
Inclination to Inquire and Lifelong Learning	Need for Cognition Scale (Cacioppo, Petty, Feinstein, & Jarvis, 1996)
Integration of Learning	No Quantitative Assessment was used for this outcome
Intercultural Effectiveness	Miville-Guzman Universality-Diversity Scale (Fuertes, Miville, Mohr, Sedlacek, & Gretchen, 2000
	Openness to Diversity/Challenge (Pascarella, Edison, Nora, Hagedorn, & Terenzini, 1996)
Leadership	Socially Responsible Leadership Scale (Tyree, 1998)
Moral Character	Defining Issues Test-2 (Rest, et al., 1999)
Well-Being	Scales of Psychological Well-Being (Ryff, 1989; Ryff & Keyes, 1995)

In addition to data collected as part of the WNS, this study also employs institutional data collected by IPEDS, which is a data collection and dissemination program administered by the National Center for Education Statistics, a part of the Institute for Education Sciences in the U.S. Department of Education. Each year, all colleges and universities (both public and private) that receive any federal funding are required to report institutional data on topics including institutional costs and financial aid, student retention and graduation rates, enrollment statistics, and aggregated student demographic data. These data are then made publicly available. For this study, I have downloaded data from this Web site for each institution for the year in which the

institutional cohort began its first year of college. This provides a consistent snapshot of institutional characteristics that existed when each cohort began college. For colleges and universities with more than one institutional cohort, the data for the multiple cohorts are averaged.

Dependent Variable

The dependent variable for all analyses in this study is the student's moral judgment. This outcome is measured by the DIT-2 (Rest, Narvaez, Bebeau, & Thoma. 1999b), the most common tool for assessing moral judgment in higher education research (King & Mayhew, 2002; King & Mayhew, 2004; Pascarella & Terenzini, 2005; Rogers, 2002). Students' moral judgment was assessed at each time they were surveyed.

The DIT-2 presents respondents with a series of five moral scenarios that are designed to activate the respondents' moral schemas. For example, in one scenario, students are presented with the case of a sick woman, Mrs. Bennett, who wants to die, and a doctor who must decide whether to give her enough medication to commit suicide. Students are then asked how important certain considerations are when determining how the doctor should handle the dilemma, including whether Mrs. Bennett's family approves of her decision to die, and whether the doctor could be held legally liable for her death. The DIT-2 is an assessment instrument rooted in the theoretical underpinnings of cognitive-structural theory; thus, instead of emphasizing the content of a decision (here, what the doctor should do) in the scoring, it instead examines the cognitive structures respondents use to make meaning of the dilemma, as evaluated through the factors they see as important in making a moral decision. The instrument is used to determine the extent to which individuals use Personal Interest, Maintaining Norms, and Postconventional moral schemas when responding to these scenarios. The designers of this

instrument did not assume that a person uses only one schema for all moral dilemmas he or she encounters; accordingly, respondents may apply different schema to different types of problems. The theoretical underpinnings of this instrument are discussed in detail in Chapter II.

For this study, I use the DIT-2's N2 score, which assesses the extent to which respondents both employ postconventional moral schemas and eschew more simplistic (i.e., personal interests and maintaining norms) ways of evaluating moral dilemmas (Bebeau & Thomas, 2003). The N2 score has internal consistency reliability statistics ranging from .77 to .81 (Rest et al., 1999) and has been found to better represent the moral schema that respondents are utilizing for meaning-making than its predecessor, the P score, includes only the first criterion and was used extensively in research on college students before the introduction of the N2 score (Rest, Thoma, Narvaez, & Bebeau, 1997). When comparing the two measures, the N2 score outperformed the P score on six different criteria for construct validity: sensitivity to educational intervention; age- and education-group differences; change over time in longitudinal studies; and correlation with measures of moral comprehension, prosocial behavior, and civil libertarian attitudes.

Independent Variables

Independent variables for this study include those consistent with college impact, developmental, and integrated frameworks for studying the development of moral judgment. I present the variables for each framework in turn and then present student pre-college characteristics comprise the remaining variables.

College Impact Framework

The independent variables used in the college impact framework represent aspects of the college experience that are supported by both moral judgment theory and the college impact

tradition of outcomes research. These variables represent students' exposure to institutional conditions and practices that may affect their moral judgment by increasing students' encounters with disequilibrium and complex, nuanced consideration of moral issues, as well as institutional characteristics that are often considered in college impact research. The research on the development of moral judgment using a college impact framework that I described in the previous chapter provides support for the potential of a wide range of institutional conditions and practices and formal experiences to encourage that development. In following the traditions of college impact research described by Pascarella and Terenzini (2005), variables described in this section place emphasis on the presence of institutional conditions and practices and students' patterns of participation in and engagement with formal institutional activities and structures. The selection of variables is guided (but not solely determined) by that previous body of research.

The selection of independent variables for the college impact models in this study follows Terenzini and Reason's (2005) Comprehensive Model of Influences on Student Learning and Persistence, which was discussed in detail in Chapter II; each dimension is discussed separately below. Independent variables in this framework comprise four dimensions: organizational context, peer environment, individual student experiences, and student precollege characteristics and experiences. In each dimension, the model-building process begins with the variables listed below; however, some variables may be removed from the analysis to maximize parsimony and statistical power and to minimize multicollinearity. It should be noted that decisions about what to include in the WNS data were not guided by Terenzini and Reason's model, and although there are variables in the WNS student survey that apply to each dimension in the model, these

variables do not address all aspects of each dimension. Table 3.4 provides descriptive data of college impact framework variables.

Table 3.4. Means and Proportions of College Impact Framework Variables, for the First- and Fourth-Year Samples.

	First-Year Sample	Fourth-Year Sample
Service Learning Class	0.431	0.734
Honors Program	0.168	0.302
One Diversity Class	0.229	0.380
More than one diversity class	0.382	0.768
Student Organization Leader	0.244	0.707
Leadership Training Program	0.127	0.346
Religious Congregation	0.318	0.343
Social/Political Lecture	0.684	0.844
Community Service	0.523	0.804
Greek Organization	0.135	0.172
Varsity Athlete	0.161	0.204
Race Workshop	0.464	0.664
Biology	0.120	0.128
Business	0.107	0.084
Education	0.061	0.040
Humanities	0.122	0.075
Physical Sciences	0.061	0.072
Professional	0.105	0.248
Social Sciences	0.187	0.050
Engineering	0.023	0.077
Other Major	0.102	0.128
Undecided[a]	0.047	0.000
College Average Academic Ability[b]	25.815	26.692
Perceived Faculty Interest in Students[b]	0.020	0.001

Note. Only variables in final estimated models are included in this table. [a]No students reported being undecided at the end of the fourth year. [b]Individual variables that are aggregated and used as institutional characteristics are presented as the unweighted average of institutional averages.

Organizational context.

Internal structures, policies, and practices. This construct of Organizational Context is a broad construct that "implies a rich variety of internal organizational structures and processes that can have some influence on students' experiences and, consequently, learning outcomes" (Terenzini & Reason, 2005, p. 8). To operationalize the construct, I rely on structural variables

that are commonly used in college impact research. This dimension include variables that represent the students' institutional type (community college, liberal arts college, comprehensive university, research university), sector (public, religious private, non-religious private), religious affiliation number of students, proportion of minority students, and faculty/student ratio. Colleges and universities participating in the WNS provided researchers much of this data as part of their applications to be chosen as partner institutions. However, to ensure consistency in the reporting of the numbers and to aid in the standardization between institutions in this study, I use data collected from each institution and published by the IPEDS.

Academic and co-curricular programs, policies, and practices. This dimension provides information about both the intended and enacted curriculum and co-curriculum. As discussed in the previous chapter, the existing research on moral judgment in college does not address the impact of this dimension. One reason is likely the difficulty in collecting institutional data on these practices. Terenzini and Reason (2010), for example, suggest that this construct might include such factors as:

> whether an institution has a formal or core set of courses common to all lower-division students; a summer reading program (with subsequent discussion sections in the fall term) for new students; first-year seminars (the content, credit value, and academic standing of the instructor can be important); learning communities; student-faculty research opportunities; service-learning courses; academic and co-curricular programs specifically designed to provide opportunities for new students to encounter and learn about diverse peoples and cultures; an emphasis on (and faculty development support for) active and collaborative pedagogies; and a new student orientation program jointly developed and delivered by academic and student affairs divisions. (p. 14)

As these types of data are available in neither the WNS nor IPEDS data sets, I examine three IPEDS measures that describe the allocation of teaching resources at the institution. Institutions differ based on the way resources are organized for instruction and the relative importance of undergraduate education in the academic focus of the institution. I provide two

measures of the emphasis on undergraduate education at the institution: undergraduate to graduate student ratio and faculty to student teaching assistant ratio. In addition, I use a measure, from IPEDS, of the percent of students who earn degrees in vocationally focused majors. These are, admittedly, very rough proxies for ideal variables in that dimension (which are unavailable for this study). They do, however, represent, in a limited way, the emphasis and resources – both monetary and in faculty time – expended on undergraduate education, and the emphasis of the institution on vocational fields rather than traditional disciplines.

Faculty culture. In Terenzini and Reason's (2005) model, Faculty Culture is defined as "the dominant philosophies of education to which most (or a significant number of) faculty members subscribe, as well as their perceptions of their roles and what it means to be a faculty member at 'this' institution" (p. 10). Most importantly, they say, this refers to the extent to which this faculty culture is student- and learning-centered. Because the WNS includes only data collected from students, it is not possible to examine the way that faculty themselves see their culture. Instead, I rely on student perceptions of faculty as a proxy measure for faculty culture.

I employ two factors to operationalize the faculty culture, which were developed as part of the WNS (Pascarella & Colleagues, 2007). The first is a five-item factor that represents the extent to which the student has had positive out-of-class interactions with faculty. The second is a five-item factor that represents the student's perception of faculty interest in students' learning and development. Taken together, these factors provide information about the faculty emphasis on faculty-student interaction, teaching, and student learning, all of which would be expected to affect student outcomes. Since these are institutional factors, in addition to the students' assessment, I also include a separate institutional mean for each factor as an institutional-level

variable in the analysis, representing an overall reflection of faculty culture on faculty culture. See Table 3.5 for information on both factors.

Table 3.5. Survey Items and Factor Loadings for Faculty Culture Factors in the College Impact Framework

Factors	Loading
Quality of non-classroom interactions with faculty[a] (α=.852)	
The extent R agrees that non-classroom interactions with faculty have had a positive influence on intellectual growth and interest in ideas	.867
The extent respondent agrees that non-classroom interactions with faculty have had a positive influence on personal growth, values, and attitudes	.866
The extent respondent agrees that non-classroom interactions with faculty have had a positive influence on career goals and aspirations	.811
The extent respondent agrees that he or she is satisfied with the opportunities to meet and interact informally with faculty members	.730
The extent respondent agrees that since coming to this institution, he or she has developed a close, personal relationship with at least one faculty member	.724
Faculty Interest in Teaching and Student Development[a] (α=.856)	
Most faculty with whom respondent had contact are genuinely interested in teaching	.828
Most faculty with whom respondent had contact are genuinely interested in students	.818
Most faculty with whom respondent had contact are outstanding teachers	.801
Most faculty with whom respondent had contact are interested in helping students grow in more than just academic areas	.778
Most faculty with whom respondent had contact are willing to spend time outside of class to discuss issues of interest and importance to students	.768

Note. [a]This scale was constructed by Pascarella and Colleagues (2007) as part of the WNS.

Peer environment.

Terenzini and Reason (2005) describe the peer environment as embodying "the system of dominant and normative values, beliefs, attitudes, and expectations that characterize a campus' student body" (p. 12); it affects students by providing " a perhaps semi-conscious grasp of what other students value and of what one's peers expect behaviorally, whether in the student's social or academic world" (p. 12). Instead of limiting the peer environment to student's friends and

limited peer group, Terenzini and Reason widened their conceptualization to refer instead to "a broader, more general, and subtle set of influences" (p. 12). To reflect this, I operationalize the peer environment using this broader lens, referring here to characteristics of the student body as a whole, rather than just the friends and others peers which whom a student is mostly likely to interact on a regular basis.

For this dimension, I include data collected as part of the WNS student survey, as well as data collected and published by IPEDS. Items in this dimension assess selected aspects of the peer environment, again focusing on those that would be expected to create situations in which students encounter disequilibrium or create a culture that emphasizes experiences that encourage or discourage the development of moral judgment. (In the college impact tradition, these variables assess only if the institution had these characteristics, not whether students did encounter disequilibrium because of these characteristcs). Four variables comprise the peer environment factor. These include variables from IPEDS on the make-up of the student's entering class at his or her institution: the percent of students of color and the percent of international students in each institutional cohort. I also include from the WNS survey the percent of students who are members of social fraternities and sororities and who have done or plan to do community service as students. Independent variables in this dimension also include the mean SAT/ACT of students in the institutional cohort (collected via IPEDS) and the mean reported highest intended academic degree of the institutional cohort (collected via WNS survey) as variables to represent the orientation toward academics of the Peer Environment.

Individual student experiences.

Classroom experiences. Classroom experiences refer to the pedagogical practices students' instructors employ, the type of assignments they complete, their interactions with their

classmates, and other experiences that take place in the formal classroom. For this study, these are operationalized with two factors that assess the extent which students are challenged by their coursework. The first is a six-item factor that assesses the extent to which a student reports that his or her classes were challenging and faculty had high expectations. The second is a four-item factor assessing the amount of higher order exams and assignments the student was asked to complete in his or her coursework. See Table 3.6 for information about all three factors.

Table 3.6. Items and Loadings for Classroom Experiences Factors in the College Impact Framework

Factors	Loading
Challenging Classes and High Faculty Expectations[1] *(α=.836)*	
How often faculty challenged respondent's ideas in class	.787
How often faculty asked respondent to point out any fallacies in basic ideas, principles, or points of view presented in the course	.769
How often faculty asked respondent to argue for or against a particular point of view	.767
How often students challenged each other's ideas in class	.726
How often faculty asked respondent to show how a particular course concept could be applied to an actual problem or situation	.722
How often faculty asked challenging questions in class	.676
Frequency of Higher Order Assignments[1] *(α=.780)*	
How often exams or assignments required respondent to point out the strengths and weaknesses of a particular argument or point of view	.845
How often exams or assignments required respondent to argue for or against a particular point of view and defend an argument	.824
How often exams or assignments required respondent to compare or contrast topics or ideas from a course	.811
How often exams or assignments required respondent to write essays	.685
How often exams or assignments required the respondent to use course content to address a problem not presented in the course.	.466

Curricular experiences. Curricular experiences refer to students' course-taking patterns and their other activities that are part of the academic structure of their institution or their own academic departments or programs. A plethora of studies (described in Chapter II) have shown

the potential for a range of types of course to affect students' moral development. For this dimension, I include variables that describe students' enrollment in several types of courses that are often designed with goals that encourage some facet(s) of moral development, providing students a space in which to confront disequilibrium, and a deeper consideration of societal and moral issues. As discussed in Chapter II, extant literature has provided evidence for the potential of many of these curricular experiences to encourage the development of moral judgment. This dimension includes students' reports of the number of courses they have taken that focused on diversity, women's or gender studies, and social justice; and participation in service-learning or an honors college or program.

This dimension also includes students' majors as reported at in the WNS data collection at the end of their fourth year of college. These majors have been coded into ten categories: Arts and Humanities, Biological Sciences, Business, Education, Physical Science, Professional, Social Science, Other, and Undecided.

Out-of-class experiences. This final dimension of Terenzini and Reason's (2005) model represents the activities and experiences in which students engage outside of the formal academic programs of the institution. The central focus of college impact frameworks is to examine the way institutionally-sponsored programs and contexts affect student outcomes, so this dimension includes only activities, programs, or organizations sponsored and/or organized by the institution and its actors. These variables include students' participation in activities that have been shown the studies reviewed in the previous chapter demonstrated to have potential to encourage (or discourage) students' development of moral judgment or other activities with the potential to put students in a position to encounter moral issues. Variables for this dimension are: holding a leadership position in a student organization; participating in a leadership training

program; belonging to a religious congregation; attending a debate or lecture on a current political social issue; engaging in community or volunteer work; living in an on-campus residence hall; belonging to a social fraternity or sorority; and participating as a varsity student-athlete; and attending a racial/cultural awareness workshop.

Terenzini and Reason (2005) designed their Comprehensive Model of Influences on Student Learning and Persistence to provide a framework study the impact on student outcomes of a broad range of characteristics of institutional culture and student characteristics within that culture. The set of independent variables selected for this study follow the organization of this framework, and apply them to the examination of factors affecting moral judgment development.

Developmental Framework

For the developmental framework, I have chosen variables that cognitive developmental theory and research (described in Chapter II) suggest would encourage or inhibit the development of moral judgment. The review of existing literature also in Chapter II demonstrates that studies that employ a developmental framework do so by examining the effects of students' encountering disequilibrium or engaging with social and moral issues in deep, nuanced ways. In keeping with the tradition of a developmental framework as one that focuses the processes that lead to "key features of developmental changes in the ways people make meaning of their experiences over time in increasingly complex and inclusive ways" (King & Baxter Magolda, 2010, p. 214), variables in this framework represent the cognitive activities in which students engage rather than the institutional experiences in which they engage. For example, the college impact framework described above includes variables that indicate whether a student has taken certain kinds of classes where they might encounter the kinds of processes that King and Baxter Magoda describes (such as classes focused on professional ethics or diversity issues), but the

developmental framework includes variables that indicate the extent to which students have engaged in those processes across courses and other contexts while in college. Table 3.7 presents descriptive statistics for the variables of developmental framework.

Table 3.7. Means and Proportions of College Impact Framework Variables, for the First- and Fourth-Year Samples

	First-Year Sample	Fourth-Year Sample
Classroom encounters with disequilibrium	0.060	0.018
Self-initiated encounters with disequilibrium	0.048	0.039
Meaningful conversations with diverse others	0.034	-0.009
Negative interactions with diverse others	-0.123	-0.143
Experiences with Higher Order Learning	0.030	-0.002
Experiences with Integrative Learning	0.022	0.011

Note. Only variables in final estimated models are included in this table.

To operationalize students' encounters with disequilibrium during their time in college, I use three factors constructed from WNS student survey data. The first is a three-item factor that represents students' classroom encounters with disequilibrium. The second is a three-item factor that represents the frequency with which students challenged themselves to encounter disequilibrium. The third and fourth represent students' interactions with diverse others; the third is an eight-item factor that represents students' frequency of engaging in meaningful conversations about challenging topics and with diverse others. This factor differs from measures of structural diversity and those of proximity or contact with diverse others (consistent with a college impact framework) by moving to substantive interactions, focusing on depth of experience and actual discussion across difference, especially around topics with more potential for dissonance. If this factors results in the overestimation of students' meaningful encounters

with diversity, however, any statistically significant effects of this variable on the outcome would be an underestimate of the true effect.

Fourth is an four-item factor that represents students negative interactions with diverse others Table 3.8 includes further information about all three factors. Together, these three factors represent students' encounters with disequilibrium in their classes, in their interactions with others, and by their own initiative.

To operationalize students' deep and nuanced considerations of experiences and ideas, I use three factors originally presented by Nelson Laird, Shoup, Kuh, and Schwarz (2008) that together represent students' engagement in deep learning while in college. These are a four-item factor representing students' engagement in higher order learning, a five-item factor representing experiences with integrative learning, and a three-item factor representing reflective learning. Mayhew, et al. (2012) examined the effects of deep learning on moral judgment during the first year of college, and they found different directions and significance of the effects of the three factors; because of this, I do not combine the three into one larger factor combining these three types of deep learning. A fourth factor (one that is not part of the deep learning scale), is the extent to which students have integrated information, ideas, and experiences (Pascarella & Colleagues, 2007). Table 3.9 provides more information about all four factors.

Table 3.8. Items and Factor Loadings for Encounters with Disequilibrium Factors in the Developmental Framework

Factor	Loading
Classroom encounters with disequilibrium (α=.706)	
How often faculty challenged respondent's ideas in class	.833
How often students challenged each other's ideas in class	.823
How often exams or assignments required respondent to argue for or against a particular point of view and defend an argument	.727
Self-initiated encounters with disequilibrium (α=.777)	
How often respondent tried to better understand someone else	.862
How often respondent examined strengths and weaknesses of own views on a topic or issue	.831
How often respondent learned something that changed the way he or she understands an issue or concept	.803
Meaningful conversations with diverse others (α=.885)	
How often respondent had serious discussions with other students about major social issues such as racial diversity, human rights, equality, or justice	.793
How often respondent had serious discussions with other students about different lifestyles and customs	.786
During current school year, how often has respondent had serious conversations with students of a different race or ethnicity than his or her own	.724
How often respondent had meaningful and honest discussions about issues related to social justice with diverse students while attending this college	.711
During current school year, how often has respondent had serious conversations with students who are very different from him or her in terms of their religious beliefs, political opinions, or personal values	.689
How often respondent made friends with a student whose race was different than own	.661
How often respondent had serious discussions with faculty whose political, social, or religious opinions were different from his or her own	.623
How often respondent encounters diverse perspectives outside the classroom	.610
Negative interactions with diverse others[a] *(α=.686*)	
How often R had tense, somewhat hostile interactions with diverse students while attending this college	.804
How often R had guarded, cautious interactions with diverse students while attending this college	.773
How often R had hurtful, unresolved interactions with diverse students while attending this college	.453
How often R felt silenced by prejudice and discrimination from sharing personal experiences with diverse students while attending this college	.816

Note. [a] This factor is presented and applied to moral judgment by Mayhew and Engberg (2010).

Table 3.9. Items and Factor Loadings for Engagement in Deep Learning Factors in the Developmental Framework

Factor	Loading
*Higher Order Learning (*α=.697)[a]	
Time respondent spent analyzing basic elements of an idea, experience, or theory, such as examining a particular case or situation in depth and considering its components	.734
Time respondent spent synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships	.762
Time respondent spent making judgments about value of information, arguments, or methods, such as examining how others gather or interpret data and assessing soundness of conclusions	.747
Time respondent spent applying theories or concepts to practical problems or in new situations	.658
*Reflective Learning (*α=.771) [a]	
How often respondent examined strengths and weaknesses of own views on a topic or issue	.831
How often respondent tried to better understand someone else's views by imagining how issue looks from his/her perspective	.867
How often respondent learned something that changed the way R understands an issue or concept	.787
*Experiences with Integrative Learning (*α=.697) [a]	
During current school year, how often has respondent worked on a paper or project that required integrating ideas or information from various sources	.669
During current school year, how often has respondent included diverse perspectives (different races, religions, genders, political beliefs, etc.) in class discussions or writing assignments	.682
During current school year, how often has respondent put together ideas or concepts from different courses when completing assignments or during class discussions	.704
During current school year, how often has respondent discussed ideas from readings or classes with faculty members outside of class	.657
During current school year, how often has respondent discussed ideas from readings or classes with others outside of class (students, family members, co-workers, etc.)	.660

Note. [a] These factors are constructed and presented by Nelson Laird, Shoup, Kuh, & Schwarz (2008).

Integrated Framework

In this study, I move beyond the traditional developmental and college impact frameworks to also investigate the development of moral judgment during college using an integrated framework, which combines the two. For the first step of creating the integrated framework, I include all variables from the final development and college impact models (with variables removed if needed to increase power and decrease multicollinearity). Using variables that are parts of each of the developmental and college impact frameworks in the same estimated hierarchical linear models (see below for more information about the modeling process) is one method of integrating the two frameworks, with variables from each being present in one model. As described in Chapter II, this method was the most common way of using an integrated framework in the previous literature on moral judgment in college.

An additional way to integrate the developmental and college impact frameworks is to use independent variables that combine aspects of the two frameworks into one variable. The college impact framework focuses on institutional conditions and student participation in formal institutional programs and experiences; the developmental framework focuses on the students' engagement with cognitive tasks that are theoretically supportive of development. Thus, a researcher can include in analysis variables that combine these two foci, asking students' how they have engaged with these cognitive tasks because of specific institutional conditions or formal programs. I do this by using two factors that assess the extent to which students believe their institutions provided opportunities for them to engage in two types of these developmentally-supportive cognitive tasks. The first is a four-item factor that represents students' belief that the institution supports social interaction and activity, especially among diverse others. The second is a five-item factor that represents the extent to which students

believe the institution has contributed to their personal, moral, and spiritual development. Table 3.10 provides more information about both factors. Table 3.11 provides the descriptive statistics for each variable exclusive to the integrative framework.

Table 3.10. Items and Factor Loadings for Factors in the Integrated Framework

Factor	Loading
Institution supports social interaction with diverse others (α=.718)	
Extent to which respondent's institution emphasizes encouraging contact among students from different economic, social, and racial or ethnic backgrounds	.773
Extent to which respondent's institution emphasizes providing support he or she needs to thrive socially	.756
Extent to which respondent's institution contributes to knowledge, skills and personal development in terms of understanding people of other racial and ethnic backgrounds	.716
Extent to which respondent's institution emphasizes attending campus events and activities	.665
Institution contributes to personal, moral, spiritual development (α=.809)	
Extent to which respondent's institution contributes to knowledge, skills and personal development in terms of developing a personal code of values and ethics	.834
Extent to which respondent's institution contributes to knowledge, skills and personal development in terms of contributing to the welfare of your community	.763
Extent to which respondent's institution contributes to knowledge, skills and personal development in terms of solving complex real-world problems	.748
Extent to which respondent's institution contributes to knowledge, skills and personal development in terms of understanding yourself	.747
Extent to which respondent's institution contributes to knowledge, skills and personal development in terms of developing a deepened sense of spirituality	.658

Table 3.11. Means of Integrated Framework Variables, for the First- and Fourth-Year Samples

	First-Year Sample	Fourth-Year Sample
Institution supports social interaction	0.004	-0.029
Institution contributes to my development	-0.027	-0.026

Note. Only variables in final estimated models are included in this table.

Student Precollege Characteristics

The focus of this study is the examination of the effect of the college experience on moral judgment. In addition to the sets of independent variables used for the developmental, college impact, and integrated frameworks, I also include several variables to control for students' pre-college differences. The control variables include characteristics that previous literature has shown to affect the moral judgment of college students as well as socio-demographic traits often controlled for in college impact research: gender, race, U.S. citizenship status, age, and self-reported political orientation. Although citizenship status, age, and political orientation can change during a student's time in college – as might the way students identify and report their gender and race – using the pre-college reports of these variables follows the tradition of college impact research by acknowledging that the factors that lead to these changes may also contribute to students' changes in moral judgment over the same periods of time. Table 3.12 presents descriptive data for the precollege characteristics.

A measure of academic ability also is included here, operationalized by using students' institution-reported scores on the ACT, SAT, or COMPASS college entrance exam. When institution-reported measures were not available, students' self-reported scores are used. Since the scores come from three different instruments, they have been transformed to a common scale (Pascarella & Colleagues, 2007).

Table 3.12. Means and Proportions Precollege Characteristics, for the First- and Fourth-Year Samples

	First-Year Sample	Fourth-Year Sample
DIT-2 N2 Baseline	35.452	37.170
Entering Academic Ability	26.183	
Female	0.632	0.622
Male	0.368	0.378
African American	0.069	0.050
Native American	0.004	0.001
Asian/Pacific Islander	0.061	0.062
Latino/a	0.049	0.051
White	0.780	0.796
Politically Conservative	0.218	0.196
Neither Conservative nor Liberal	0.370	0.364
Politically Liberal	0.412	0.439

Note. Only variables in final estimated models are included in this table.

I have also included a baseline measure of the dependent variable (the N2 score on the DIT-2) assessed as part of the first survey. In order to attribute the differences in moral judgment at the end of the fourth year of college to experiences during college, it is necessary to control for students' pre-college differences in the outcome.

Missing Data

Missing survey data limit the validity of responses if these data are not missing at random (Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau, 2004), and data are missing for both samples in this study. Table 3.13 presents the number and proportion of students who have missing values for each of the independent variables in the final models for both the first-year and fourth-year samples. Across both samples, the highest percentage of missing of data for any variable is whether students in the fourth-year sample participated in a service learning class, with 15.3% of students not providing a response. No other variable in either data set is missing for more than 10% of students. In fact, only four other variables have missing data for more than 4% in their respective sample: meaningful conversations with diverse others (7.2% in the first-

year sample and 4.2 in the fourth-year sample); negative interactions with diverse others (5.1% in the first-year sample); and major (6.4% in the first-year sample and 8.5% in the fourth-year sample). Because samples were identified based on students having valid data for both the baseline DIT2 N2 score and the application outcome N2 score, these variables have no missing data in either sample. Similarly, there are no missing data for students' gender or precollege academic ability, as these data originated from administrative records provided each college.

In all model estimation in this study's analyses, I used multiple imputation to account for missing data. The assumption underlying this procedure is that data are missing not at random, meaning that accurate values for the missing values cannot be calculated based on those cases with complete data with complete certainty. Multiple imputation deals with this problem by estimating multiple values for each missing value rather than just one. This creates multiple data sets; parallel analyses are conducted on each imputed dataset and then these results are combined for a final set of imputed results. In the final combination of the five sets of results, parameter estimates are averaged together, and standard errors are presented so that they account for both the expected variance within the sample (within-imputation variance) and variation across the imputed values for the variable in each dataset (between-imputation variance). The addition of the between-imputation variance to the within-imputation variance for the regression coefficient leads to standard errors for coefficients estimated using multiple imputation that are larger (and, thus, *t* values that are smaller) than analyses with non-imputed data. This represents the uncertainty that missing data cause (Rubin, 1987; Schafer & Graham, 2002; Wayman, 2003; Yuan, 2000).

Table 3.13. Number and Percent of Students with Missing Data on Variables in Final Analysis, in the First-Year and Fourth-Year Samples

	First-Year Sample		Fourth-Year Sample	
Variable	Number Missing	Proportion Missing	Number Missing	Proportion Missing
Precollege Characteristics				
Race	121	0.037	105	0.039
Developmental Framework				
Classroom Encounters with Disequilibrium	116	0.035	22	0.008
Self-Initiated Encounters with Disequilibrium	83	0.025	46	0.017
Meaningful conversations with diverse others	235	0.072	112	0.042
Negative interactions with diverse others	166	0.051	42	0.016
Experiences with Higher Order Learning	83	0.025	40	0.015
Experiences with Integrative Learning	86	0.026	40	0.015
College Impact Framework				
Service-Learning Class	79	0.024	407	0.153
Honors Program	99	0.030	14	0.005
One Diversity Class	95	0.029	6	0.002
More Than One Diversity Class	95	0.029	6	0.002
Student Organization Leader	93	0.028	15	0.006
Leadership Training Program	109	0.033	28	0.010
Religious Congregation	101	0.031	26	0.010
Community Service	90	0.027	60	0.022
Greek Organization	80	0.024	43	0.016
Varsity Athlete	84	0.026	24	0.009
Race Workshop	94	0.029	11	0.004
Major	211	0.064	227	0.085
Nonclassroom Interactions with Faculty	108	0.033	13	0.005
Faculty Interest in Students	106	0.032	13	0.005
Integrated Framework				
Institution supports social interaction	107	0.033	59	0.022
Institution contributes to my development	106	0.032	53	0.020

Note. Only variables with missing data are included in this table.

To conduct multiple imputation in these analyses, I relied on the explanation and examples for the SAS PROC MI procedure presented by Yuan (2000). All analyses were conducted with five imputed datasets.

Data Analysis

In order to address the five research questions for this study stated in Chapter I, I use hierarchical linear modeling (HLM) to conduct analyses using each of the three frameworks (college impact, developmental, and integrated) to examine the effects of variables described above on moral judgment. I then compare the model's pseudo-R^2 statistics and confidence intervals around the models' coefficients to compare the estimated effects and explanatory power of the three frameworks.

To examine four-year changes in students' moral judgment (measured by the N2 score of the DIT-2), I use a three-stage HLM process utilizing each of the three frameworks. HLM is an appropriate method both because of its statistical properties and its potential to address the influence of the institutional context on student outcomes. First, students surveyed in this study are clustered within the colleges and universities they attend. This means that the error between students' estimated and observed outcomes is correlated among students who attend the same institution, which is a violation of the assumptions of linear regression. Violating this assumption can lead to a bias in regression coefficient standard errors, increasing the likelihood of committing Type 1 Error and inappropriately rejecting the null hypothesis for individual coefficients. Second, I undertake this study with the assumption that colleges and universities provide different environments in which students encounter these types of experiences (as measured within either a developmental or college impact framework), and that students will do so with different frequencies and consistency and with differing effects based on the institution

they attend. Third, the college impact and integrated frameworks both incorporate institutional-level variables with the same values for all students attending the same institution (such as an institution's graduation rate or proportion of students who participate in community service). The conceptual models suggest that the effects of these institutional-level variables can affect moral judgment either directly (affecting the predicted value of the outcome itself) or indirectly (affecting the slope of individual-level effects on the predicted outcome). As a statistical method, HLM accounts for the clustered nature of the data without biasing standard errors, and it allows for the estimation of the between-institution variance in both the outcome and the individual-level effects (Raudenbush & Bryk, 2002).

In all HLM analyses for this study, independent variables are centered at the grand mean and further standardized so that they have a standard deviation of one. This means that the variables are centered on the mean value for all students in the sample rather than the mean for the school they attend.

Unconditional Models

First, I use a wholly unspecified model to estimate the Interclass Correlation Coefficient (ICC) to determine whether the data provide evidence of institutional differences in students' moral judgment at the end of the fourth year of college. Hierarchical linear modeling is an appropriate statistical technique for analyzing these data. Using unspecified models with no independent variables allows me to estimate the amount of variance in the dependent variables that is accounted for in individual and group differences, partitioning the variances into their within-school (individual differences) and between-institution (institutional differences) components. Based on those estimates, I then calculate the ICC to determine the total proportion of variance in the independent variable that occurs between institutions (in other words, the

proportion of the variance that is attributable to observable and unobservable institutional characteristics). Researchers are in disagreement as to how much variance should occur at the group level to support the use of HLM for the data; suggested thresholds for this value of the ICC include .25 (Heinrich & Lynn, 2001; Guo, 2005), .10 (Lee, 2000; Sarin & McDermott, 2003), and .05 (Porter, 2005; Raudenbush & Bryk, 2002). Others argue that as long as the ICC is greater than 0, HLM should be used to limit the underestimation of standard errors that would be present in an ordinary least squares model (McCoach & Adelson, 2010; Roberts, 2007). Since the third wave of data (senior year) for the third cohort of students are not yet available, it is impossible to conduct this analysis at this time to provide support for the design of this study. However, other WNS data inform this decision: using the first wave of WNS data, Mayhew (2012) found that more than 18% of the variance DIT-2 N2 score occurred between institutions, well above all but the most conservative thresholds noted above. Because the ICC estimation includes no independent variables, it will be the same for each of the three frameworks used in the next steps of analysis.

Within-Institution Models

Second, I construct within-institution models to estimate changes in moral judgment using each of the three frameworks. These models estimate the effects of individual-level variables on four-year changes in moral judgment. Although within-institutions models estimate the effects of individual-level variables, these models vary from ordinary least squares regression techniques in that they restricts the estimation to the effect in independent variables on just the proportion of the dependent variable that varies within institutions while not considering the between-institution proportion of the variance. Because of the nested nature of the data, these estimates also include adjustments to the standard error of each coefficient to correct for the

downward bias of that error that is a concern with other estimation techniques (Raudenbush & Bryk, 2002).

The within-institution models also allow me to determine whether the effect of any of the independent variables on the dependent variables varies between institutions, in other words, whether the size of the effect of student characteristics or experiences on moral judgment differs depending on the institution students attend. Because of the limited number of institutions in the sample, parsimony in the model is vital to ensure an appropriate amount of statistical power for the model estimations. In order to maximize parsimony and limit the number of estimated parameters, I construct the models using an iterative process. I begin by estimating random effects for each independent variable; however, if those random effects do not estimate statistically significant differences in the coefficient, I substitute fixed effects for those independent variables in the final analytic models.

Between-Institution Models

Third, after estimating the within-institution models for each of the three frameworks, I estimate between-institution models that incorporate institutional characteristics for the college impact and integrated frameworks (the developmental framework includes no institutional-level variables). These between-institution models build on the results of the within-institution models in two ways. First, they allow for the estimation of direct effects of the institutional-level variables on students' moral judgment; due to the partitioning of the variance into within-institution and between-institution models, these effects account for between institution variance only. Second, they allow for of slopes-as-outcomes estimation for the within-school (individual) independent variables. For independent variable coefficients that are estimated to vary between institutions in the within-institution model, the between-institution model allow for estimation of

the effect of institutional variables on that coefficient, illustrating how institutional-level variables can increase or decrease the effect of individual-level variables in both the college impact and integrated frameworks. As in the between-institution models, I use an iterative process to develop the models in order to maximize parsimony and statistical power at both the individual and institutional levels. This is especially important in the between-institution models, because estimated parameters include the estimate of the effect of each institutional-level independent variable on each individual-level slope in the model, which can yield a high number of parameters when compared to models that do not estimate these effects.

Differences among Frameworks

For the fourth step in the analysis, I compare the three models using the developmental, college impact, and integrated frameworks; this addresses the study's fourth and fifth research questions, which ask how the explanatory power of the models and estimated sizes of the effects of the independent variables change when the developmental and college impact frameworks are integrated. I do so by comparing the models' pseudo- R^2 values and the confidence intervals around estimated effect.

It is impossible to calculate a true R^2 statistic to determine the total variance explained in an HLM model because R^2 formulas are based on ordinary least squares regression with only one level of analysis. Because HLM estimates variance at two levels simultaneously, an R^2 calculation that does not account for the multi-level nature of the model leads to incorrect and sometimes impossible estimates of the explained variance (Snijers & Bosker, 1994; Kreft & De Leeuw, 1999; Luke 2004). The most common method of calculating pseudo-R^2 values in multi-level models is to calculate separate explained variance estimates for each level in the model separately (Gelman & Pardoe, 2006). Multiple methods have been suggested for calculating the

pseudo-R^2, and I use the method proposed by Kreft and De Leeuw (1999) and Singer (1998) and endorsed by HLM creators Raudenbush and Bryk (2002). Using this method, I estimate variance explained at each level by dividing the difference in the error in the unconditional model and the error in the final restricted model by the unconditional model error (See Figure 3.1). The student-level statistic (within-institution) is an estimate of how well the independent variables explain variance in moral judgment; the institution-level (between-institution) statistic be an estimate of the amount of between-institution variance in moral judgment explained by the model. I will then compare these two statistics for models estimated using the three frameworks to explain how explained variance at each level differs.

$$\frac{Unconditional\ Error - Restricted\ Error}{Unconditional\ Error}$$

Figure 3.1. Equation for calculating within-school and between-school pseudo-R^2.

Since the integrated framework includes estimates of the effects of independent variables used in both the developmental and college impact frameworks, the inclusion of new variables in the model could change the estimated effects of each variable by controlling for a different set of variables in the estimation. To understand the difference in estimated effects across models, I examine the estimated coefficients and the confidence intervals around those coefficients for the same variables across different frameworks and models.

Taken together, these analytic steps allow me to investigate the development of moral judgment during college using developmental, college impact, and integrated frameworks. In summary, I propose to first estimate separate models using the development and college impact frameworks. Next, I integrate the two frameworks, including variables from each, as well as new

variables that integrate the two frameworks. This integrated model allows me to provide the most complete picture of what encourages development of moral judgment; by comparing the integrated model with the two previous models, I examine how integrating them changes the estimates of effects and the explanatory power of the model.

Limitations

This study examines the effects of students' college experiences on moral judgment using developmental, college impact, and integrated frameworks, which has the advantage of contrasting the ways of conceptualizing predictors of moral judgment. However, like all studies, this design comes with several limitations.

Although the WNS Data are appropriate for the study for many reasons outlined earlier in this chapter, they were not collected for the purpose of this study and therefore do not include a full complement of measures that align with each of the major elements of the three frameworks. In the developmental framework, the factors created and utilized in the analysis address some of the ways that students may engage in the cognitive tasks that encourage development, but there are many other ways this can happen that are not included in the study.

In the college impact framework, although data related to every dimension of Terenzini and Reason's (2005) Comprehensive Model of Influences on Student Learning and Persistence were available through the WNS or IPEDS, some dimensions could be addressed more completely with additional data collection. For example, in this study, I rely solely on data collected from students, making it impossible to measure Faculty Culture from the perspective of the institution's faculty themselves. Further, questions posed to students about faculty are directed toward the institution's faculty as a whole rather than faculty with whom they worked most closely. For example, students were not given the opportunity to acknowledge the faculty

culture within their home academic department. Using this approach, an engineering student might interact with a completely different set of faculty than an elementary education student at the same university. Although those faculty may have presented very different types of culture to the two students, the WNS data can draw no distinction between those faculty groups. Further, students are unlikely to have knowledge of some other dimensions of Terenzini and Reason's model, such as academic priorities or institutional policies. Terezini and Reason offer examples the kinds of data that may be important to fully assessing these organizational context dimensions of their model; these include: the presence of administrators, faculty, or staff with explicit responsibilities for the first-year experiences; the presence of extensive collaboration across academic and student affairs units on campus; and different types of financial aid policies. One can imagine a plethora of similar organizational and structural factors that could influence students' development of moral judgment during college. To collect these data for an institution, however, would require extensive additional efforts at each institution that was focused specifically on these institutional context dimensions. Such data collection was beyond the scope of the WNS, is not a part of the IPEDS federal mandate, and the collection of this type of supplemental data is beyond the scope of this dissertation.

In the integrated framework, I created and estimated the effects of factors that represent students' belief that the institution supports social interaction with diverse others and contributes to their personal, moral, and spiritual development. A survey designed to provide data for an integrated framework could ask students for their perspective on a wider range of ways that institutions provide the opportunities for the cognitive activities that support development through institutional conditions and formal programs and experiences that would be appropriate for an integrated framework. Additionally, a fully integrated framework would also include data

that address whether and how students engage in specific types of cognitive tasks – such as encountering disequilibrium and thinking about social and moral issues in complex ways – as part of specific formal college experiences. For example, did students re-evaluate their understanding of poverty in a service-learning program? Did their fraternities or sororities conduct social events that made them reconsider the morality of facilitating binge-drinking or serving alcohol to minors? This type of data would require the systematic collection of data for this purpose and informed by an integrated framework, which was not part of the conceptual model underlying the quantitative portion of the WNS study nor the type of information collected by IPEDS.

Beyond limitations relating to the variables in the study, there is also a concern of error in the survey data due to the length of the survey. As respondents complete long or cognitively taxing surveys, the quality of the data they provide can deteriorate over the course of the survey. For example, Galesic and Bosnjak (2009) found that for longer surveys, respondents responding to items positioned in the later in the survey took less time to complete each item and provided answers that were more uniform than they did for items earlier in the survey. This can lead to respondents satisficing in their responses and more frequently choosing neutral responses (O'Muircheartaigh, Krosnick and Helic, 2000) or extreme responses (Holbrook, Cho, and Johnson, 2006). These effects can be more pronounced for items that take more thought to complete (Subar et al., 2000). The WNS surveys were very long, taking an hour or two hours to complete, depending on the survey. In addition, they were also cognitive taxing surveys, asking students to recall a wide range of experiences and assessing five disparate outcomes. The DIT-2 is a long and taxing instrument, including more than 100 items and requiring the reading of five

vignettes and detailed instructions. Survey length and the effort required to complete it may threaten the validity of the data collected by the WNS in general, and the DIT-2 in particular.

Although the sample of institutions for is among the largest used in quantitative studies of the development of moral judgment during college, it still comprises a small number of institutions for estimating between-institution effects. More statistical power would allow for the estimation of more between-institution parameters, which form an important focus of the Terenzini and Reason (2005) model. This would be especially limiting in a study that included the type of institutional data outlined earlier in this section.

Finally, the institutions in this study were not chosen via random sampling; instead colleges and universities were chosen based in part on their institutional commitment to liberal arts outcomes and their experience attempting to promote their achievement. If students attend institutions that are making concerted efforts to educate for these outcomes, including the development of moral character, they will likely have experiences that are the results of those efforts. In that case, the effects of the college experience on their moral judgment may be different than for students who do not attend institutions that place the same focus on these liberal arts outcomes; this would limit the generalizability to other colleges and universities, even those that otherwise seem similar.

These limitations are important to consider for this study, but they do not undermine the value of this study and its potential to contribute to the research on the development of moral judgment in college. As an exploratory study, the data examined here can establish the value of examining the development of moral judgment with developmental, college impact, and integrated frameworks. Although no one study be able to answer every question about that development, it provides a broader picture of how students' experiences in college affect their

moral judgment than has been done in previous research, and it provides direction for future researchers in choosing new ways of using these three frameworks.

CHAPTER IV: RESULTS

In this chapter, I present results from all analyses used to address the five research questions as described in Chapter III. First, I address whether there are significant changes in students' moral judgment during the first year and the first four years of college and whether that moral judgment varies by which college a student attends. I do this by providing the results of a series of *t*-tests investigating differences in moral judgment over time and unspecified hierarchical linear models investigating differences between institutions. Second, I present the results of the hierarchical linear models estimated using both developmental and college impact frameworks, including outlining the model specification process. After presenting these two separate models, I then present the results of the hierarchical linear model estimated using an integrated framework, which will integrate the developmental and college impact models. Third, using these estimated models, I present differences in the pseudo-r^2 values of each model, which will provide information about the explanatory power of each model. Finally, I present the changes in the standardized effects of independent variables across models using the three frameworks, examining the potential extent of omitted variable bias in the models using only the developmental or college impact frameworks. Taken together, these results will provide evidence of the nature of the change in students' moral judgment over time and between colleges and the relative efficacy of three different frameworks in understanding that change. In Chapter V, I will discuss the implications of these results, examining the efficacy of an integrated framework in examining changes in moral judgment during college.

Changes in Moral Judgment during College

Before beginning analyses that address the research questions, I first determined whether students' levels of moral judgment changed while they were in college. I conducted dependent sample *t*-tests to determine whether the differences between pre-test and post-test DIT-2 N2 score averages are statistically significant. As with the descriptive statistics in the previous section, I draw on two different samples for these *t*-tests: the first-year sample, with the DIT-2 N2 post-test measure of moral judgment assessed at the end of students' first year of college, and the fourth-year sample, with the DIT-2 N2 post-test assessed at the end of students' fourth year of college. Only students with a valid N2 score for both the baseline assessment at the designated outcome assessment (at the end of the first or fourth year of college) are included in the samples; some students are in both samples, but others are in only the first- or fourth-year samples. These different samples result in two samples having different baseline averages.

Table. 4.1. Results of Dependent *t*-Tests Between Average Baseline and Outcome N2 Scores

Sample	Mean Baseline N2 Score	Mean N2 Score at End of First Year	Mean N2 score at End of Fourth Year	Difference	*t*-Statistic of Difference
First-Year (*n*=3277)	32.580 (sd=15.444)	36.928 (sd=15.998)		4.348***	20.96
Fourth-Year (*n*=2667)	35.119 (sd=15.407)		43.470 (sd=15.168)	8.351***	28.97

Note. *** $p<.001$

Table 4.1 shows the results of those *t*-tests. Both the differences between the beginning and end of the first year (first-year sample) and the differences between the beginning of the first year and the end of the fourth year (fourth-year sample) are statistically significant ($p<.001$). The mean increase in N2 score over the first year of college (4.348) is more than one-quarter of a

standard deviation of the average first-year change; the mean increase in four years (8.351) is just over one-half of a standard deviation of the average four-year change for students in the sample. These results indicate that students do show development in moral judgment as they persist through college.

Although these results indicate that students do develop in their assessed moral judgment during both the first year and the first four years of college, these results say nothing about the way individual students changed. This study assumes that students develop moral judgment during college at different rates, and that the extent to which they develop is influenced (at least, in part) on their college experiences. Results presented later in this chapter explore the extent to which those college experiences affect students' development of moral judgment. First, I examine whether students develop in different amounts during college. If all students simply experience a similar change in their N2 score, then students would rank in a similar order at the baseline assessment as they do at the outcome assessment. Alternatively, if students experiences different amounts of development during their time in college, then these students would be expected to rank them differently at the outcome assessment than they did at the baseline assessment.

In Tables 4.2 and 4.3, I present transition matrices comparing the distribution of students across baseline scores and outcomes scores. Each matrix indicates the percent of students in one baseline score quintile who are in that same quintile for the outcome measure. For example, in Table 4.2, 28.66% of students in the baseline second quintile are in the first outcome quintile, while 32.47% remain are again in the second quintile and 22.56% are in the third quintile. For all baseline quintiles in both samples, more students are in the same quintile for the outcome than any other quintile. However, only the first and fifth baseline quintiles (the lowest and highest

baseline N2 scores) have more than half of the students who make them up remain in the same quintile for the outcome scores; these quintiles have only one direction in which students can move on the matrix, so it is unsurprising that they see less movement. That most students are not in the same quintile of scores at the baseline as they are at the outcome supports the assumption that students are developing at different rates, and the distribution of students is quite different between these two points of assessment.

Table 4.2. Transition Matrix for Baseline N2 Score and N2 Score at the End of the First Year of College (Presented in Quintiles)

N2 Score Quintile at Baseline Assessment	N2 Score Quintile at End of First-Year (%)				
	1	2	3	4	5
1	56.64	25.95	10.84	5.34	1.22
2	28.66	32.47	22.56	13.72	2.59
3	11.45	25.95	28.55	23.36	10.69
4	2.29	11.74	27.90	33.84	24.24
5	0.92	3.97	10.08	23.82	61.22

Note. Each quintile represents 655 or 656 students.

Table 4.3. Transition Matrix for Baseline N2 Score and N2 Score at the End of the Fourth Year of College (Presented in Quintiles)

N2 Score Quintile at Baseline Assessment	N2 Score Quintile at End of Fourth-Year (%)				
	1	2	3	4	5
1	51.03	24.77	13.51	7.50	3.19
2	29.21	29.40	22.66	13.11	5.62
3	12.57	25.70	25.70	24.02	12.01
4	5.62	14.23	24.16	29.96	26.03
5	1.5	6.00	13.88	25.52	53.10

Note. Each quintile represents 533 or 534 students.

This study is also built on the assumption not just that students' moral judgment changes during college, but that extent of that change is different based on the college or university those

students attend. The clustered nature of the data used in these analyses support the use of a multilevel approach to model estimation; however, it is these assumed differences between institutions that support the conceptual value of this approach. To examine this assumption, I expanded the *t*-tests presented in Table 4.1 to consider the differences in N2 score between the baseline and the two outcome assessments for the subsample of students at each college and university.

Unspecified Models

After determining that moral judgment does indeed develop during college, I turned next to examining influences on that development. The first step in estimating models examining the effect of college experiences on the development of moral judgment (using developmental, college impact, and integrated frameworks) is to estimate an unspecified model. These estimated HLM models indicate whether the dependent variable (N2 score) varies significantly both within colleges and between colleges. Being unspecified, the model includes no independent variables. As with the *t*-tests in the previous section, I estimated two models: one each for the first-year and fourth-year samples; these are shown in Table 4.4.

Table 4.4 Covariance Parameter Estimates and Interclass Correlations (ICC) of N2 Score for the First-Year and Fourth-Year Samples

	Estimate	Standard Error	*z*-Statistic	ICC
First-Year Model (*n*=3277)				.220
Intercept (τ)	57.74	13.761	4.20***	
Residual (σ^2)	205.10	5.102	40.21***	
Four-Year (*n*=2667)				.187
Intercept (τ)	43.54	10.586	4.11***	
Residual (σ^2)	189.08	5.221	36.22***	

Note. *** p<.001.

Results from these models indicate that for both samples, students' N2 scores vary both between schools and within schools (p<.001); in Table 4.4, between-school variance is represented by τ and within-school variance is represented by σ^2. Further, the interclass correlation (ICC) indicates that 22.0 percent of the variance in N2 score at the end of the first year and 18.7 percent of the variance in N2 score at the end of the fourth year occur between rather than within colleges. That variance is statistically significant both between and within schools indicate that simply estimating ordinary least squares models, which would not account for between-school variance, would be inadequate. To do so would increase the chance of committing Type 1 error (concluding that a significant relationship exists between an independent and dependent variable when one does not) by underestimating standard errors and, thus, overestimating *t*-statistics. Although developing moral judgment is an individual outcome for students, the interclass correlation demonstrates that any model of it that includes only individual-level variables would ignore a portion part of the variance, potential leading to incorrect and incomplete theory-building. Thus, both statistically and conceptually, HLM is the appropriate method for estimating students' development of moral judgment during college.

Precollege Characteristics

The first step in creating prediction models of students' development of moral judgment in college is to examine the effects of student's precollege characteristics. All models – both college impact and developmental – that inform this study include an important role for baseline data in both measuring change during college and better isolating the effects of college experiences. All models discussed in Chapter II include precollege characteristics in addition to in-college experiences; this approach allows researchers to better account for differences in the outcome (in this case, moral judgment as measured by the DIT-2 N2 score) that can be

attributable to changes while in college rather than differences in the outcome that are simply attributable to student differences before they arrived on campus.

The baseline measure of a student's moral judgment is highly predictive of a student's DIT-2 N2 score at the end of the both the first year (b=0.555, p<.001) and the fourth year (b=0.503; p<.001). This means that even when accounting for other precollege factors, a student's moral judgment when entering college accounts for a large portion of the differences between students as they progress through college. In fact, an increase in one standard deviation of the baseline measure accounts for more than half of a standard deviation in the outcome measure, even after controlling for students' other precollege characteristics. It is also worth noting that because of the inclusion of a student's N2 score at the beginning of college, this model controls for baseline development. Thus, coefficients for other variables represent the effect of that variable on students' *change* in moral judgment during college.

Similarly, students' precollege academic is also significantly predictive of students' development of moral judgment (See Table 4.5). Like the effect of baseline N2 score, the effect of precollege academic ability declines from the end of the first year (b=.230; p<.001) to the end of the fourth year (b=0.195; p=0.001).

In addition to these measures of precollege moral judgment and academic ability, I also found several student demographic characteristics to be significant predictors of moral judgment at the end of the first and fourth year of college. Male students had lower N2 scores than female students at the end of the first (b=-.172; p<.001) and fourth (b=-.209; p=.001) years, even when accounting for other precollege characteristics in these models. No dummy variables representing racial or ethnic groups are significant predictors of moral judgment at the end of the first year; however, by the end of the fourth year there are statistically significant negative effects

on moral judgment for students who are African American (b=-.122; p<.10), Asian or Pacific Islander (b=-.178; p<.01), and Latino/a (b=-.189; p<.05).

Table 4.5. Standardized Effects of Precollege Characteristics on N2 Score at the End of the First and Fourth Years of College

	First-Year Sample (n=3277) x^2=41.14*** Adjusted r^2=.457			Four-Year Sample (n=2667) x^2=91.72*** Adjusted r^2=.344		
	Estimate	Standard Error	t-value	Estimate	Standard Error	t-value
Intercept	0.014	0.300	0.48	0.096*	0.043	2.21
DIT-2 N2 Baseline	0.555***	0.014	39.07	0.503***	0.018	28.05
Precollege Academic Ability	0.230***	0.017	13.97	0.195***	0.022	8.92
Male	-0.172***	0.024	-7.08	-0.211***	0.031	-6.73
African American	-0.073	0.054	-1.36	-0.122$^+$	0.072	-1.70
Native American	-0.209	0.235	-0.89	0.381	0.506	0.75
Asian/Pacific Islander	-0.046	0.054	-0.85	-0.176**	0.062	-2.84
Latino/a	0.021	0.058	0.36	-0.189*	0.078	-2.41
Politically Conservative	0.122*	0.032	2.56	0.045	0.042	1.09
Politically Liberal	0.230***	0.017	4.58	0.032	0.034	0.92

Note. $^+$ p<.10; * p<.05; ** p<.01; *** p<.001. Reference groups indicate students who are female, White, and "neither liberal nor conservative."

Finally, student's self-reported political orientation is a large and statistically significant predictor of moral judgment at the end of the first year. A student identifying his or her political viewpoint as moderately conservative or extremely conservative (b=0.122; p<.05) or as moderately liberal or extremely liberal (b=0.230; p<.001) had significant positive effects when compared to those who describe themselves as "neither liberal nor conservative." These effects do not persist in the fourth year.

A student's age and U.S. citizenship status were included in earlier iterations of these models, but neither had a significant effect on moral judgment at the end of the first or fourth years of college. In the interest of parsimonious model estimation, those variables were not included in any of analyses reported in the remainder of this chapter.

Developmental Framework Models

The first research question asks what experiences, within a *developmental framework*, predict students' development of moral judgment during college. To address this question, I present the results of two sets of estimated HLM models (one each for the first- and fourth-year samples; see Tables 4.6 and 4.7). The results of each are presented as block regressions, with the first block consisting of only the college experience variables of the developmental framework and the second block adding precollege characteristics.

Results of these models support the use of a developmental framework when examining the effects of college experiences on the development of moral judgment; all but one of the developmental variables are significant predictors of students' moral development during college. Students who report more frequent classroom encounters with disequilibrium exhibit higher levels of moral judgment, even after accounting for precollege characteristics, including their baseline N2 score, and this effect persisted throughout college. The effects of more classroom encounters with disequilibrium (first-year b=0.029; p<.05; fourth-year b=0.041; p<.05) and self-initiated encounters with disequilibrium (first-year b=0.110; p<.001; fourth-year b=0.099; p<.001) are statistically significant for both the first-year and fourth-year samples.

Table 4.6. Standardized Effects of College Experiences on N2 Score at the End of the First Year of College – Developmental Framework

	Developmental Framework Variables x^2=481.62*** r^2=.056			Developmental Framework and Precollege Variables x^2=21.78*** r^2=.480		
	Estimate	Standard Error	*t*-statistic	Estimate	Standard Error	*t*-statistic
Intercept	-0.071	0.068	-1.04	0.011	0.028	0.40
Classroom encounters with disequilibrium	0.046*	0.019	2.40	0.029*	0.014	2.00
Self-initiated encounters with disequilibrium	0.110***	0.020	5.62	0.049***	0.015	3.29
Meaningful conversations with diverse others	0.012	0.021	0.56	-0.011	0.015	-0.73
Negative interactions with diverse others	-0.180***	0.017	-10.68	-0.095***	0.013	-7.50
Experiences with Higher Order Learning	-0.055**	0.018	-3.08	-0.033*	0.013	-2.48
Experiences with Integrative Learning	0.054*	0.021	2.54	0.087***	0.016	5.61
DIT-2 N2 Baseline				0.544***	0.014	38.86
Precollege Academic Ability				0.221***	0.016	13.75
Male				-0.151***	0.024	-6.26
African American				-0.038	0.052	-0.73
Native American				-0.283	0.230	-1.23
Asian/Pacific Islander				0.004	0.053	-0.08
Latino/a				0.016	0.057	0.28
Politically Conservative				0.062*	0.031	1.99
Politically Liberal				0.108***	0.026	4.11

Note. + p<.10; * p<.05; ** p<.01; *** p<.001. Reference categories refer to a student who is female, White, and "neither liberal nor conservative."

These models also include two factors that represent students' encounters with diversity. The first, meaningful encounters with diverse others, was hypothesized the positively affect students' moral development by representing opportunities they have had to interact with peers

across types of differences, thus creating more opportunities for encounters with disequilibrium. However, in both models, this variable was not a statistically significant predictor of moral judgment. Conversely, negative interactions with diverse others were hypothesized negatively affect students' moral judgment by providing opportunities to reinforce students' belief systems rather than challenging them. This variable was a significant negative predictor in both the first-year (b=-.095; p<.001) and fourth-year (b=-.070; p<.001) models.

Two additional developmental variables represented aspects of deep learning that would indicate classroom engagement with academic material more cognitively aligned with more complex moral judgment schema; these were also significant predictors of moral judgment. Students who reported more experiences of integrative learning in their coursework exhibited higher levels of moral judgment in both the first-year (b=0.087; p<.001) and fourth-year (b=0.058; p<.05). Students who reported more experience with higher order thinking in their coursework, however, exhibited lower levels of moral judgment in first-year sample (b=-0.033; p<.05) and no effect in the fourth-year sample.

Table 4.7. Standardized Effects of College Experiences on N2 Score at the End of the Fourth Year of College, Using a Developmental Framework

	First-Year Sample x^2=477.40*** r^2=.047			Fourth-Year Sample x^2=99.76*** r^2=.364		
	Estimate	Standard Error	*t*-value	Estimate	Standard Error	*t*-value
Intercept	-0.136	0.079	-1.71	0.010*	0.044	2.29
Classroom encounters with disequilibrium	-0.046*	0.022	-2.06	-0.041*	0.020	-2.04
Self-initiated encounters with disequilibrium	0.139***	0.023	5.92	0.099***	0.024	4.82
Meaningful conversations with diverse others	0.001	0.027	0.05	0.029	0.021	1.38
Negative interactions with diverse others	-0.148***	0.021	-7.02	-0.070***	0.019	-3.69
Experiences with Higher Order Learning	-0.032	0.022	-1.46	-0.001	0.019	-0.03
Experiences with Integrative Learning	0.097***	0.029	3.35	0.058*	0.024	2.47
DIT-2 N2 Baseline				0.488***	0.018	27.51
Precollege Academic Ability				0.191***	0.021	8.76
Male				-0.183***	0.031	-5.87
African American				-0.131^{+}	0.072	-1.82
Native American				0.113	0.500	0.23
Asian/Pacific Islander				-0.178**	0.062	-2.88
Latino/a				-0.229**	0.078	-2.95
Politically Conservative				0.016	0.042	0.39
Politically liberal				0.009	0.034	0.27

Note. $^{+}$ p<.10; * p<.05; ** p<.01; *** p<.001. Reference categories refer to a student who is female, White, and "neither liberal nor conservative."

Results from these models support the value of using a developmental framework to investigate the development of students' moral judgment in college. As discussed in Chapter II, a large majority of the prior research on college students' moral development has employed

college impact frameworks, with very few studies incorporating developmental variables. These results, with five of six developmental variables found to have a statistically significant effect on students' development of moral judgment, support the position that studies that do not employ a developmental framework or variables consistent with such a framework are likely missing aspects of the college experience that affect students' development. Although the estimated effects in these models appear modest, they should be taken within the context of the size of students' first-year and four-year changes in N2 score. For the first-year sample, the average increase in N2 score is 0.282 standard deviations; that number rises to 0.542 standard deviations (see Table 4.1). In this context, the effect of a one standard deviation increase in negative interactions with diverse others in the first-sample (b=-0.095) is more than one-third of the mean change, and the effect a one standard deviation increase in self-initiated encounters with disequilibrium in the fourth-year sample (b=0.099) is almost one-fifth of the mean change. These results indicate that these variables consistent with a developmental framework play a role in the development of moral judgment; they also indicate that the size of their impact, compared to the size of total change, is significant.

College Impact Framework Models

The second research question asks what experiences, within a *college impact framework*, affect students' development of moral judgment. To address this question, I once again estimated HLM models using a block regression technique, first estimating the effects of individual student experiences, subsequently adding student major, institutional characteristics, and precollege characteristics (See Table 4.8 for first-year sample models and Table 4.9 for fourth-year sample models).

Results of these models suggest that student major is a significant predictor of moral judgment in the first and fourth years of college, but that once precollege characteristics are controlled for, few of students' other curricular and cocurricular experiences within this framework are statistically significant predictors. The difference between the estimates in the models that do and do not include precollege characteristics is especially stark for the first-year sample. In this model (Table 4.8), in even the third block of the model (when institutional characteristics are added) participating in a service-learning class (b=-0.108; p<.01), academic honors program (b=0.242; p<.01, leadership training program (b=-0.131; p<.05, community service (b=-0.096; p<.001), varsity intercollegiate athletics (b=-0.157; p<.01), and student organization leadership (b=-0.072; p<.10) all were significant predictors of moral judgment. These are not small effects; in all but one of these cases, the experience predicted a change in N2 score of approximately one-tenth of a standard deviation or more. However, once precollege characteristics were added to the model, only participating in community service (b=0.051; p<.05) remains statistically significant, and the standardized effect for that variable is much smaller. For the fourth-year sample, taking one class related to diversity or social justice issues (b=0.065; p<.05), participating in a leadership training program (b=0.099; p<.01), and participating in community service activities (b=0.071; p<.10) all remain significant predictors of N2 score after precollege characteristics are taken into account.

Although the results of these models point to negligible effects of many curricular and cocurricular experiences in the development of moral judgment during college, the models do support the effect of college major, a component of Terenzini and Reason's (2005) curricular experiences. Even after precollege characteristics are taken into account, every category of major has a statistically significant *negative* effect on moral judgment for the first-year sample when

compared to the humanities referent: biology (b=-0.114; p<.05), business (b=-0.256; p<.001), education (b=-0.153; p<.01), physical sciences (b=-0.165; p<.05), professional fields (b=-0.194; p<.001), social sciences (b=-0.131; p<.001), engineering (b=-0.218; p<.001), and undecided (b=-0.223; p<.001). Students who reported their major as "other" also had N2 scores that were marginally significantly lower than the humanities reference group (b=-0.081; p<.10). The fourth year sample offers a clearer look into the institutional effects of major since students had more exposure to the environments of their majors; in this analysis, four of majors continue to show significant negative effects on moral judgment: business (b=-0.239; p<.001), social sciences (b=-0.214; p<.001), engineering (b=-0.376; p<.001), and other (b=-0.180; p<.01). No major demonstrated a positive effect in either sample compared to the reference category of humanities majors.

Table 4.8. Standardized Effects of College Experiences on N2 Score at the End of the First Year of College – College Impact Framework

	College Experiences x^2=527.78*** Adjusted r^2=0.022			Add Major x^2=419.23*** Adjusted r^2=0.041		
	Estimate	Standard Error	*t*-value	Estimate	Standard Error	*t*-value
Intercept	-0.118	0.083	-1.43	0.152	0.087	1.76
Service Learning Class	-0.118***	0.036	-3.29	-0.098**	0.036	-2.75
Honors Program	0.259***	0.045	5.71	0.256***	0.045	5.70
One Diversity Class	-0.022	0.041	-0.54	-0.026	0.041	-0.65
More than one diversity class	-0.003	0.038	-0.08	-0.016	0.038	-0.43
Student Organization Leader	-0.092*	0.042	-2.22	-0.080^{+}	0.042	-1.93
Leadership Training Program	0.123*	0.056	2.21	0.134*	0.055	2.41
Religious Congregation	0.029	0.037	0.78	0.040	0.037	1.07
Social/Political Lecture	0.034	0.036	0.95	0.023	0.036	0.64
Community Service	0.105**	0.034	3.09	0.106**	0.034	3.13
Greek Organization	-0.070	0.051	-1.38	-0.071	0.050	-1.42
Varsity Athlete	-0.204***	0.050	-4.04	-0.148**	0.050	-2.94
Race Workshop	0.076*	0.036	2.14	0.053	0.019	1.29
Biology				-0.290***	0.065	-4.46
Business				-0.584***	0.063	-9.30
Education				-0.401***	0.077	-5.22
Physical Sciences				-0.176*	0.088	-2.00
Professional				-0.359***	0.065	-5.53
Social Sciences				-0.247***	0.057	-4.32
Engineering				-0.273**	0.105	-2.62
Other Major				-0.279***	0.062	-4.49
Undecided				-0.388***	0.080	-4.84
	Institutional Characteristics x^2=66.71 Adjusted r^2=0.047			Add Precollege x^2=34.02*** Adjusted r^2=0.461		
Intercept	0.188**	0.062	3.03	0.090^{+}	0.047	1.90
Service Learning Class	-0.108**	0.037	-2.96	-0.026	0.027	-0.96
Honors Program	0.242***	0.045	5.42	-0.020	0.034	-0.58
One Diversity Class	-0.021	0.041	-0.53	-0.005	0.031	-0.16
More than one diversity class	0.000	0.038	0.01	0.027	0.028	0.95
Student Organization Leader	-0.072^{+}	0.042	-1.70	-0.008	0.032	-0.26
Leadership Training Program	0.132*	0.056	2.37	0.054	0.041	1.32

Religious Congregation	0.017	0.036	0.48	-0.001	0.028	-0.04
Social/Political Lecture	0.006	0.036	0.18	0.007	0.028	0.25
Community Service	0.096**	0.034	2.84	0.051*	0.025	2.02
Greek Organization	-0.081	0.050	-1.62	0.021	0.037	0.57
Varsity Athlete	-0.157**	0.050	3.12	-0.022	0.037	-0.58
Race Workshop	0.054	0.036	1.50	0.023	0.014	1.36
Biology	-0.293***	0.064	-4.55	-0.114*	0.048	-2.35
Business	-0.555***	0.064	-8.66	-0.256***	0.048	-5.39
Education	-0.363***	0.077	-4.71	-0.153**	0.058	-2.69
Physical Sciences	-0.184*	0.087	-2.11	-0.165*	0.066	-2.51
Professional	-0.334***	0.066	-5.03	-0.194***	0.049	-3.99
Social Sciences	-0.243***	0.057	-4.26	-0.131***	0.043	-3.06
Engineering	-0.238*	0.104	-2.28	-0.218**	0.078	-2.79
Other Major	-0.235***	0.062	-3.80	-0.081^{+}	0.046	-1.74
Undecided	-0.347***	0.082	-4.21	-0.223***	0.062	-4.07
College Average Academic Ability	0.327***	0.033	9.93	-0.028	0.023	-1.20
Perceived Faculty Interest in Students	0.097***	0.215	4.52	0.025	0.021	1.21
DIT-2 N2 Baseline				0.547***	0.014	38.14
Precollege Academic Ability				0.230***	0.019	12.35
Male				-0.148***	0.025	-5.88
African American				-0.074	0.055	-1.34
Native American				-0.267	0.235	-1.14
Asian/Pacific Islander				-0.023	0.055	-0.41
Latino/a				0.007	0.058	0.11
Politically Conservative				0.083**	0.032	2.58
Politically Liberal				0.111***	0.027	4.11

Note. $^{+}$ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$. Reference groups indicate a student who is female, White, "neither liberal nor conservative," a humanities major, and who did not report any of the experiences included in the models.

Table 4.9. Standardized Effects of College Experiences on N2 Score at the End of the Fourth Year of College – College Impact Framework

	College Experiences x^2=388.70*** Adjusted r^2=0.017			Add Major x^2=341.65 Adjusted r^2=0.045		
	Estimate	Standard Error	*t*-value	Estimate	Standard Error	*t*-value
Intercept	-0.261*	0.105	-2.49	-0.010	0.109	-0.09
Service Learning Class	-0.125**	0.045	-2.75	-0.137**	0.045	-3.04
Honors Program	0.110**	0.042	2.60	0.122**	0.042	2.93
One Diversity Class	0.028	0.038	0.74	0.043	0.037	1.17
More than one diversity class	0.052	0.047	1.10	0.036	0.047	0.76
Student Organization Leader	-0.026	0.045	-0.58	-0.014	0.044	-0.32
Leadership Training Program	0.123**	0.043	2.89	0.147***	0.042	3.48
Religious Congregation	-0.070^{+}	0.041	-1.74	-0.081*	0.040	-2.01
Social/Political Lecture	0.022*	0.053	0.42	0.014*	0.053	0.27
Community Service	0.184***	0.048	3.80	0.179***	0.048	3.74
Greek Organization	-0.094^{+}	0.055	-1.70	-0.068	0.054	-1.24
Varsity Athlete	-0.068	0.054	-1.25	0.009	0.055	0.16
Race Workshop	-0.010	0.043	0.23	0.027	0.043	0.64
Biology				-0.223**	0.072	-3.15
Business				-0.570***	0.074	-7.80
Education				-0.231*	0.108	-2.22
Physical Sciences				-0.162	0.097	-1.62
Professional				-0.111	0.084	-1.58
Social Sciences				-0.396***	0.062	-6.60
Engineering				-0.597***	0.117	-5.25
Other Major				-0.444***	0.079	-5.71
	Institutional Characteristics x^2=55.28 Adjusted r^2=0.046			Add Precollege x^2=71.72*** Adjusted r^2=0.355		
Intercept	0.101	0.091	1.11	0.107	0.078	1.37
Service Learning Class	-0.126**	0.048	-2.82	0.009	0.037	0.25
Honors Program	0.134**	0.042	3.21	-0.050	0.035	-1.43
One Diversity Class	0.042	0.037	1.12	0.065*	0.031	2.11
More than one diversity class	0.038	0.047	0.81	0.044	0.039	1.14
Student Organization Leader	-0.020	0.044	-0.45	-0.012	0.037	-0.32
Leadership Training Program	0.151***	0.042	3.59	0.099**	0.035	2.81
Religious Congregation	-0.083*	0.040	-2.08	-0.029	0.034	-0.85

Social/Political Lecture	0.008	0.052	0.15	-0.021	0.043	-0.48
Community Service	0.170***	0.048	3.56	0.071^{+}	0.040	1.79
Greek Organization	-0.066	0.054	-1.23	0.056	0.044	1.27
Varsity Athlete	0.001	0.054	0.01	0.024	0.045	0.54
Race Workshop	0.024	0.043	0.57	0.002	0.029	0.07
Biology	-0.204**	0.072	-2.85	-0.077	0.059	-1.29
Business	-0.552***	0.074	-7.49	-0.239***	0.062	-3.88
Education	-0.203^{+}	0.107	-1.90	-0.105	0.088	-1.19
Physical Sciences	-0.162^{+}	0.097	-1.67	-0.109	0.080	-1.37
Professional	-0.093	0.083	-1.12	-0.072	0.069	-1.05
Social Sciences	-0.396***	0.067	-6.47	-0.214***	0.051	-4.22
Engineering	-0.573***	0.114	-5.04	-0.376***	0.094	-3.98
Other Major	-0.419***	0.079	-5.34	-0.180**	0.065	-2.77
College Average Academic Ability	0.350***	0.041	8.61	0.030	0.035	0.86
Perceived Faculty Interest in Students	0.060**	0.022	2.73	0.049	0.018	2.70
DIT-2 N2 Baseline				0.487***	0.018	26.83
Precollege Academic Ability				0.201***	0.024	8.41
Male				-0.173***	0.032	-5.33
African American				-0.099	0.074	-1.34
Native American				0.363	0.503	0.72
Asian/Pacific Islander				-0.136*	0.062	-2.18
Latino/a				-0.197*	0.079	-2.50
Politically Conservative				0.037	0.043	0.87
Politically Liberal				0.032	0.035	0.93

Note. $^{+}$ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$. Reference groups indicate a student who is female, White, "neither liberal nor conservative," a humanities major, and who did not report any of the experiences included in the models.

Beyond students' individual experiences and majors, college impact frameworks also encompass the effects of institutional characteristics on students' outcomes. The model by Terenzini and Reason (2005), which guided the development of the college impact framework models in this chapter, specifies two types of institutional characteristics that potentially affect student outcomes: the organizational context and the peer environment. Only three variables from these dimensions of Terenzini and Reason's model are significant in any of the blocks of college impact framework models included in this study: a student's perceptions of faculty interest in teaching and student development (representative of faculty culture) and a school-level variable representing the mean precollege academic ability of an institution's study cohort (representing peer environment). Before controlling for precollege characteristics, both of these were significant predictors of students' N2 score at the end of the first year (faculty interest b=0.327; p<.001 and average precollege ability b=0.097; p<.001); at the end of the fourth year, two of these remained significant (faculty interest b=0.350; p<.001; average precollege ability b=0.060; p<.001). However, like many of the individual student experience variables in this framework, when the block of precollege characteristics are added to the models, both institutional characteristics ceased to be statistically significant predictors of N2 score.

Integrated Models

The third research question asks what college experiences predict change in students' moral judgment when an integrated framework (which incorporates the tenets of both a developmental and college impact framework) is used. To address this question, I estimated HLM models for both samples using the same block modeling process as the previous models. The first block comprises two items unique to the integrated framework, the second block adds

all college experience variables from the developmental and college impact framework models, and the third block adds students' precollege characteristics.

As discussed in Chapter III, the integrated model includes two variables that do not appear in the other two models. These are factors representing students' perceptions of the extent to which their institution: a) supports social interactions with diverse others; and b) contributes to the personal, moral, and spiritual development of its students. It was hypothesized that the first of these factors would yield a positive effect on students' development of moral judgment because it would indicate that institutions were providing students opportunities to interact with diverse others and developing the skills to do so in ways that promote their learning. However, this variable has a marginally significant negative effect on N2 score for the first-year sample (b=-0.028; p<.10) (Table 4.10) and larger statistically significant negative effects on the fourth-year sample (b=-.053; p<.05) (Table 4.11). The second variable is not statistically significant in either model.

The variables from the developmental framework models are statistically significant predictors of moral judgment in ways that are very similar to the original models. As they were in the developmental framework model, most of the developmental variables have significant positive (self-initiated encounters with disequilibrium, experiences with integrative learning) or significant negative (negative interactions with diverse others) effects for both the first- and fourth-year samples. Also like the original developmental framework models, reported classroom encounters with disequilibrium has a significant positive effect for the first-year sample (b=0.033; p<.05) and a negative effect for the fourth-year sample (b=-0.051; p<.01), and experiences with higher order learning has a significant effect for the first-year sample (b=0.024; p<.10) and no effect in the fourth-year sample. The one change between the results of the

developmental framework models and the effects of the developmental variables in the integrated framework models relate to the effects of meaningful conversations with diverse others. In the developmental framework models, there are no significant effects for this variable for either sample; however, in the integrated framework models, there was a marginally significant positive effect on N2 score for the fourth-year sample (b=0.041; p<.10).

Comparing the effects for major across the college impact framework models and the integrated framework models, there are several differences between the models for the effects of other curricular and out-of-class experiences. For the first-year sample, the only experience (excluding student major) that was a significant predictor of moral judgment was participating in community service. In the integrated framework model, however, this variable is not a significant predictor. In other words, none of the variables of the type most frequently posited by researchers using the college impact framework to examine moral judgment significantly predicted development during the first year of college when modeled using an integrated framework.

In the college impact framework model for the fourth-year sample, taking one class that focuses on diversity or social justice and participating in community service exhibit significant positive effects on moral judgment; however, the effect of neither variable is significant in the integrated framework model. Although being a member of a social fraternity or sorority did not have a significant effect in the college impact framework model, it was marginally significant (b=0.076; p<.10) when the precollege characteristics were considered. Overall, participating in leadership training and belonging to a social fraternity or sorority are the only curricular or out-of-class experiences (excluding student major) that have even marginally significant effects for

either sample when modeled using an integrated framework (after controlling for precollege characteristics).

Table 4.10. Standardized Effects of College Experiences on N2 Score at the End of the First Year of College – Integrated Framework

	Variables Included only in Integrated Framework x^2=632.50*** Adjusted r^2=0.001			Integrated Framework x^2=28.67*** Adjusted r^2=0.094		
	Estimate	Standard Error	*t*-value	Estimate	Standard Error	*t*-value
Intercept	-0.074	0.074	-0.96	0.158**	0.058	2.70
Institution supports social interaction	-0.005	0.010	0.14	-0.050*	0.020	-2.48
Institution contributes to my development	-0.025	0.006	-0.74	-0.063*	0.023	-2.73
Classroom encounters with disequilibrium				0.054**	0.020	2.75
Self-initiated encounters with disequilibrium				0.113***	0.020	5.76
Meaningful conversations with diverse others				0.010	0.021	0.47
Negative interactions with diverse others				-0.179***	0.016	-11.20
Experiences with higher order learning				-0.045*	0.018	-2.48
Experiences with integrative learning				0.049*	0.023	2.09
Service Learning Class				-0.082*	0.035	-2.35
Honors Program				0.226***	0.043	5.24
One Diversity Class				-0.033	0.040	-0.84
More than one diversity class				-0.005	0.037	-0.13
Student Organization Leader				-0.061+	0.041	-1.43
Leadership Training Program				0.119*	0.055	2.18
Religious Congregation				0.032	0.035	0.92
Social/Political Lecture				-0.001	0.035	-0.04
Community Service				0.041	0.033	1.23
Greek Organization				-0.058	0.048	-1.20
Varsity Athlete				-0.122*	0.049	-2.50
Race Workshop				0.078*	0.035	2.27
Biology				-0.253***	0.063	-4.02
Business				-0.492***	0.063	-7.80
Education				-0.352***	0.076	-4.66
Physical Sciences				-0.123	0.085	-1.44
Professional				-0.288***	0.064	-4.50

Social Sciences	-0.186***	0.056	-3.35
Engineering	-0.235*	0.101	-2.32
Other Major	-0.241***	0.060	-4.00
Undecided	-0.346***	0.080	-4.34
College Average Academic Ability	0.297***	0.026	11.25
Perceived Faculty Interest in Students	0.078**	0.023	3.45

	Integrated Framework with Precollege Characteristics x^2=18.76*** Adjusted r^2=.479		
Intercept	0.102*	0.047	2.19
Institution supports social interaction	-0.028^{+}	0.015	-1.87
Institution contributes to my development	-0.017	0.016	-1.01
Classroom encounters with disequilibrium	0.033*	0.016	2.08
Self-initiated encounters with disequilibrium	0.046**	0.016	2.97
Meaningful conversations with diverse others	-0.009	0.016	-0.58
Negative interactions with diverse others	-0.103***	0.012	-8.36
Experiences with higher order learning	-0.024^{+}	0.014	-1.73
Experiences with integrative learning	0.087***	0.017	5.14
Service Learning Class	-0.024	0.026	-0.91
Honors Program	-0.025	0.033	-0.73
One Diversity Class	-0.018	0.030	-0.61
More than one diversity class	-0.004	0.028	0.15
Student Organization Leader	-0.010	0.031	-0.33
Leadership Training Program	0.045	0.041	1.10
Religious Congregation	0.012	0.028	0.45
Social/Political Lecture	-0.006	0.028	-0.23
Community Service	0.020	0.025	0.74
Greek Organization	0.049	0.037	1.33
Varsity Athlete	-0.002	0.037	-0.05
Race Workshop	0.045	0.028	1.59
Biology	-0.107*	0.048	-2.22
Business	-0.229***	0.047	-4.85

Education	-0.162**	0.057	-2.86
Physical Sciences	-0.119^{+}	0.065	-1.83
Professional	-0.179***	0.048	-3.73
Social Sciences	-0.110**	0.042	-2.61
Engineering	-0.212**	0.077	-2.73
Other Major	-0.086^{+}	0.046	-1.88
Undecided	-0.226***	0.060	-3.80
College Average Academic Ability	-0.024	0.021	-1.15
Perceived Faculty Interest in Students	0.003	0.021	0.13
DIT-2 N2 Baseline	0.538***	0.014	38.05
Precollege Academic Ability	0.219***	0.018	11.87
Male	-0.134***	0.053	-5.38
African American	-0.042	0.230	-0.78
Native American	-0.309	0.054	-1.34
Asian/Pacific Islander	0.006	0.058	0.12
Latino/a	0.025	0.071	0.43
Politically Conservative	0.067***	0.032	2.10
Politically Liberal	0.101***	0.027	3.78

Note. $^{+}$ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$. Reference groups indicate a student who is female, White, "neither liberal nor conservative," a humanities major, and who did not report any of the experiences included in the models.

Table 4.11. Standardized Effects of College Experiences on N2 Score at the End of the Fourth Year of College – Integrated Framework

	New Integrated Framework Variables x^2=462.64*** Adjusted r^2=0.005			Integrated Framework x^2=56.90*** Adjusted r^2=0.090		
	Estimate	Standard Error	*t*-value	Estimate	Standard Error	*t*-value
Intercept	-0.131	0.080	-1.63	0.111	0.094	1.18
Institution supports social interaction	-0.072***	0.022	-3.25	-0.109***	0.023	-4.63
Institution contributes to development	0.087***	0.024	3.57	0.012	0.026	0.45
Classroom encounters with disequilibrium				-0.054*	0.023	-2.35
Self-initiated encounters with disequilibrium				0.143***	0.024	5.99
Meaningful conversations with diverse others				0.021	0.025	0.81
Negative interactions with diverse others				-0.143***	0.024	-5.86
Experiences with higher order learning				-0.029	0.022	-1.35
Experiences with integrative learning				0.086**	0.028	3.10
Service Learning Class				-0.140**	0.045	-3.11
Honors Program				0.150***	0.041	3.65
One Diversity Class				0.024	0.037	0.66
More than one diversity class				0.020	0.047	0.44
Student Organization Leader				-0.031	0.044	-0.71
Leadership Training Program				0.139***	0.042	3.29
Religious Congregation				-0.064	0.039	-1.63
Social/Political Lecture				0.012	0.052	0.22
Community Service				0.143**	0.048	2.98
Greek Organization				-0.038	0.053	-0.71
Varsity Athlete				0.021	0.062	0.40
Race Workshop				-0.029*	0.025	0.68
Biology				-0.233**	0.072	-3.24
Business				-0.488***	0.074	-6.60
Education				-0.173	0.106	-1.64
Physical Sciences				-0.109	0.097	-1.13
Professional				-0.106	0.083	-1.27

Social Sciences	-0.367***	0.061	-6.04
Engineering	-0.496**	0.115	-4.31
Other Major	-0.340**	0.079	-4.32
College Average Academic Ability	0.353	0.040	8.89
Perceived Faculty Interest in Students	0.061	0.229	2.65
	Integrated Framework with Precollege Characteristics x^2=71.18*** Adjusted r^2=0.371		
Intercept	0.168	0.081	2.07
Institution supports social interaction	-0.053**	0.020	-2.67
Institution contributes to development	-0.010	0.022	-0.48
Classroom encounters with disequilibrium	-0.051**	0.019	-2.68
Self-initiated encounters with disequilibrium	0.096***	0.019	4.99
Meaningful conversations with diverse others	0.041+	0.021	1.92
Negative interactions with diverse others	-0.072***	0.018	-3.89
Experiences with higher order learning	0.002	0.018	0.13
Experiences with integrative learning	0.057*	0.023	2.50
Service Learning Class	-0.018	0.038	-0.47
Honors Program	-0.035	0.035	-1.01
One Diversity Class	0.054	0.030	1.77
More than one diversity class	0.021	0.039	0.55
Student Organization Leader	-0.026	0.037	-0.71
Leadership Training Program	0.092*	0.037	2.50
Religious Congregation	-0.015	0.034	-0.45
Social/Political Lecture	-0.026	0.043	-0.60
Community Service	0.058	0.040	1.46
Greek Organization	0.076+	0.045	1.73
Varsity Athlete	0.035	0.051	0.79
Race Workshop	-0.006	0.080	-0.19
Biology	-0.096	0.060	-1.60
Business	-0.209***	0.062	-3.37
Education	-0.088	0.088	-1.00

Physical Sciences	-0.084	0.081	-1.04
Professional	-0.077	0.070	-1.11
Social Sciences	-0.207***	0.050	-4.10
Engineering	-0.341***	0.096	-3.55
Other Major	-0.137*	0.065	-2.10
College Average Academic Ability	0.043	0.035	1.22
Perceived Faculty Interest in Students	0.052**	0.019	2.77
DIT-2 N2 Baseline	0.477***	0.018	26.54
Precollege Academic Ability	0.186***	0.024	7.75
Male	-0.156***	0.032	-4.79
African American	-0.124^{+}	0.074	-1.68
Native American	0.191	0.498	0.38
Asian/Pacific Islander	-0.140*	0.062	-2.26
Latino/a	-0.215**	0.077	-2.77
Politically Conservative	0.019	0.043	0.45
Politically Liberal	0.004	0.035	0.13

Note. $^{+}$ $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$. Reference groups indicate a student who is female, White, "neither liberal nor conservative," a humanities major, and who did not report any of the experiences included in the models.

A Comparative Analysis of the Explanatory Power

This dissertation is focused not just on the specific college experiences that encourage or inhibit students' development of moral judgment, but also on ways researchers can most effectively investigate these relationships. Thus, the fourth research question addresses whether an integrated framework that comprises both the developmental and college impact frameworks has more explanatory power than either of those frameworks by itself. To address this question, I present the pseudo-r^2 and adjusted pseudo- r^2 of each of the frameworks presented above; these values represent the proportion of variance in the outcome explained by each model (see Table 4.12).

These values indicate that the integrated framework explains a greater proportion of the variance of N2 score for both the first-year and fourth-year samples than either the developmental or college impact frameworks alone. In the first-year sample, these differences are greater. For the first-year sample, the developmental framework, not including the block of precollege characteristics, accounts for 0.058 of the total variance and the college impact framework accounts for 0.054; when these two frameworks are integrated, the total proportion of variance explained is 0.103. Although the inclusion of additional independent variables will, by definition, increase a model's pseudo-r^2 value, the adjusted pseud-r^2 value for the models show a greater proportion of the variance explained by the integrated framework even after accounting for the larger number of independent variables. The adjusted pseudo r^2 for the integrated framework is 0.094, compared to 0.056 and 0.047, respectively, for the developmental and college impact frameworks on their own.

This pattern is similar for the fourth-year sample, with the integrated model having an adjusted pseudo r^2 of 0.090 compared to 0.047 and 0.046, respectively, for the developmental

and college impact frameworks (all without the precollege characteristics). The integrated framework accounted for almost twice as much variance in both samples as either of the other frameworks by themselves.

Table 4.12. Pseudo-r^2 and Adjusted Pseudo-r^2 for Estimated Models

	Number of Independent Variables	Pseudo r^2	Adjusted Pseudo r^2
First-Year Sample			
Precollege Characteristics	10	0.459	0.457
Developmental Framework	6	0.058	0.056
Developmental Framework with Precollege Characteristics	16	0.482	0.480
College Impact Framework	23	0.054	0.047
College Impact Framework with Precollege Characteristics	33	0.467	0.461
Integrated Framework	31	0.103	0.094
Integrated Framework with Precollege Characteristics	41	0.486	0.479
Fourth-Year Sample			
Precollege Characteristics	10	0.345	0.344
Developmental Framework	6	0.048	0.047
Developmental Framework with Precollege Characteristics	16	0.367	0.364
College Impact Framework	22	0.052	0.046
College Impact Framework with Precollege Characteristics	32	0.362	0.355
Integrated Framework	30	0.098	0.090
Integrated Framework with Precollege Characteristics	40	0.379	0.371

Note. The first-year sample includes one additional category of major (undecided). This option was not available for students in the fourth-year sample.

Table 4.12 also shows that the adjusted pseudo-r^2 values increased dramatically when precollege characteristics were added to the model. By themselves, the precollege variables explained 0.457 and 0.343 of the variance, respectively, in the first- and fourth-year samples, after adjusting for the number of independent variables. When precollege characteristics were added to the integrated framework models, the adjusted pseudo r^2 values increased to 0.479 and

0.367, respectively. These proportions of explained variance are still higher than the variance explained by the developmental or college impact frameworks by themselves, but the magnitude of the difference is smaller once the effects of precollege characteristics are taken into account. Given the large effect of the baseline measure of moral judgment on the later measures of the outcome, as well as large effect of precollege academic ability, the size of the adjusted pseudo r^2 values for models including the precollege characteristics is not surprising; however, it puts into context the relatively small effects of college experiences included in these models on moral judgment when compared to students' characteristics before entering college. All of the resources put into these college programs and the other experiences and institutional characteristics explain virtually no additional variance in moral judgment than the characteristics with which students enter.

Omitted Variable Bias and Changes in Effect Sizes

The previous section demonstrates that the integration of the developmental and college impact frameworks result in greater explanatory power in the estimated models, and this held for both the first-year and fourth-year samples. However, increasing explanatory power is not the only way that integrating the two models could alter the results and conclusions drawn from the model. In the integrated models, college experiences from both the developmental and college impact frameworks are estimated to have statistically significant relationships with changes in students' moral judgment. This indicates that both the developmental and college impact framework models are underspecified (they exclude significant independent variables) and are at risk for omitted variable bias. Omitted variable bias results when an underspecified model excludes at least one variable that is correlated with both the outcome and one or more of independent variables included in the model; this bias can cause parameter estimates of those

correlated independent variables to be either overestimated or underestimated (Chamberlain, 1979; Clark, 2005).

For example, in the developmental framework, negative interactions with diverse others has a significant negative effect in both the first- and fourth-year samples. In the college impact framework, being an engineering major also has a negative effect across both samples. It is possible, that engineering students have more negative interactions with diverse others than other students, perhaps do to the relatively high number of international students often in engineering programs and the emphasis on completing assignments in teams. These diverse teams could lead to tension and negative perceptions across differences that extend to non-engineering contexts. In this case, part of the effect of being an engineering major in the college impact framework could in fact be the negative effect of negative interactions with diverse peers, which would lead to an overestimation of the engineering effect. Conversely, engineering students might have fewer negative interactions with diverse peers, again, perhaps because of a larger number of international students and an emphasis on group projects. These diverse teams could, instead, lead to a more interaction and better understanding across difference that also extend to non-engineering contexts leading to fewer negative interactions with diverse peers. In this case, the positive effect of engineering on another negative predictor excluded from the college impact mode could lead to underestimating the effect of being an engineering major in the model.

Omitted variable bias resulting from the lack of integration of the developmental and college impact frameworks is particularly relevant in light of these findings. In this study, the variables included in the developmental framework were much more likely to be statistically significant compared to the variables in the college impact framework; therefore in a body of literature that rarely utilizes a developmental framework (see examples in Chapter II),

researchers using the college impact framework run the risk of consistently underestimating or overestimating the effects of college experiences.

To examine the potential of omitted variable bias in the developmental and college impact frameworks, I present the parameter estimates for statistically significant parameters in the three different models, as well as confidence intervals for the parameters, in Tables 4.13 and 4.14. Changes in the value of the parameter between the two models, suggests a threat of omitted variable bias. (Because a parameter that is not statistically significant is not statistically different from zero, two nonsignificant parameters, by definition, cannot be said to be different from one another. Those parameters have been removed from these analyses). These results indicate that there are differences between the parameter in the original frameworks and the integrated framework in 16 of 17 parameters in the first-year sample and all 16 in the fourth-year sample; for 15 of the parameters across the two samples, this difference was at least 0.01, or 1% of a standard deviation in the outcome.

Table 4.13. Standardized Coefficients and Confidence Intervals for the Estimated Effect of College Experiences on First-Year N2 Score

	Original Framework		Integrated Framework	
	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
Developmental Framework Parameters				
Intercept	0.011	[-0.043, 0.065]	0.102*	[0.011, 0.193]
Classroom encounters with disequilibrium	0.029*	[0.001, 0.059]	0.033*	[0.002, 0.064]
Self-initiated encounters with disequilibrium	0.049***	[0.020, 0.081]	0.046**	[0.016, 0.077]
Negative interactions with diverse others	-0.095***	[-0.122, -0.076]	-0.103***	[-0.127, -0.079]
Experiences with higher order	-0.033*	[-0.059, -0.005]	-0.024+	[-0.052, -0.003]
Experiences with integrative learning	0.087***	[0.055, 0.117]	0.087***	[0.054, 0.120]
College Impact Framework Parameters				
Intercept	0.090+	[-0.003, 0.182]	0.102*	[0.011, 0.193]
Community Service	0.051*	[0.002, 0.101]	0.020	[-0.029, 0.069]
Biology	-0.114*	[-0.209, -0.019]	-0.107*	[-0.200, -0.012]
Business	-0.256***	[-0.349, -0.163]	-0.229***	[-0.321, -0.136]
Education	-0.153**	[-0.267, -0.042]	-0.162**	[-0.274, -0.051]
Physical Sciences	-0.165*	[-0.293, -0.036]	-0.119+	[-0.246, -0.008]
Professional	-0.194***	[-0.290, -0.099]	-0.179***	[-0.273, -0.085]
Social Sciences	-0.131***	[-0.214, -0.047]	-0.110**	[-0.192, -0.027]
Engineering	-0.218**	[-0.372, -0.648]	-0.212**	[-0.363, -0.060]
Other Major	-0.081+	[-0.172, 0.010]	-0.086+	[-0.176, 0.004]
Undecided	-0.223***	[-0.343, -0.104]	-0.226***	[-0.034, -0.110]

Note. + $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$.

In addition comparing the parameter estimates themselves, I also calculated 95% confidence intervals for each of the estimates (also shown in Table 4.13 and 4.14). If the original parameter estimate from the developmental or college impact framework models falls outside the confidence interval of the estimate of the same parameter in the integrated framework model, the two estimates can be assumed to be statistically different than one another ($p<.05$). In both the first-year and fourth-year samples, none of the estimates from the original models are statistically significantly different than the integrated models. However, because both the multiple

imputation process and the estimation of multilevel models inflate the standard errors of the parameter estimates (which in turn leads to wider confidence intervals), the lack of significant differences between the estimates in this study should not provide a false sense of confidence that this would be the case in other studies with less conservative confidence intervals.

Table 4.14. Standardized Coefficients and Confidence Intervals for the Estimated Effect of College Experiences on Fourth-Year N2 Score

	Original Framework		*Integrated Framework*	
	Estimate	95% Confidence Interval	Estimate	95% Confidence Interval
Developmental Framework Variables				
Intercept	0.100*	[0.015, 0.186]	0.168	[0.009, 0.327]
Classroom encounters with disequilibrium	-0.041*	[-0.075, -0.003]	-0.051**	[-0.088, -0.014]
Self-initiated encounters with disequilibrium	0.099***	[0.057, 0.131]	0.096***	[0.059, 0.134]
Meaningful conversations with diverse others	0.029	[-0.012, 0.064]	0.041+	[-0.001, 0.083]
Negative interactions with diverse others	-0.070***	[-0.101, -0.027]	-0.072***	[-0.109, -0.035]
Experiences with integrative learning	0.058*	[0.020, 0.110]	0.057*	[0.012, 0.103]
College Impact Framework				
One Diversity class	0.065*	[0.004, 0.124]	0.054	[-0.006, 0.113]
Leadership Training	0.099**	[0.036, 0.182]	0.092*	[0.020, 0.164]
Community Service	0.071+	[-0.007, 0.148]	0.058	[-0.020, 0.135]
Greek Organization	0.056	[-0.028, 0.150]	0.076+	[-0.010, 0.166]
Race Workshop	0.002	[-0.068, 0.70]	-0.006+	[-0.076, 0.062]
Business	-0.239***	[-0.361, -0.120]	-0.209***	[-0.330, -0.087]
Social Sciences	-0.214***	[-0.313, -0.115]	-0.207***	[-0.306, -0.108]
Engineering	-0.376***	[-0.562, -0.192]	-0.341***	[-0.529, -0.152]
Other Major	-0.180**	[-0.308, -0.053]	-0.137*	[-0.265, -0.009]
Perceived Faculty Interest in Students	0.049	[0.016, 0.086]	0.052**	[0.015, 0.089]

Note. + $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$.

In this chapter, I have presented results that address the development of moral judgment during college, as examined by utilizing developmental, college impact, and integrated frameworks. The results presented above point to the value of the integrated framework, indicating that the framework combines components of the variance in moral judgment predicted in the two other frameworks, that the integrated framework has more greater explanatory power than the individual developmental or college impact frameworks, and that the developmental and college impact framework are at greater risk of omitted variable bias. These results also point to the importance of precollege characteristics in the development of college even in college, as well as the relative limited effects of most college activities. I will discuss these results more thoroughly in the next chapter.

CHAPTER V: DISCUSSION AND IMPLICATIONS

This study examines the value of using an integrated framework to examine the effect of college experiences on students' development of moral judgment. In this chapter, I discuss in detail the major findings of this study, outlining the factors potentially contributing to these results and drawing conclusions about the methodologies used in this study. I also suggest methodological implications for research on the development of moral judgment in college and directions for this research.

Changes in Moral Judgment during College

Results from this study indicated that students do demonstrate positive change in their moral judgment during both the first year and the first four years of college, as evidenced by their DIT-2 N2 scores. Consistent with previous research, these findings showed that students develop in moral judgment during college, becoming more likely to use moral judgments based on postconventional schema and to eschew moral reasoning based on a maintaining norms schema.

In addition to establishing that the development of moral judgment did happen during college for these samples, the results of these initial analyses also provide an indication of the magnitude of that change for these samples. In the first year of college, the mean change in N2 score was 4.348 points. Comparing that mean change to the standard deviation of baseline N2 score (15.444) indicates that the average change in the first year is equivalent to 28.2% of a standard deviation in the baseline measure of moral judgment; the four-year increase in N2 score

of 8.351 was twice as large, 54.2% of a standard deviation in the baseline measure for the four-year sample. (For both samples, the standard deviations for the N2 outcome score were virtually identical to that of the baseline score, meaning that the relationship between the mean change and dispersion of the outcome is also virtually identical, whether referring to the baseline or the outcome measures.)

The size of the changes over both the first year and the first four years of college provide an important lens through which any regression coefficients for independent variables predicting changes in N2 score should be seen. Effects that are statistically significant but seem too small to have practical significance should be considered relative to the size of the average change for the sample. Likewise, large effects may be even more significant than they initially appear. For example, in this study, all model coefficients are reported in standardized coefficients, meaning that each coefficient can be interpreted as the proportion of a standard deviation that the model predicts the moral judgment outcome to increase or decrease given a one-unit change in the independent variable. Therefore a coefficient of 0.070 in a model of moral judgment at the end of the first year refers to 7% of a standard deviation in the first-year outcome measure, which seems like a small amount, particularly when determining how a college or university should employ limited resources. However, that same coefficient of .070 refers to nearly 25% of the average increase in moral judgment during that first year. A program that predicts an increase in development equivalent to one-quarter of the average increase is suddenly worth much more in a discussion of those resources.

The small average change also has implications for the educators' expectations for any individual intervention aimed at encouraging the development of students' moral judgment. If the average change for students over the course of four years is as modest as these numbers

suggest, an academic or student affairs educator should not expect an individual class or program to have a large impact on moral judgment.

Individual and Institutional Differences in Moral Judgment

This study employed a multilevel approach to modeling the development of moral judgment during college. I determined this to be the appropriate approach based first on conceptual reasons, with statistical reasons serving to reinforce the appropriateness of this method. The findings outlined in Chapter IV (Table 4.6) indicate that 22.0% of the variance in N2 scores at the end of the first year and 18.7% at the end of the fourth year can be attributed to between-school differences; this means that approximately one-fifth of the differences in students' moral judgment at each point can be attributed to institutional differences rather than to their individual characteristics and experiences within their institution. I will discuss the implications of specific institutional-level variables later in this chapter; however, this approximately 20% institutional-level variance has implications for research beyond the effects of specific variables investigated in this study.

These findings show that research that does not consider the influence of the institution on students' development of moral judgment is ignoring an important component of that development. There are statistical reasons to argue for the use (or avoidance) of multilevel analysis of student outcomes (such as, but not limited to, HLM), and those are certainly important. However, the results of this study also support the use of these methods for conceptual reasons. It is also not enough to simply use a multilevel modeling technique to partition variance into within-college and between-college components, and then use only individual-level independent variables to model moral judgment; this ignores the institutional-level variance component. Instead, it is essential for researchers to actively examine the

institutional factors that influence students' development of moral judgment. To focus only on individual-level variables while disregarding institutional variables ignores potentially powerful forces that are known to influence moral development in significant ways, either positively or, perhaps more importantly, negatively affecting moral development.

Unfortunately, it is all too common in the extant research on moral judgment development to ignore institutional level differences. Of the 55 studies that I reviewed in Chapter II that investigate how college experiences influence the development of moral judgment using the DIT or DIT2, only 14 involved students at multiple institutions, and six of those reported recent data from the WNS. Half of these multi-institutional studies examine the impact of institutional characteristics, focusing almost exclusively institutional type as defined by Carnegie classification (Padgett, Johnson, & Pascarella, 2012; Mayhew, 2012; Martin, Hevel, Asel, & Pascarella, 2011) or religious affiliation (Good & Cartwright, 1998; Elm, Kennedy, & Lawton, 2001; Maeda, Thoma, & Bebeau, 2009; Traiser & Eighmy, 2011). Only Maeda, Thoma, & Bebeau focused on other factors: they also examined the effect of a conservative orientation among the student body and geographic location in the southern United States. Taken together, these studies demonstrate that most current research on moral judgment development in college has ignored institutional features, and the few that have considered them used rather blunt measures of institutional categories.

The individual-level variance found here at the end of both the first and fourth year of college also has implications for educators and administrators, who should also be aware of the potential influence that characteristics of the institution can have on students. These include understanding the ways that classes, programs, and other individual-level experiences can influence students' moral judgment, and using this information to plan curricula and allocate

resources accordingly. Beyond this, it is also important for educators and administrators at colleges and universities to understand how the characteristics of their institutions (including the faculty culture and peer environment (as detailed by Terenzini and Reason, 2005) may also be influencing development. This would allow these educators to target aspects of the college or university for change that could positively affect student moral development. Further, if educators know that their institution exhibits specific characteristics that research suggests inhibit development, they can target those classes and programs or broader cultural norms in ways that they hope will counteract those negative effects. This, however, is not possible without an understanding that – and how – institutions affect moral development.

Despite these recommendations that institutional effects on moral judgment be considered by both researchers and educators, it is important to acknowledge a limitation in the results discussed above. The outcomes for this study were unadjusted end-of-the-first-year and end-of-the-fourth-year N2 scores, and controls for the baseline N2 scores were included as independent variables when estimating models presented here. This is in contrast to an outcome that represents the amount of change in N2 score between these two points. (This decision is the common approach in the literature on changes in moral judgment during college, as well other college student outcomes; it is discussed in more detail in Chapter V. Accordingly, the interclass correlations (ICC) reported in Table 4.6, refer to the amount of variance in these unadjusted outcomes rather than in the amount of change between the baseline and outcome assessments. This means that when using the ICCs to determine the proportion of variance in moral judgment that is accounted for at the institutional (between-college) level (22% in the first year and 18.7% in the fourth year), the statistic does not account for the difference in student baseline measures by college. Since baseline N2 and outcome scores are highly correlated, it is possible that any

large differences in baseline N2 score by college account for some or all of the ICC. It is very important, then, that the ICC not be interpreted as describing institutional differences in the amount that students' moral judgment *changes* during college.

To better understand the institutional differences in change, I examined the institutional differences in baseline and outcome N2 score. Appendix B presents the mean baseline and outcome scores for each institution in each sample, along with the results of a *t*-test to determine the statistical significance of the two in each institutional sample. The data in this table were organized by grouping the 44 colleges into quartiles based on the average baseline N2 measure for the first year sample (with the first quartile including those insitutions with the least difference). Within each quartile, the colleges are arranged by the size of the difference between that baseline measure and outcome for the first-year sample (differences that are not statistically significant at least 95% after conducting the series of *t*-tests are considered to be zero). Instution labeles are included to allow comparison with other tables, and other studies that used the WNS institution labels.

Although these data do not provide the proportion of variance in the change in moral judgment that can be attributed to college differences like an ICC would (that is beyond the scope of this study), they suggest the existence of institutional differences, even though most of these differences are small. The institutional differences between the average baseline N2 score and the average N2 score at the end of the first year range from 0.412 to 16.63; however, the interquartile range is only 2.901. For the fourth-year sample, the differences range from 1.642 to 20.644, and the interquartile range of the differences is 3.262. Beyond these differences in the difference, not all institutions had a difference that was statistically significant between the baseline and the outcome for one of both of the samples. For the first-year sample, statistically

significant differences between the mean baseline and the outcome N2 scores were not found for 14 of the 44 colleges; four colleges do not have statistically significant differences in the fourth-year sample. These results suggest that, although the differences are small, change in students' scores differ depending on the college they attend, a conclusion which cannot be drawn from the ICCs presented in Chapter IV. How much of the total variance in changes in moral judgment that is accounted for by the institutional differences, however, remains unknown.

Precollege Characteristics

To understand the effect of college experiences on outcomes, it is necessary to consider students' precollege characteristics. This does two important things to improve the conclusions drawn from research on student outcomes. First, it allows researchers to account for differences in students prior to starting college that affect the differences in students at the end of the time period being studied; otherwise results may just reflect precollege differences rather than effects associated with college experiences. Second, considering precollege characteristics provides perspective for the findings about the effects of college experiences; researchers can examine what proportion of students' outcomes are attributable to their college experiences and how much is attributable to the characteristics in place when they started college. The first of these is common in research; studies frequently include precollege characteristics as control variables when examining student outcomes. The second, however, is much rarer.

In this study, as outlined in Chapters III and IV, I began the process of modeling students' moral judgment at the end of the first and fourth years of college by first estimating models with precollege characteristics as only independent variables. These demonstrated that students' baseline moral judgment score, academic ability, and being female were significant positive predictors of moral judgment at both the end of the first and fourth years of college.

Further, reporting a neutral political orientation was a significant negative predictor at the end of the first year, and belonging to one of three racial/ethnic categories (African American, Latino/a, and Asian and Pacific Islanders) was a significant negative predictor of moral judgment at the end of the first year. These results are important not just for the implications of the variables that displayed positive or negative effects, but also for the size of those effects and the proportion of the variance they explain in the moral judgment outcome.

This group of precollege characteristics (baseline score, academic ability, gender political orientation, and racial/ethnic identification) account for a very large proportion of variance in moral judgment at the end of both the first and fourth years: r^2 values for the models are .457 and .344, respectively. This means that these few precollege characteristics that students brought with them when they started college accounts for almost half of the variance in moral judgment after the end of the first year of college, and for more than one-third of the variance by the end of fourth year. The importance of precollege characteristics in moral judgment at these later points of time can especially be seen by the large effect of students' baseline N2 score. For the first-year sample, the standardized coefficient for the baseline score was .555, and the standardized coefficient for the fourth-year sample was .505; a standard deviation increase in baseline score predicted more than half of a standard deviation increase in both outcome scores. It is not surprising, especially for the first-year sample, that the baseline score has such a large influence on the outcome score. Although a student's attitudes or beliefs about specific moral issues may change dramatically in a year of college, structural development typically unfolds more slowly, and it is unrealistic to expect a student's experiences to trigger a major substantive structural change in thinking in a few months (e.g., Rest, 1986; Barber, King, & Baxter Magolda, 2013). It is, perhaps, more surprising that the baseline score still has such a large influence after four years

of college; however, despite the high values of the coefficients for baseline N2 scores for both samples, they are still far from perfectly correlated, especially when other characteristics are taken into account. So while students' baseline scores continue to exert a large influence on N2 scores through the college years, other characteristics and experiences still account for almost half of the variance in moral judgment after four years of college.

In addition to the large effect of baseline N2 score on students' outcomes for both the first- and fourth-year samples, precollege academic ability also had a large independent effect on the outcomes. A one-standard deviation difference in precollege academic ability accounts for just more than one-fifth of a standard deviation change in N2 score for the first-year sample and just less than a one-fifth of a standard deviation change for the fourth-year sample. Even when accounting for baseline N2 score, with which it is correlated, students' precollege academic ability (as operationalized by scores on college entrance exams) is predictive of moral judgment at the end of the first and fourth years of college.

An assessment of students' academic ability is rarely addressed in the previous literature on the development of moral judgment on college. Of the 55 studies analyzed and discussed in Chapter II, only nine included any sort of measure of students' academic ability as an independent variable. Of these, six used the same WNS data as this study (Martin, Hevel, Asel, & Pascarella, 2011; Mayhew, 2012; Mayhew, Seifert, & Pascarella, 2010; Mayhew, Seifert, & Pascarella, 2012; Mayhew, Siefert, Pascarella, Nelson Laird, & Blaich, 2012; Padgett, Johnson, & Pascarella, 2012), and these showed similarly large positive effects of college entrance exam scores on students' N2 score at the end of the first year of college. The three additional studies operationalized academic ability with students' college grade-point averages (GPAs); Traiser and

Eighmy (2011) found a positive relationship with moral judgment, but Brown-Liburd (2011) and Fleming, Romanus, & Lightner (2009) found no significant relationship between the two.

This small number of studies, many of which are very similar, make it difficult to make a generalized claims about the relationship between academic ability the development of moral judgment. The results concerning GPA are inconsistent and arise from only three studies, providing very little evidence from which to draw conclusions. The larger number of studies (seven, including this study), that operationalize academic ability using college entrance test scores all rely on the same data from the same sample. The colleges and universities that comprise these studies were not chosen at random, and they may vary from the population of colleges in ways that would change this relationship.

Further, that these studies operationalize the academic ability using students' scores on college entrance exams (such as the ACT and SAT), meaning that a more precise description of these studies is to say they show a positive relationship between those test scores and N2 score. The pros and cons of using the ACT or the ACT as a proxy of academic ability or preparedness have been well examined elsewhere; to do so here is beyond the scope of this study. What is clear is that regardless of what the tests measure (and what they do not), the scores are predictive of moral judgment. Further, the relationship between college entrance exam scores and N2 outcome scores persist even when controlling for the baseline N2 score; these mean that the relationship is not just between precollege academic ability and moral judgment in general, but specifically between precollege academic ability and the *change* in moral judgment during the first year and the first four years of college. If the relationship between the two measures was simply because they are both measuring students' verbal ability or test-taking prowess, it would be expected for that relationship also to exist with the baseline measure. In fact, since the

baseline N2 score is more temporally proximal to the college entrance exam, one would expect the relationship to be stronger with the earlier measure of moral judgment. Instead, there is an independent relationship between the precollege academic ability and change in moral judgment, beyond the relationship with the baseline assessment.

It might also be true that in the relationship between precollege academic ability and development of moral judgment, precollege academic ability is functioning as a proxy for unmeasured institutional differences based on a college having a student body with a higher levels of academic ability and preparedness. In their model of student outcomes in college, Terenzini and Reason (2005) point to the importance of the peer environment in students' outcomes. Beyond these peer effects, more selective colleges (which employ higher scores on college entrance exams as a gatekeeping feature of the admissions process) may be better at creating the kind of experiences that encourage the development of moral judgment. If either of these institutional explanations for the relationship were true, however, one would expect to find evidence of that supports these explanations elsewhere in the estimated models. If even a portion of the relationship was because of these institutional factors, there would be a significant relationship with the college's average precollege academic ability. In the models estimated for this study, that relationship is not statistically significant once a student's individual precollege ability is accounted for in the models.

Although the generalizability is limited and the mechanism is unclear, the existing evidence does consistently point to a positive relationship between students' scores on college entrance exams and the development of moral judgment, at least among students in the WNS sample. Although it is true that this does not necessarily mean that this relationship is the same in the larger population, it is the only available evidence about this relationship and it should not be

ignored. Short of additional evidence that contradicts these findings, researchers and educators should at least consider that students' academic abilities may play a larger role than expected in development of moral judgment. A more thorough examination of these effects is beyond the scope of this study, but that examination is an essential direction for future research.

The results of the effects of political orientation also tell a positive story about students' development during college. For the first-year sample, students who report being moderately or extremely conservative or liberal politically show statistically significantly higher levels of moral judgment when compared to students who report being in the political center. However, that significant relationship between political orientation and the development of moral judgment does not persist for the fourth-year sample. These results suggest that for the students in this sample, time spent in college decreases rather than increases the impact of political orientation on moral development.

The negative impact of gender on changes in moral judgment among male students should be concerning for researchers and educators. Carol Gilligan (1977, 1982) and others have criticized Kohlberg's justice-oriented approach as privileging an inherently male perspective of morality over what Gilligan described as a care-oriented female perspective. In Chapter II, I reported research demonstrating that when gender differences are found, women tend to score at higher levels on the Defining Issues Test than their male counterparts (e.g., Thoma, 1986; King & Mayhew, 2006), and the findings of this study are consistent with those. In this study, there were large negative effects on moral judgment for male students in both the first-year (b=-0.172; p<.001) and the fourth-year samples (b=-0.211; b<.001). This indicates male students demonstrated lower scores than did women at both the end of the first and fourth years of

college, even when controlling for their lower average baseline DIT2 scores and other experiences included in these models.

Also distressing is the appearance of negative effects for racial and ethnic minority students. After controlling for other precollege characteristics, there is no significant relationship between race and moral judgment for any racial or ethnic minority group at the end of the first year of college. However, for the fourth year sample, significant, and negative effects are present for students who are Latino/a (b=-0.189; b<.05), Asian or Pacific Islander (b=-0.176; b<.01), African American (b=-0.122; b<.010). (There were not enough Native American students in the sample to determine a relationship with moral judgment for this group.) This indicates that for this sample, something happened *after the first year of college* that led to negative effects for minority students. Even after accounting for different baseline levels of moral judgment and different precollege measures of academic ability, male students and students of color fell farther behind their female and White peers over the course of their time in college. This gender and race gap in the development of moral judgment that persists and increases during college should be a significant concern for colleges and universities, as well as for researchers in this field.

For researchers, these variables are all too often framed as control variables without thorough discussion. Additional research is essential to understand why these gaps occurred during college. Prior research (e.g, Kuh, 2008; Sweat, Jones, Han, & Wolfgram, 2013; Harper, 2009), has indicated that under-represented minority students are less likely to experiences what Kuh calls high-impact experiences, such as undergraduate research experiences, collaborative assignments, and service- and community-based learning, which are predictive of learning and college success. Further, Harper argued that institutions do not approach student engagement in these high-impact experiences in race-conscious ways, which he posits would lead to more

equitable participation in these experiences and their benefits. It is certainly possible that this discrepancy in experiences (some, but not most of which are examined in this study) exist among students based on race and gender and that this contributes to the increase in race and gender gaps in moral judgment during four years of college. If male students and students of color have fewer opportunities for the kind of cognitive activities that have been empirically found to encourage moral development, this is unacceptable. Educators have an imperative to investigate the practices on their own campus in order to guarantee that all students have access to the same opportunities, regardless of race or gender.

It may also be the case that male students and students of color have access to the same opportunities in college, but that they respond differently to them than their female and White counterparts, such that the same experiences are less likely to promote their development. Researchers would be well advised to consider these precollege characteristics as important and worthy of study in their own right rather than simply relegating them to the status of a variable to be statistically controlled for when examining the programs and educational efforts that receive the bulk of the attention in the literature.

Taken together, the results of the analyses of the effects of precollege characteristics on the development of moral judgment during college show that student characteristics at the time they enter college is a major indicator of what their moral judgment will be at the end of both their first and fourth years of college. The r^2 statistics from both models indicate that these handful of variables – baseline moral judgment, academic ability, gender, race, and political orientation – account for more than half of the variance in students' moral judgment at the end of both the first and fourth years of college. Additionally, the standardized effects of these precollege characteristics are quite large, even when controlling for baseline N2 score; and, with

the exception of political orientation, these effects see little decline over the course of students' time in college. Neither the proportion of explained variance nor size of standardized effects decrease significantly when precollege characteristics are included in the developmental, college impact, and integrated frameworks discussed later in this chapter; indeed, these precollege variables consistently indicate a much larger role in students' moral judgment at the end of college than that played by their college experiences examined in this study.

The existing research addresses the effects of precollege characteristics in incomplete and sometimes seemingly haphazard ways. Although college outcomes research traditionally employs baseline assessments, race/ethnicity, and gender as controls, these are not often the focus of research questions and systematic analysis and theory building when considering them in the context of influencing change during college. This is also true of the research on the development of moral judgment. Further, despite its large effect on moral judgment, students' academic ability is absent from all but nine studies (Cummings, Dyas, Maddux, & Kochman, 2001; Drake, Griffin, Kirkman, & Swann, 2005; Elm, Kennedy, & Lawton, 2001; Herrington & Weaver, 2007; Jeffrey, 1993; Kaplan, 2006; Livingstone, Derryberry, King, & Vendetti, 2006; McNeel, Abou-Zeid, Essenburg, Smith, Danforth, & Weaver, 1996; Snodgrass & Behling, 1996). In this study, I operationalized precollege academic ability as a students' performance on the SAT, ACT, or comparable college admissions test, acknowledging the imperfection of these measures. These data were part of the students' institutional records provided for this study; this is a potential source of data for future researchers, despite the limitations.

Similarly, despite political orientation being assessed as part of the DIT2, the characteristics was reported in only 12 previously reviewed studies (Drake, et al,. 2005; Grunwald & Mayhew, 2008; Klimek & Wenell, 2011; Maeda, Thoma, & Bebeau, 2009;

Mayhew, 2012; Mayhew & Engberg, 2010; Mayhew & King, 2008; Mayhew, Seifert, & Pascarella, 2010; Mayhew, Seifert, & Pascarella, 2012; Mayhew, Seifert, Pascarella, Nelson Laird, & Blaich, 2012; Lies, Bock, Brandenberger, & Trozzolo, 2012); notably, two-thirds of these were written by Mayhew and colleagues. These are large holes in the existing research; addressing questions related to these precollege characteristics will provide researchers a large opportunity to better understand the development of all students during college, contributing both in building theory and aiding institutions in employing limited resources in ways most likely to aid all students.

Examining Moral Judgment with a Developmental Framework

In the extant literature, few published studies have utilized a developmental framework when examining the effect of college experiences on students' moral judgment. In the studies described in Chapter II, only four utilized a solely developmental framework, compared to 43 that utilized a solely college impact framework. Only twelve studies include independent variables consistent with a developmental framework, with four of those using only such variables. It is easy to see that developmentally-focused variables and developmental frameworks are greatly under-represented in the existing literature the development of moral judgment in college. (These studies are discussed in greater detail in the final section of Chapter II.) Therefore, one of the goals of this study was to bring an intentional and wider-ranging examination of developmentally-focused variables to the study of moral judgment in college. The results of the estimated models using this developmental framework support the value of a variable that represent the kinds of cognitive activities that theory suggests would encourage or inhibit development; this is an area rich in potential for future research.

The results from the models employing a developmental framework indicated that several developmental variables were estimated to have effects that were positive (classroom encounters with disequilibrium and experiences with integrative learning) or negative (negative interactions with diverse others and experiences with higher order learning) on students' moral judgment in one or both of the samples. These findings support the position that variables contribute to the development of moral judgment in college, and these variables should be addressed in the research.

Two of these developmental variables (self-initiated encounters with disequilibrium and experiences with integrative learning) yielded statistically significant positive relationships with moral judgment in both samples, and another variable (classroom encounters with disequilibrium) was positively related to moral judgment in the first-year sample (see Tables 4.6 and 4.7). Not only are these positive effects statistically significant, some of them are relatively large, especially when compared to effects of variables in other frameworks. For example, even when controlling for students' precollege characteristics, the effect of integrative learning in the first-year sample (b=0.087, p<.001) is the equivalent of more than one-quarter of the mean change in moral judgment during the first year of college. Similarly, the effect of self-initiated encounters with disequilibrium in the fourth-year sample (b=0.099, p<.001) is the equivalent of almost one-fifth of the mean change in moral judgment that students displayed between the beginning of their first year of college and the end of their fourth year. That these variables were consistently significant across models and the size of their effects speak to the potential of a developmental framework to investigate the development of moral judgment during college. Efforts to understand how and when students encounter disequilibrium and how it can be used more effectively to promote moral are an important direction for future research.

Validity of Developmental Variables

In the developmental framework, I employed five factors; I created one for this study, and the other four were created by other researchers and have been previously used in the WNS (Pascarella & Colleagues, 2008). All of these factors had strong psychometric properties. Cronbach's alpha statistics, for example, ranged from .686 to .885, indicating strong internal consistency among individual items included in each factor. Further, factor loadings for all but one variable across the five factors are greater than .600 and most factor loadings are greater than .700, indicating strong correlations between factors scores and students' responses to individual items comprising that factor. The factor loadings were highest for the factor that I created specifically for this study, students' classroom encounters with disequilibrium. In addition to strong reliability measures, these variables also have strong face validity for the concepts they are intended to operationalize, and are supported by expert review (Patricia King). In each case, these were the best available options in the WNS survey data; a survey designed specifically to gather data on developmentally-focused variables might well have included different items.

Concerns about validity of the variables in the developmental framework are illustrated by the Classroom Encounters with Disequilibrium factor. The construct validity of this factor was called into question due to the negative relationship with moral judgment found in the fourth-year sample. Developmental theory suggests that, with appropriate supports, encounters with disequilibrium can encourage moral development. Since the factor representing Self-Initiated Encounters with Disequilibrium is included in the model (with a positive effect), it may also be that as students progress through college, they respond less positively to encounters with disequilibrium that they see as being placed on them by instructors rather than those that they seek out themselves. This may also simply be a case of Type II error, in which a model estimates

statistical significance for a sample that doesn't exist in the population; a confidence level can never be 100 percent.

This finding of a negative relationship between this factor and moral judgment in the fourth-year sample is could be explained by another developmentally-related condition: these students may have lacked the kind of support they needed to productively deal with the disequilibrium. Students at earlier levels of development may be particularly at risk from reacting to disequilibrium by withdrawing from dissonant situations rather than by accommodating the new dissonant knowledge or experience into their own worldview. King, Baxter Magolda, and Massé (2011) provide examples of students who rely on external meaning-making systems withdrawing from uncomfortable interactions with people of different races rather than reflecting upon them or their own reactions. Beyond withdrawing, some students may use their experiences with dissonance to further crystalize their less-developed ways of making meaning. Boyle-Baise (1998) and Boyle-Baise and Sleeter (1998) describe students reacting this way when working with students from impoverished families in a service-learning project. When presented with information about these students that conflicted with their previously held ideas, some of the college students in these studies constructed ways for them to fit anyway, such as saying that the children must have really bad parents to be the disadvantaged situation. Like the students studied by King, Baxter Magolda, and Massé, the external meaning-making of these students would not allow them to construct new ideas one their own.

This is why support is so important when students encounter disequilibrium. If the professors and other instructors are providing disequilibrium without effective supports, a variable that demonstrates a negative relationship between those exposures to disequilibrium in the classroom and the development of moral judgment may in fact be demonstrating construct

validity; that is, development is more likely to occur if the nature of the disequilibrium is appropriate for the student's level of developmental readiness. Additional research is necessary to understand the most appropriate way to gather information about students' encounters with disequilibrium and to fully understand how the process of encountering that disequilibrium in the classroom.

Even if some of the variables are not the best measures of the constructs that I sought to assess, they yielded effects that were significant both practically and statistically. Measurement error in regression models biases coefficients toward zero (Fuller, 2009), thus it is reasonable to expect that measurement error in the variable would lead to these models producing conservative estimates of the effect of these cognitive experiences that are statistically significant in these models. If that is the case and these effects are even larger than they appear, it is even more important to consider the use of a developmental framework and developmentally-focused variables when researching moral development.

Examining Moral Judgment with a College Impact Framework

In addition to the developmental framework discussed in the previous section, a second goal of this study was to apply a broad college impact framework to the examination of the development of moral judgment during college. For this study, that meant not conceiving of college experiences as simply participation in specific programs, courses, and activities, as is common in the previous research on moral judgment using a college impact framework; rather, I employed Terenzini and Reason's (2005) Comprehensive Model of Influences on Student Learning and Persistence. This model follows in the tradition of college impact models as discussed in Chapter II, focusing on participation in formal and informal activities and exposure to the characteristics of institutions, faculty, and other students, rather than a developmental

framework's explicit focus on the cognitive tasks and activities in which students engage that predict development. What separates this model from other college impact models is its comprehensive nature, as it encompasses a broad range of organizational factors, individual student experiences, and features of the peer environment in a way that is not done by other models and has rarely been incorporated in moral judgment research. With this study, I have aimed to develop a more comprehensive view of the college impact factors affecting moral judgment than exists in the extant literature on the development of moral judgment in college.

Institutional Characteristics

Structural-demographic characteristics are most often included in student outcomes research (such as institutional size, sector, and selectivity); this line of research has consistently indicated that these factors have little predictive power on student impacts (e.g., Dey, Hurtado, Rhee, Inkelas, Wimsatt, & Guan, 1997; Milem & Berger, 2000; Pascarella & Terenzini, 2005). Accordingly, I included several additional institutional variables to these standard factors. The goal for these variables is to better represent the following dimensions of the Terenzini and Reason's model: internal structures, policies, and practices; academic and co-curricular programs, policies, and practices; faculty culture; and the peer environment. Incorporating data from the WNS survey and the Integrated Postsecondary Education Data System (IPEDS), I included in my analyses variables such as undergraduate to graduate student ratio; faculty to student teaching assistant ratio; student perceptions of faculty interest in teaching and student development; and percent of students who had done or who planned to do community service during college. By broadening the focus of institutional variables beyond structural-demographic characteristics, I was able to investigate institutional and cultural factors that are more malleable and actionable.

Unfortunately, in terms of identifying specific aspects of institutional organization and culture that encourage or inhibit moral development, the findings for these factors yielded few statistically significant results. Unsurprisingly (given prior research), none of the structural-demographic variables were significant predictors of moral judgment at the end of either the first or fourth years of college. However, institutional variables representing organizational and cultural characteristics also failed to predict moral judgment. Of the 15 institutional variables included in each of the models employing a college impact framework, only two were significant predictors when controlling for other variables in the college impact model: the average precollege academic ability of the students at the institution (a factor of the peer environment) and students' perceptions of faculty interest in student learning and development (a factor of faculty culture). The positive effect of the institution's average academic ability was that a standard deviation increase in average academic ability account for an increase of 0.327 standard deviation of moral judgment at the end of the first year and .350 at the end of the fourth year. The effect of perceived faculty interest in students is considerably smaller for both samples, but it indicates that a higher level of faculty interest in students' learning and development predicts higher moral judgment at the end of both the first and fourth years of college. However, neither of these effects persisted in the final models for either sample after controlling for individual students' precollege characteristics.

This suggests that while these two institutional characteristics influence moral judgment, their influence is not distinct from that of student characteristics when they enter college. For example, as discussed earlier in this chapter, the precollege academic ability of individual students has a large significant effect on their moral judgment at the end of both one and four years of college. Since most students' own academic ability will be close to their average of their

institution, this individual effect may overwhelm the institutional effect once both are included in the same model. Further, without the inclusion of precollege characteristics, the significance of the two institutional characteristics could be an effect of selection. College impact models postulate that one way students' precollege characteristics affect their college outcomes is that they contribute to the college choice students make and to their access to different colleges. In this way, precollege characteristics lead to students attending different colleges where they will be exposed to different environments and experiences. It is possible that students whose precollege characteristics predict higher levels of moral judgment at the end of the first and fourth years of college are more likely to attend colleges with higher levels of both institutional characteristics in the model. In that case, the precollege characteristics could predict both the institutional variables (by influencing the college a student attends) and the outcome, explaining the lack of significance of institutional variables in the final college impact models. In a later section of this chapter, I explore these possible selection effects more in more depth.

Although it is not surprising that the structural-demographic variables of an institution would not have a statistically significant effect on the development of moral judgment, it is more surprising – and more disappointing – that the other institutional-level variables were not significant. The lack of significance of any institutional characteristics in the final models of either sample raises questions about the efficacy of measuring the institutional aspects of college that affect the development of moral judgment through the current common methods and data used to examine these effects. Constrained by the particulars of the WNS data set, I attempted to utilize and construct variables that were reasonable proxies for the constructs included in Terenzini and Reason's (2005) model. That these variables demonstrated no predictive power on the development of moral judgment in the analyses for this study is an intriguing finding. The

data that I used from this study were all part of either the institutionally-reported IPEDS data or were taken from the student surveys and administrative data comprising the WNS quantitative dataset; these data are similar to those that are available to most researchers who are conducting student outcomes research with large, multi-institutional data sets. The IPEDS data are publicly available, and the WNS surveys rely primarily on items from the National Study of Student Engagement (NSSE) survey and Cooperative Institutional Research Program (CIRP) surveys, both widely used in student outcomes research. These results – coupled with Terenzini and Reason's postulation that more detailed and organizationally focused data are needed to more fully examine institutional effects on student outcomes – suggest that the most commonly available and most commonly used institutional variables for student outcomes research are not effective in predicting moral judgment.

This provides a serious challenge to researchers; it suggests that new methods, data sources, and variables are necessary to effectively examine the institutional effects on moral judgment in particular and, potentially, student outcomes in general. It also presents an opportunity to researchers: developing these methods would contribute to significant and necessary broadening of student outcomes and development within institutional contexts. Although such research may be costly and is resource- and time-intensive work, it is needed in order to better understand and measure these factors for subsequent quantitative research on student learning outcomes.

Individual Student Experiences

Beyond institutional characteristics, the analyses employing a college impact framework also typically include variables representing students' individual experiences. Terenzini and Reason (2005) identify three dimensions of individual experiences (curricular, classroom, and

out-of-class) that influence students' outcomes during college. In moral judgment research, these categories most commonly refer to major and participation in a specific course or formal program, such as diversity courses or volunteer programs. A smaller portion of the research focuses on participation in other types of cocurricular activities, such as varsity athletics or Greek life. I included a range of similar variables in my study, and discuss each separately below.

The large influence of majors. Although there are no institutional characteristics and few individual experiences that demonstrated statistically significant positive or negative effects on students' development of moral judgment during college, students' majors did have significant and comparatively large effects on moral judgment at the end of both the first and fourth year of college. Even after controlling for observed precollege characteristics, in the first-year sample all major categories (biology, business, education, physical sciences, professional, social sciences, engineering, "other," and undecided) demonstrated statistically significant negative effects when compared to humanities majors, ranging from -0.081 for "other" majors and -0.114 for biology to -0.256 for business. In the first-year sample, there are few majors with significant effects (when compared to humanities majors), but the negative effects that do persist are quite large: other (b=-0.180), social sciences (b=-0.214), business (b=-0.239), and engineering (b=-0.376). It is important to reiterate that these numbers are not simply demonstrating that students in these majors have lower levels of moral judgment at the end of the first year than do humanities majors. Indeed, that these effects persist even when controlling for baseline measurement indicates that these gaps are attributable to differences that develop during the first year college. Regardless of whether students in these majors start college with higher or lower levels of moral judgment, these models indicate that even when controlling for other

characteristics and experiences, students in non-humanities majors exhibited smaller increases in moral judgment. These effects are large, and I suggest three potential reasons for them.

Major and precollege characteristics. First, students who enter different majors may be different in ways that are unobserved in this study. In a regression model, unobserved differences among students would bias the effects of major in these estimated models if those unobserved differences are correlated with both the independent variable and the dependent variable. In this case, that would mean that some precollege characteristics that would make a student more likely to choose a specific major would also make them less likely to develop moral judgment during college. For example, a student who values highly paid employment might be more likely to major in higher paying majors such as engineering or business; she might also be less likely to develop postconventional moral schema that relies on standards of justice (i.e., where the rights and responsibilities of societal interactions are justly distributed).

This possibility of unobserved precollege characteristics contributing to the effects of majors is supported by the alleviation of some of the effects between the first and fourth years of college. Whatever unobserved precollege characteristics biasing the effects would be more proximal to the end of the first year than the end of the fourth year. Many researchers have demonstrated that students' attitudes, beliefs, and values change during college, and so it makes sense that as these change, the influence on moral schema of students' initial attitudes could diminish as well.

It is important to note that unobserved characteristics are not necessarily unobservable. In this study, I used the precollege variables that are most commonly used in the college outcomes literature (baseline outcome measure, precollege academic ability, gender, race/ethnicity, political orientation, age, and international or domestic status). Most common national student

surveys, including the data used in this study, ask students about their attitudes, beliefs, and values when beginning college. This is beyond the scope of this study, but understanding how these attitudes, beliefs, and values contribute to students choosing their majors could be a useful step in understanding how precollege factors and majors affect students' later outcomes, including the development of moral judgment.

Major and other college experiences. Second, students in different majors may participate in different experiences in college in unobserved ways. Due to myriad reasons, students in different majors may be more or less likely to engage college activities and other experiences that encourage or discourage moral development in systematic ways. For example, this study does not consider the effect of students' internships and other out-of-class work experiences related to their field of study. These experiences could encourage or inhibit the development of moral judgment. Due to the culture of some majors and the job markets in some fields, these work experiences are not distributed randomly across majors; for example, students in majors such as engineering, business, and communication are more likely to participate in internships and other pre-professional work experiences. One of the goals of most internship experiences is for a student to become socialized into her chosen professional field, and students are usually encouraged to interact with more senior workers; these interactions could include exposure to the way other workers approach moral and ethical dilemmas in their field. This kind of experience could result in students seeing the complicated nature of professional ethics and understanding the nuanced ways that professionals in their field grapple with competing interests and attempt to balance them in ethical ways. Based on developmental theory, this type of activity would be expected to stimulate more advanced moral judgment in students. However, students on an internship might instead interact with professionals around moral and ethical issues in

ways that encourage them to get around rules and regulations, to weigh company profits over larger issues of justice, or follow rules and regulations with understanding that these professionals don't agree with them and only follow them because to do so is demanded. This kind activity would be expected to have the potential to inhibit more advanced moral judgment in students.

If these work experiences affect the development of moral judgment and are distributed among students in this study differentially based on their major, then their effects would appear in this study as effects of majors. If students in other majors participating in internships at the same rate and those internships have the effects on the development of moral judgment, then the distribution of that effect would no longer be based on student major, meaning it would lessen the size of the effect of those majors on the outcome.

This example of a potential effect of internships manifesting itself as an effect of students' majors is not meant to explain away the large effects of major in this study. Rather, it provides an example of how students in different majors could have different experiences in ways that are not based on their classes and curricula. Faculty and other educators are often focused on classroom practices and curricula, including when they consider ways to educate for moral judgment. If students in some majors are exposed to out-of-class experiences that encourage or inhibit moral development more or less than in other majors, faculty must take that into consideration when planning educational programs around moral or professional issues within their department. These experiences may provide valuable ways to supplement the moral education happening in the classroom, but they may also provide obstacles for students that faculty should address in systematic ways. Thus, understanding the differential experiences of students based on their major is not only valuable for researchers trying to better understand the

ways students develop during college, but also for educators in understanding ways to better educate different types of students.

Potential causal effects of majors. These first two ways discussed here that student major might predict moral judgment rely on selection bias for an explanation. In the first, students' unobserved precollege characteristics contribute to sorting them into majors. In the second, students in different majors sort themselves into college experiences in unobserved ways. A third way that student major can predict differences in the development of moral judgment might have nothing to do with selection bias; instead, there may well be a causal relationship between the educational practices in some majors and the degree of development of moral judgment for students in those majors.

For example, in this study, engineering students showed the largest negative effects of any major at the end of the fourth year. Previous research has shown that engineering students primarily learn about professional ethics in ways that emphasize black-and-white discussions of laws, regulations, and professional codes of conduct (e.g., Finelli, Holsapple, Ra, Bielby, Burt, Carpenter, Harding, & Sutkus, 2012; Holsapple, Carpenter, Sutkus, Finelli, & Harding, 2012). This is different than the more nuanced philosophical approach one might find in the humanities or the social justice-oriented approach more common in education. Engineering is a field in which the education surrounding ethics and moral development at the undergraduate level is one that privileges conventional moral schema rather than encouraging the post-conventional moral reasoning indicative of higher N2 scores. In this case, evidence would suggest that at least part of the difference in the development of moral judgment between engineering students and their peers in humanities or education majors is being caused by the differences in the ways that moral and ethical dilemmas are defined and discussed within the curriculum of each major. Rather than

students characteristics being the difference in these outcomes (meaning that selection is the root cause of the differences), this kind of educational difference in pedagogy or focus) would be a causal effect of the educational approaches.

It is, of course, impossible in the current study to identify a causal effect of these class- and curriculum-based differences on the development of moral judgment between different majors, especially since the larger WNS study was not designed to facilitate this type of causal analysis. Isolating these causal effects would be difficult, and would take a different approach than the kind of large-scale surveying common in college impact research. It is not possible to randomly assign students to majors, and even with quasi-experimental methods, such as propensity score matching based on likelihood of choosing a particular major or a regression discontinuity design based on entrance exams or program requirements, it would not be possible to distinguish between the effects of class differences and the different kinds of experiences that students have based on major (discussed above). While it might be possible to randomly assign students within the same major and institution to differently designed units on professional ethics, this would only isolate the effects of individual interventions rather than explain the causal effect of the major as a whole.

A different approach would be to examine in detailed and in-depth ways how students experience their classes and curriculum in different majors and programs, focusing on the voices of students and their perspectives on their experiences. Holsapple, Carpenter, Sutkus, Finelli, and Harding (2012), for example, asked students in focus groups about their experiences with ethics education in their engineering courses. We found that students were experiencing these educational efforts in ways that were very different from faculty members' descriptions of their intentions. This kind of research could provide researchers and educators information about the

ways students are learning in their majors and the ways that these may encourage or inhibit growth. This would help faculty to make more informed decisions about the ways they organize their programs.

It is beyond the scope of this study to empirically examine the three potential explanations of the effects of student major on moral judgment presented here. Given the large effects of major, even when accounting for some precollege characteristics common to college impact research, it is important for future research to consider these effects in rigorous and systematic ways. The standardized effects of student major in this study dwarf the effects of any individual activities or programs, yet these activities and programs dominate the research on the development of moral judgment during college, and little is known about the sources of these large effects. I have suggested three potential reasons, and these are paths for future research.

In addition, these results suggest that student major be used as an important control when investigating the effects of other college experiences on moral judgment. My analysis of previously published literature shows that it is frequently ignored. This failure to control for the statistically significant effects of major lead to a threat of omitted variable bias (discussed in depth in Chapter IV) and to misestimation of the effects of the other variables in the studies (if these effects are also correlated with student major).

It is also important to note that in these analyses, students in humanities majors consistently demonstrate the highest levels of moral judgment, even after controlling for other factors. For the first-year sample, all other majors have statistically significant negative effects on moral judgment compared to the humanities; for the four-year sample, students majoring in business, social sciences, and engineering continue to show significantly lower levels of moral judgment than they peers majoring in the humanities, even after controlling for many other

characteristics and experiences. This provides an opportunity for educators to learn from the successful methods of encouraging moral development that may be in play in humanities classrooms.

The third way that student major may affect the development that discussed above is that there may be causal effects relating to the different content studied and different pedagogies used in the courses of different majors. If this is true, examining and adapting those methods used in humanities classes may provide insight to educators in other fields on how to adapt these practices to engineering, business, and other majors that do not have the same positive effects on their students' moral development.

Better understanding the effects of majors – and the underlying reasons for these effects – will also help educators ensure that students across all majors have the same access to in-class and out-of-class experience that encourage moral development. If deemed appropriate by the research, for example, students in "at-risk" majors can be targeted for interventions that might address precollege differences that inhibit moral development or the methods of teaching about moral and ethics dilemmas can be redesigned in problematic majors to better encourage post-conventional moral reasoning.

Courses, programs, and activities. In the college impact framework models, I also included as independent variables indicators of students' participation in a range of courses, formal programs, and other cocurricular activities. These results are noteworthy not for the specific experiences that are predicted to positively or negatively affect the development of moral judgment, but rather for the almost complete absence of significant effects for these variables. In the first-year sample, after precollege characteristics were included in the model, only one independent variable, participating in community service, was significantly related to

moral judgment. In the fourth-year sample, two additional independent variables were significant predictors: taking a class dealing with diversity issues, and participating in a leadership training program. None of the other activities (which were chosen because of their presence in the existing literature) demonstrated statistically significant effects on students' moral judgment.

These results are perhaps unsurprising. To demonstrate the difficulty in gauging the impact of these experiences on students' development, consider one of non-significant experiences, participation in a service-learning course. As detailed in Chapter II, previous research is mixed about role of service-learning participation in encouraging development. So this finding of no significant effect of service-learning participation on moral judgment in either sample is not an outlier in regards to prior studies. This result should not be interpreted to mean that service-learning programs are not useful tools in encouraging students' development of moral judgment. A service-learning course, like any sort of college experience included in a college impact framework, can be conducted in a virtually infinite number of ways. Some of these ways would be expected to encourage development, others to inhibit it, and others to have no effect. If a researcher asks, "Does participating in a service-learning class improve students' moral judgment?" the answer will almost always be "Maybe." More specifically, the answer is likely "Some of them, sometimes, for some students." It can be a useful finding to know that, over multiple studies, students who take a service-learning course demonstrate on average higher or lower levels of moral judgment (the current study suggests neither), but on its own, that finding does little to direct educators to create the conditions that lead to development, nor does it provide the kind of detail necessary for researchers to develop more complex theories about how to encourage that development in students.

These results, and the results of previous research, call into question the value of participation-focused approach. Simply finding that participation in a given activity (e.g., service-learning, membership in a professional organization, taking a women's studies class) predicts increases in moral judgment among a specific sample of students at specific set of schools provides little information for researchers to build theory or for educators to improve moral development educational efforts. It is not effective to tell educators to simply increase the number of service-learning classes or require a course in women's studies for all students, basing that recommendation in research that provides little to no guidance to help make those classes more developmentally effective. Dedicating resources to an intervention that has the potential to be successful if designed correctly is not an optimal strategy if the critical aspects of the design are not known; instead resources should go to designing those interventions in ways that are most likely to have the greatest positive effect. These variables of the typical college impact study do little in that regard.

The Threat of Selection Bias

In the two previous sections, I have discussed the implications of the results of this study's models estimated using the developmental and college impact frameworks. One implication is that these results suggest that certain college experiences do have the potential to encourage the development of moral judgment, and I have identified several of these experiences that are significant predictors of higher or lower moral judgment (N2) scores at the end of the first and fourth years of college in this study.

It is important to note, however, that this study do not allow for claims of causality from these results; due to these design of this study, I can only state that these experiences predict certain outcomes in this study; in order to make that claim of causality, it would be necessary to

limit the threat of selection bias. Selection bias refers to the biasing of treatment effects arising from non-random assignment of study participants into a treatment; this limits claims of causality because it is unclear whether a treatment is causing an outcome or whether some other characteristic is causing both the outcome and the sorting into the treatment (e.g., Heckman, 1979).

The threats to validity from selection bias arise from two possibilities: 1) that outcome differences arise not from the treatment, but from the same underlying variable that caused some students to be exposed to the treatment; and 2) heterogeneity of treatment effects, with students exposed to the treatment being affected by it differently than the students who were not (Brand & Yu, 2010). For example, in this study I found that participating in a leadership training program has a positive effect on students' development of moral judgment. It could be that leadership training programs actually caused the increase in moral development, or it could be an artifact of selection bias (or, perhaps most likely, a combination of the two). In this case, selection bias would be an issue if an underlying characteristic that is not included in the model (such as participating in a student organization) both makes a student more likely to participate in a leadership training program and to experience higher increases in moral judgment. It would also be an issue if some underlying characteristic (such as an openness to personal reflection) makes the student both more likely to participate in the program and more likely to benefit from it.

This issue is not specific to research on moral judgment in particular or to college outcomes in general. Barrow and Rouse (2005), for example, argue that despite decades of educational research, we can draw surprisingly few conclusions about what causes student success "because research has not emphasized isolating causal relationships between education inputs and student outcomes" (p. 1). Although I would argue that Barrow and Rouse and other

economists often go too far in claiming that that non-causal research does not produce actionable results, the threats of selection bias on results should not be ignored. This may be particularly true of the types of college experiences that are mostly often examined through a college impact framework in research on moral judgment because, according to Russell (2004) of the way that students choose activities while in college.

> This is a twofold problem in a free market of student choice: the impact of student experience and background and the impact of student aspirations. Especially in the character development area, programs try to attract, and are attractive to, students who are similar to the ones already there, or who want to be like them. It is here that program design and evaluation design may part company. Evaluation design assumes some random variation in students, but character development programs apparently thrive on nonrandomness (p. 106).

The existing literature on moral judgment has done little to account for selection bias other than to control for a range of covariates in regression models. This method, (also used in the current study) is insufficient for controlling for selection bias. Raudenbush and Bryk (2002) offer the following explanation:

> When random assignment of subjects to treatments is impossible, an attempt must be made to identify and control for individual background differences that are related to group membership and also to the outcome. This poses two problems: First, one can never be confident that all of the relevant background variables have been identified and controlled. Second, reasonable people can disagree about proper models for computing adjustment coefficients, and this choice of adjustments can have a substantial impact on inferences about the individual school effects. One general principle does emerge,

> however, in considering adjustments: The more dramatically different the groups are on background characteristics, the more sensitive inferences are likely to be to different methods of adjustment and the less credible the resulting inferences. (p. 155)

Despite this limitation in the method, the use of covariates is a method that is used in the vast majority of studies that make an attempt to limit selection bias in the study of the effects of college experiences on the development of moral judgment. This leads to the potential for misleading or inaccurate results from these studies, and is something that researchers must do a better job of addressing. It is an essential step in the advancement in research on moral judgment in college, especially if the goal of such research is to provide educators with information to help them plan and design more effective efforts for encouraging that development.

An important exception to this approach is a study by Grunwald and Mayhew (2008), in which they used propensity score matching to estimate causal effects of taking four different classes on students' moral judgment. When comparing the propensity score method with the more traditional covariate regression method, the propensity score matching analysis yielded treatment effects that were less than half of the size of the more traditional regression approach, calling into question the accuracy of results found using the more traditional method. This method can only account for observable differences; a true experimental design might well yield even smaller treatment effects (or none at all). Their paper is important because it was the first to examine the relationship between college experiences and moral judgment that used quasi-experimental methods to attempt to address selection bias. However, it has not ushered in a new era of causal analysis in moral judgment research. Despite calls by Grunwald and Mayhew, Mayhew & King (2004), and others for more research into the causal effects of college

experiences on moral judgment, the study by Grunwald and Mayhew remains the sole contribution to this approach.

In this study, I was not in the position to make causal claims of treatment effects in any models, largely because the purpose of this study was not to isolate treatment effects, but to compare and combine different frameworks that researchers use in investigating moral judgment outcomes. Nevertheless, selection bias can still lead to misestimation. Although I could not control for selection bias in this study, in this section I have considered the way that students may not be randomly distributed across the experiences that this study suggests positively or negatively influence the development of moral judgment. Although I cannot determine the extent to which students are not even distributed across experiences based on unobserved characteristics, I can examine the extent to which this is true based on observed characteristics. To do this, I conducted a series of significance tests to determine whether students of different races, genders, political orientations, or majors participate in the experiences at different rates using developmental, college impact, or integrated frameworks. Unequal distribution across experiences based on these characteristics would give an indication that students are not distributed randomly, leading to concerns about selection bias from other unobserved characteristics.

Table 5.1 presents the results of these tests for the first-year sample, and Table 5.2 presents the fourth-year results. These tests demonstrate that most of these college experiences are unequally distributed among students in at least one of these categories, and many experiences are unequally distributed among most or all of the categories. This indicates that selection bias may well be at play, with students' precollege characteristics and major (which were found in this study to predict moral judgment) playing roles in whether they are exposed to

the college experiences that also predict moral judgment. In other words, the precollege characteristics are correlated with both the other independent variables and the dependent variable in the models.

Table 5.1. *F*-Statistics for ANOVA Tests Examining the Distribution of Student Characteristics across Experiences that Predict Moral Judgment (First-Year Sample)

	Race (df=4)	Gender (df=1)	Political Orientation (df=2)	Major (df=10)
Classroom Encounters with Disequilibrium	3.25*	11.12***	4.70**	15.22***
Self-initiated encounters with disequilibrium	1.45	0.26	27.19***	11.85***
Negative Interactions with diverse peers	17.99***	27.11***	0.12	2.50**
Experiences with higher order learning	2.84*	0.44	2.80+	5.99***
Experiences with integrative learning	4.70***	0.18	18.44***	11.74***
Service-learning	3.47**	2.15	0.71	5.20***
Honors Program	0.53	2.46	1.65	2.79**
Student Organization Leader	2.09	1.35	4.67**	6.00***
Leadership training program	2.97	2.89+	5.20**	3.15***
Community Service	1.64	4.30*	5.96**	5.49***
Varsity Athlete	8.45***	81.30***	7.88***	5.13***
Race Workshop	7.55***	0.52	15.96***	3.90
Institutions supports social interaction	5.58***	2.85+	2.74+	2.35**
Institution contributes to my development	2.48*	2.35	11.17***	4.55***

Note. + $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$.

Table 5.2. *F*-Statistics for ANOVA Tests Examining the Distribution of Student Characteristics across Experiences that Predict Moral Judgment (Fourth-Year Sample)

	Race/Ethnicity (df=4)	Gender (df=1)	Political Orientation (df=2)	Major (df=9)
Classroom Encounters with Disequilibrium	0.45	5.07*	2.67+	37.08***
Self-initiated encounters with disequilibrium	2.16+	0.24	9.11***	8.42***
Negative Interactions with diverse peers	13.53***	15.36***	0.51	4.49***
Experiences with integrative learning	4.73***	4.73*	17.32***	18.45***
Service-learning class	2.58*	36.86***	5.72**	7.07***
Honors Program	0.22	19.61***	4.85**	1.81+
One diversity class	0.35	3.51+	2.75+	2.13*
Leadership Training Program	3.47**	0.01	0.82	3.08**
Religious Congregation	2.95*	0.20	94.82***	1.51
Social/Political Lecture	1.09	1.12	15.32***	16.59***
Community Service	0.14	23.86***	2.17	3.78***
Greek Organization	0.83	36.75***	19.55***	5.15***
Institution supports social interaction	2.15+	2.02	3.06*	4.41***
Institution contributes to my development	2.22+	14.62***	8.14***	2.27*

Note. + $p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$.

It is important to note that these significant relationships do not just exist for the variables consistent with a college impact framework, but also for those consistent with the developmental and integrated framework. It might be unsurprising that students with different characteristics are more or less likely to participate in experiences such as community service, belong to a Greek social organization, or participation in a leadership training program. These experiences exemplify Russell's (2004) assertion that students choose – and are accepted by – organizations and programs into which they fit.

It is perhaps more surprising that variables consistent with the developmental framework are unevenly distributed. Why, for example, would one expect classroom encounters with disequilibrium to differ by gender, or for experiences with integrative learning to differ by race? This emphasizes the point that students may be unequally distributed into experiences in both expected and unexpected ways, and reminds researchers that their own assumptions about the limitations – or lack thereof – from selection bias are just that, assumptions. It is beyond the scope of this study to further examine this selection bias, but accounting for bias in rigorous and systematic ways should be a priority of researchers investigating the development of moral judgment during college in future studies.

Examining Moral Judgment with an Integrated Framework

This dissertation is informed by a central research question: *How does the integration of developmental and college impact frameworks explain changes in students' moral judgment during college?* To address that question, I considered the ways that developmental and college impact frameworks explain moral judgment development for each framework individually, and then compared and contrasted the results of those analyses to that of the results of analyses employing an integrated framework. These results (presented in detail in Chapter IV) provide compelling evidence to support the use of an integrated framework in future research. This framework provides information about the ways both types of experiences affect development and explains the variance that was previously only explained by one or the other frameworks. Additionally, the integrated framework has a greater explanatory power than either individual framework, and results from the individual frameworks are at greater risk for omitted variable bias. Using an integrated framework provides both a broader and more detailed picture of the development of moral judgment in college.

When using an integrated framework to re-examine the effects of the college experiences that were previously included in either the developmental or college impact frameworks, the effects of each experience are largely similar. (The specific effects of each of those college experiences on moral judgment are discussed in detail earlier in this chapter.) What is important to note for this discussion of the integrated framework is that the coefficients corresponding to almost all of these experiences are quite similar in terms of significance, direction, and size. As in the models estimated with the individual frameworks, under the integrated framework, the variables from the developmental framework provide several significant effects, and the variables from the college impact framework provide few significant effects, with the exception of student major after controlling for precollege characteristics.

This similarity of effects is important in supporting the use of an integrated framework in the study of the development of moral judgment during college. If the set of variables from either the developmental or college impact frameworks had no longer yielded statistically significant effects, it would not have supported the hypothesis that the development of moral judgment would be better and more fully understood by research that integrates the two frameworks. Rather, this would imply that the frameworks were not truly different, that the two sets of variables could simply have been different ways of investigating the same experiences. For example, as discussed above, under the college impact framework, there were statistically significant negative effects on moral judgment associated with majoring in the social sciences, engineering, and business. When discussing potential reasons for the negative effects of certain majors, I posited that one of these potential reasons is that the classes across majors differ by whether or not the pedagogical methods used encourage post-conventional moral reasoning. It is possible, then, that that students in those three majors are exposed to fewer experiences with

integrative learning (which had a positive effect on moral judgment under the developmental framework) and to more negative interactions with diverse peers (which had a negative effect on moral judgment under the developmental framework). If the negative effects of social science, engineering, and business majors were completely explained by those students having different levels of exposure to those two developmental variables compared to their peers in other majors, then one could conclude that those variables in the two frameworks were examining the same experiences in different ways. With the distinct approaches of the two frameworks to defining college experiences, this is certainly possible; however, at least in this study, this does not appear to the case. The consistent effects across the models suggest that the variables in the developmental and college impact frameworks are measuring experiences that are different, and the integrated framework is measuring more than either one individually.

That the two frameworks do not represent the same experiences in this study does not mean that it would be true for every potential pair of models estimated using a developmental and college impact framework. Given a hypothetical study that investigates an infinite number of variables including every possible developmental and college impact variable, it may be possible that the variance explained by the two variables would completely overlap. However, that might also not be true; it is impossible to know in the absence of this impossible study. Regardless of the unknowable results of this hypothetical study, the results discussed here suggest that the research on moral judgment is far from approaching this point. As I discussed earlier, the current study is more comprehensive than virtually all published studies on the development of moral judgment during college and considers a wider range of independent variables.

Explanatory Power of the Integrated Framework

The results of this study also suggest that the integrated framework provides more explanatory power to our understanding of the development of moral judgment than either the developmental or college impact frameworks do individually. As discussed in the previous section, the similarity of effects for variables under the integrated framework compared to their effects under the original developmental or college impact frameworks supports the notion that these two frameworks are examining different factors that are related to the variance in moral judgment in different ways. If this is true, then integrating the two frameworks so that both of those groups of experiences are examined at the same time should lead to estimated models that predict more of the variance in moral judgment than either of the two original frameworks does on its own.

Comparisons of the pseudo- r^2 values calculated for each framework indicate that in this study, the integrated framework does indeed account for significantly more variance in moral judgment than either framework by itself, at least before precollege characteristics are taken into account. Similarly, although the differences in effects between the frameworks are modest after precollege characteristics are taken into account, the integrated framework still accounts for more variance in moral judgment at the end of both the first and fourth years that either framework alone. And though the greater number of variables would, by definition, increase the r^2compared to either of the original frameworks, the differences between the integrated model and the original frameworks persist even when comparing adjusted pseudo- r^2 values; this analysis includes a downward adjustment for the number of independent variables in the estimated models, which indicates that the increase in r^2 value is not just a factor of the larger number of independent variables.

The total adjusted pseudo- r^2 values for the integrated models was higher than for either the developmental or college impact models, with estimated models accounting for 9.4% of the variance in moral judgment at the end of the first year and 9.0% at the end of the fourth year. The models using the integrated framework account for almost the total of the sum of the variance explained by the models using the developmental and college impact frameworks (before precollege characteristics are taken into account). In the first-year sample, the adjusted pseudo- r^2 is 0.056 for the developmental framework model and 0.47 for the college impact framework model. If the models employing these two frameworks were explaining completely different portions of the total variance in moral judgment, the r^2 of the integrated framework model would be 0.103 (the sum of the variance explained by the two original models). The r^2 of the first-year integrated framework does not account for all of that summed variance in the two original models; however, it does account for the vast majority (91%) of that variance. Similarly, the 0.090 adjusted pseudo- r^2 of the integrated framework model of the fourth year data accounts for 90% of the sum of the variance of the two original frameworks. Because of the multilevel nature of the models in this study, and because of the downward r^2 adjustments for the number of independent variables, the adjusted pseudo- r^2 values do not refer precisely to the amount of variance that is explained by each model; however, this general pattern remained consistent and illustrative.

An examination of the r^2 values for each of the models of the frameworks not only points to the increased proportion of variance explained by the integrated framework, but this examination also indicates the relatively small amount of the variance that is explained by the college experiences in any of these models compared to the precollege characteristics. Before the precollege characteristics are added to the models, the integrated framework models account for

9.4 and 9.0% of the variance in moral judgment at the end of the first and fourth years of college, respectively. Those values are in comparison to the 45.7% and 34.4% of the variance, respectively, that is accounted for by the precollege characteristics. This means that the model with precollege characteristics accounts for approximately five times as much variance in moral judgment as the in-college experiences included in the model at the end of the first year. The precollege characteristics remain a powerful predictor even after four years of college, with the model including these precollege characteristics explaining almost three times as much variance in moral judgment as the model with only in-college experiences.

The difference between the role of precollege characteristics and in-college experiences in predicting moral judgment at the end of the first and fourth years of college looks even greater when comparing the variance explained by the complete integrated framework models (with all in-college experiences and the precollege characteristics) with that explained by the models that only include precollege characteristics. For both the first-year and fourth-year models, the models with precollege characteristics and in-college experiences have little more explanatory power than the models with just the precollege characteristics by themselves. This is true for all frameworks and both samples. In the first year, the adjusted pseudo- r^2 of the precollege characteristics model is 0.457, compared to 0.480, 0.461, and 0.479 for the developmental, college impact, and integrated framework models, respectively. Patterns are the same for the fourth-year sample. Adding the effects of the college experiences to the models adds very little explanatory power to the models, suggesting that even the college experiences (from either framework) that are significant predictors of moral judgment account for very little of the variance in the moral judgment after one or even four years of college.

In Chapter III, I discussed the selection of N2 score at the end of the first year and the end of the fourth year as the outcomes for this study. With this approach, the score at these two time points is used as the outcome with no adjustment based on students' entering N2 scores. Instead, the baseline score is included as a covariate (in the precollege characteristics block of independent variables), ultimately "controlling" for the large effect of the baseline assessment of moral judgment on the level of moral judgment at later points in time. Thus, after this control, the regression coefficients for other independent variables effectively refer to the effect of an independent variable on the change in moral judgment, since the baseline score has already been accounted for in the model. Another option would have been to use as an outcome the amount of change in the outcome over the course of the study, while either continuing, or not continuing to include the baseline measure as a covariate. (See the discussions by Pascarella, Wolniak, and Pierson, 2003; Pascarella and Wolniak, 2004; and Pike, 2004a; 2004b for more information about this approach.) Using this option, the outcome is the amount of change rather than a specific post-test N2 score. For this study, then, the outcomes under this approach would have been the change in N2 score between the beginning and end of the first year and the change between the beginning of the first year and the end of the fourth year. Since the outcome in this case is the change in N2 score, the regression coefficients, as in the first approach, refer to the effect of the independent variable on the change in moral judgment.

I ultimately opted to employ the first approach for two primary reasons. First, the use of the post-test score as the outcome and controlling and using a covariate to control for the baseline score is by far the most common approach in the existing body of literature examining the effects of college experiences on the development of moral judgment; no studies included in the analysis in Chapter II use the change score as an outcome. Mimicking the approach of the

existing literature as much as possible allow the results of my study to more effectively link these findings with prior research, making it easier to extrapolate from my results to the results of the existing and future body of research. Utilizing the second approach would have raised questions about the role that examining a different type of outcome played in my results and the extent to which they shed light on other research. Second, employing this approach allowed me to examine the effect of the baseline measure on later N2 scores in order to understand just how much these later outcomes are dependent on earlier levels of development. This important point would not be available for analysis using the other, less common, approach.

Despite what was gained from the approach I employed in this study, it does make it more complicated to interpret the r^2 values of models that include estimations of the effect of the baseline N2 score on the later N2 score. Since the baseline value of moral judgment is very predictive of later values (as is common among studied student outcomes), its inclusion in a model results in a very large r^2 value for that model. As such, these r^2 values are inflated indicators of the predictive power of college experiences on moral judgment. Although this r^2 value does represent the explanatory power of the model as a whole, it is inappropriate to use that value to draw conclusions about the specific predictive power of the college experiences in the model in light of these inclusive r^2 values – the model includes much more than just the college experiences. It is also important to remember when interpreting the r^2 values in these models and other studies that use this approach that the statistic used to measure the outcome is referring the proportion of variance explained in the moral judgment *outcome* accounted for by the model, not the change in moral judgment.

This is not a new problem, as this issue is often muddied in the existing literature. Of the 55 studies that I analyzed in Chapter II, most used a *t*-test or ANOVA to compare N2 scores

before and after an intervention. I discussed the problematic nature of these studies in that chapter. However, 13 of these studies used regression techniques that allow for the estimation of multiple effects at one time. These models produce an r^2 or pseudo- r^2 value that allows for the estimation of the proportion of variance in the outcome that is being accounted for in the model; of those, all by one present an r^2 or pseudo- r^2 value for their models. Almost all present a model r^2 that includes the outsized effects of the baseline moral judgment score. This is important information: it indicates that the independent variables in a model (including the baseline score) account for a certain proportion of variance and how much is still unknown. However, it provides very little information about how predictive the college experiences are for the outcome unless researchers provide a breakdown of the amount of variance explained by just those college experience variables. In multiple studies, Mayhew and colleagues (Mayhew & King, 2008; Mayhew, Seifert, & Pascarella, 2010; Mayhew, Seifert, Pascarella, Nelson Laird, & Blaich, 2012) add the baseline moral judgment measure to the estimated models separately and point specifically to the large increase in r^2 that accompanies that addition. This is the exception rather than the rule; most of these studies do nothing to indicate the amount of variance explained by the baseline measure compared to other independent variables, or even mention that the baseline accounts for a large amount of the variance.

One way to avoid this inflation of the r^2 by the baseline measure and other precollege variables that obscure the amount of variance actually explained by the college experiences is to include these variables separately in models, as is done in this study or in the studies by Mayhew and colleagues. Another way is to use the second approach described above to use the change score as the outcome variable rather than the post-test score. Not addressing this r^2 inflation in

what are otherwise rigorous studies leaves researchers and practitioners with potentially grossly misleading information about the effect on moral judgment of college experiences.

Conclusion

Since its beginnings in the American colonies, higher education in the United States has included students' moral education as an essential component of its mission. In recent years, constituencies including industry leaders, policy makers, and students themselves have responded to moral failings throughout the country by demanding higher education put more focus into this role, and colleges and universities themselves have committed extensive resources to programs, courses, and other institutional efforts aimed at encouraging students' moral development. Despite this commitment and emphasis, the research on the effectiveness of these efforts has remained inconclusive.

In this study, I aimed to provide a framework that would aid researchers in assessing these efforts and, in turn, help educators and administrators to utilize institutional resources in the most beneficial way. Student outcomes research usually employs one of two frameworks: developmental or college impact (Pascarella and Terenzini, 1991; 2005). The vast majority of research examining the effects of college experiences on moral judgment has employed a college impact model, focusing on the effects of institutional characteristics and/or formal or informal programs and experiences within the institution. Fewer employ a developmental framework, focusing on the experiences that developmental theory suggests would encourage development. A small handful of studies have relied on elements of both frameworks, but the body of research on the development of moral judgment has done little to systematically integrate these two approaches.

By comparing these frameworks, I have demonstrated that employing an integrated framework – one that includes elements of both developmental and college impact frameworks – increases the explanatory power of research compared to employing just one or the other. More importantly, this study directs attention to a broader way of thinking about educating students and researching the effects of those efforts. Pascarella and Terenzini (1991, 2005) posit that research on college student outcomes can largely be divided into that which is developmental in nature and that which is college impact; in this study I have referred to this distinction as the framework for the research. King (2014) draws a similar distinction in the way college educators view the kinds of strategies that promote positive student learning and development. King refers to the difference perspectives as using either an organizational or a student development approach; despite the different terms, she is also describing the difference between a college impact and developmental framework. In the first, these educators focus on creating the programs and other potential learning experiences for students, while in the second educators focus on creating developmental support and relationships that will trigger learning and development. As I have done here in regards to research on student outcomes, King argues that a third approach, one that focuses on the interaction of the two approaches as more desirable than either of the others alone. In other words, the integrated approach is not just applicable to research but has the potential to improve educational efforts on campuses. As in this research, this integrated approach is largely missing from the conceptualization of these efforts.

This study is a response to that type of dichotomy that both King (2014) and Pascarella and Terenzini (1991, 2005) describe as existing in both research and practice, and it provides evidence that the dichotomy is indeed a false one. The results of this study show that neither the college impact nor the developmental approach can, on its own, describe students' development

of moral judgment over the first year or the first four years of college as well as the two can when used together. Following the existing dichotomy and continuing to approach student outcomes research from one framework or the other will continue to place significant limits on what we know and what we can do about improving student outcomes in higher education. It is, instead, at the intersection of these frameworks that both researchers and educators must look to truly leverage the college experience and encourage positive student outcomes, including the development of moral judgment on which this study focuses.

These results also point to some concerns about the state of moral development in college and our ability to study it. In this study, the amount of variance in moral judgment attributed to the experiences examined in this study (comprising both the developmental and college impact frameworks) is less than 10% at its highest. It is fair to ask whether it ultimately matters what framework a researcher uses when the college experiences he or she investigates can be expected to explain so little student change. My answer is a resounding yes. Although the amount of variance in moral judgment explained by these experiences is small, even when using an integrated framework, it is still almost twice the amount that either framework describes on its own. That is not trivial: to not consider experiences consistent with one or the other of these frameworks is to ignore experiences that do contribute to students' moral judgment. Both statistically and conceptually, the evidence points to the importance of this integrated approach, both for researchers and educators.

It is also important to acknowledge how little of the total variance is being explained by college experiences. At the end of the first year (a commonly studied timeframe in college student outcomes research), students' precollege characteristics account for almost five times the amount of variance in moral judgment as is accounted for by all of the college experiences

included in the models. At the end of the fourth year, precollege characteristics still account for almost four times the amount of variance as college experiences. One explanation for this could be that when students arrive on campus, their moral judgment is mostly impervious to the effects college experiences, that no matter what colleges attempt to do, students' developmental path has mostly been set before coming to college. Previous literature (discussed earlier in this dissertation) shows that college attendance does positively affects moral judgment,; other studies show that moral judgment changes are influenced by specific college experiences. Neither set of studies supports the potential conclusion that students' moral development cannot be affected by institutional efforts. Instead, the small amount of variance in moral judgment explained by college experiences may be explained by the fact that existing methods of research do not adequately identifying the experiences that lead to that change.

A factor at play here is the type of data usesd. Here, I relied on data from two sources, both of which are representative of the dominant strands of college outcomes research. Student experience data were collected using the WNS survey instrument, which relied heavily on items from the National Study of Student Engagement (NSSE) survey, one of the two most extensively used assessments of student experiences. Institutional data were collected by the federal Department of Education and published as part of the Integrated Postsecondary Education Data, one of the only large scale sources for college-level data available to most higher education researchers. This suggests that the current practices used in student outcomes research needs to be revised. Although these two prominent sources of data may be useful for other questions, they are too narrowly focused to adequately inform important learning outcomes such as moral judgment development. We should not be content to rely on the same methods of research if those methods do not explain how students achieve essential learning outcomes. It will take new

and creative approaches to break away from these common patterns of research and identify more effective approaches.

The amount of variance in moral judgment outcomes that is accounted for by college experiences compared to precollege characteristics is disturbing for researchers and educators; so is the amount of moral judgment development that takes place in college for most students. In the first year of college, students gain on average fewer than five points on their N2 scores, and in the first four years, the average increase is fewer than nine points. This is in comparison to standard deviations of more than 15 points for both samples at the baseline assessment. Most students are seeing small improvements in moral judgment during college. This should concern everyone with an interest in preparing college students to deal with the complex moral situations in which they will find themselves after then leave college. As Sheppard, Macatangay, Colby, and Sullivan (2009) have pointed out, technology and societal needs are changing at such a rapid pace that it is impossible to know the kinds of moral issues that students will encounter. It is not optional that students be prepared for those unknown moral issues; it is an essential part of their education.

We cannot be satisfied with determining which experiences are responsible for small portions of the small changes that happening for most students' moral development when they are in college. Instead, we need a commitment from researchers to focus on understanding how to encourage considerable movements in development for large numbers of students while they and in college. As the Association of American Colleges and Universities (n.d.) has emphasized, educators and institutional leaders must commit themselves to improving our understanding of what they can do to help students achieve these outcomes; anything less than that is unacceptable.

These concerns about the effect of college experiences on students' moral judgment is less a statement about the potential of those experiences to encourage development than it is a statement about the current state of both the research and practice of those experiences as they relate to moral judgment development As they are experienced by students, the impact of these experiences is small. As they are measured by researchers, the ability to identify the effects of those experiences is limited. It will take new ideas from both researchers and educators to better leverage the power of developmentally effective cognitive activities (e.g., the key role disequilibrium in learning), This includes creating and assessing intentionally designed supportive experiences and environments to better tap the potential of a college education to encourage moral development and the achievement of other student outcomes that are essential for our changing society.

APPENDICES

Appendix A. Institutional Sample Sizes for Three Waves of Data Collection

College or University Label	Institutional Sample at Beginning of First Year	First-Year Sample		Fourth-Year Sample	
		N	Proportion of Original Sample	N	Proportion of Original Sample
3	193	170	0.881	106	0.549
4	170	131	0.771	125	0.735
5	122	90	0.738	63	0.516
6	105	76	0.724	45	0.429
7	97	57	0.588	45	0.464
8	74	57	0.770	49	0.662
9	97	58	0.598	25	0.258
10	93	76	0.817	60	0.645
11	114	89	0.781	65	0.570
12	163	90	0.552	89	0.546
13	135	116	0.859	74	0.548
14	330	214	0.648	154	0.467
15	179	133	0.743	87	0.486
16	71	44	0.620	42	0.592
17	132	60	0.455	49	0.371
18	86	58	0.674	41	0.477
22	211	125	0.592	55	0.261
31	117	42	0.359	49	0.419
32	135	17	0.126	10	0.074
33	106	89	0.840	72	0.679
34	108	37	0.343	37	0.343
35	562	291	0.518	95	0.169
36	256	6	0.023	148	0.578
41	176	74	0.420	28	0.159
42	195	80	0.410	52	0.267
43	226	161	0.712	67	0.296
44	59	13	0.220	5	0.085
45	241	113	0.469	153	0.635
46	202	44	0.218	120	0.594
48	75	3	0.040	19	0.253
49	347	106	0.305	20	0.058
50	65	36	0.554	26	0.400
51	159	122	0.767	53	0.333
52	26	4	0.154	5	0.192
53	120	52	0.433	59	0.492
55	92	24	0.261	21	0.228
58	69	47	0.681	37	0.536
59	221	69	0.312	87	0.394
61	182	53	0.291	67	0.368
62	29	14	0.483	7	0.241
63	74	19	0.257	9	0.122
64	298	3	0.010	121	0.406
67	181	20	0.110	24	0.133
70	200	94	0.470	102	0.510

Appendix B. The Results of *t*-Tests between the Baseline and Outcome N2 Score for both the First- and Fourth-Year Samples, Conducted on Institutional Subsamples

College or University Label	First-Year Sample			Fourth-Year Sample		
	Baseline N2	Outcome N2	Difference	Baseline N2	Outcome N2	Difference
			First Quartile			
41	22.790	23.202	0.412	23.202	30.13	6.928**
67	27.897	30.259	2.362	23.992	36.761	12.769***
32	21.038	24.731	3.693	17.494	28.071	10.577*
44	19.055	26.005	6.950	26.800	47.444	20.644
48	24.775	40.318	15.543	27.390	39.646	12.256**
49	27.109	30.911	3.802**	30.232	37.612	7.380*
35	26.495	30.765	4.270***	30.080	37.080	7.000***
61	25.523	30.106	4.583**	23.459	31.367	7.908***
34	25.490	30.572	5.082*	24.500	35.436	10.936***
18	28.162	34.000	5.838***	28.439	42.701	14.262***
17	26.824	32.711	5.887**	29.696	39.853	10.157***
			Second Quartile			
33	32.913	34.072	1.159	34.048	40.890	6.842***
42	30.764	33.576	2.812	29.836	36.641	6.805**
70	32.701	35.595	2.894*	31.612	39.126	7.514***
6	34.228	37.266	3.038*	36.677	44.817	8.140***
9	29.166	32.929	3.763*	28.872	40.051	11.179***
8	29.945	33.990	4.045**	29.267	39.793	10.526***
14	33.243	37.293	4.050***	32.245	39.258	7.013***
11	33.889	39.405	5.516***	34.257	46.938	12.681***
16	32.031	37.872	5.841***	30.768	40.178	9.410***
63	29.290	38.779	9.489**	30.647	39.954	9.307**
52	30.77	47.407	16.637*	41.292	43.061	1.769
			Third Quartile			
55	40.385	41.189	0.804	42.036	49.647	7.611*
46	39.498	40.318	0.820	37.847	39.489	1.642
51	37.219	38.901	1.682	37.408	49.996	12.588***
64	37.242	50.917	13.675	37.434	43.014	5.580***
53	37.782	41.653	3.871*	37.053	45.591	8.538***
15	36.433	40.685	4.252***	36.040	43.534	7.494***
12	40.352	44.954	4.602**	39.282	47.700	8.418***
13	36.895	41.760	4.865***	39.415	48.305	8.890***
5	34.579	39.695	5.116***	33.935	42.974	9.039***
31	37.296	44.772	7.476***	35.302	49.231	13.929***
43	35.463	42.966	7.503***	33.867	44.306	10.439***
			Fourth Quartile			
62	47.683	49.453	1.770	48.003	51.231	3.228
50	48.041	50.494	2.453	49.489	54.785	5.296*

36	44.474	47.427	2.953	46.821	53.158	6.337***
22	45.323	48.022	2.699*	45.622	50.300	4.678**
4	43.784	47.780	3.996***	43.530	50.687	7.157***
45	49.478	53.512	4.034***	49.271	56.691	7.420***
3	41.731	45.949	4.218***	42.160	50.592	8.432***
59	43.304	47.593	4.289**	42.817	50.006	7.189***
58	49.860	54.628	4.768**	47.464	59.673	12.209***
10	43.488	49.336	5.848***	43.514	49.251	5.737***
7	45.461	52.304	6.843***	46.739	51.746	5.007*

Note. * $p<.05$; ** $p<.01$; *** $p<.001$.

REFERENCES

ABET (2010). Criteria for accrediting engineering programs: Effective for evaluations during the 2010-2011 accreditation cycle. Retrieved February 14, 2011 from http://www.abet.org/Linked%20Documents-UPDATE/Criteria%20and%20PP/E001%2010-11%20EAC%20Criteria%201-27-10.pdf.

Collegiate Assessment of Academic Proficiency Association of American Colleges and Universities (2002). *Greater expectations: A new vision for learning as a nation goes to college.* Washington, DC: Association of American Colleges and Universities.

Astin, A.W. (1965). Effects of different college environments on the vocational choices of high aptitude students. *Journal of Counseling Psychology*, *12*, 28-34.

Astin, A. W. (1970a). The methodology of research on college impact (Part I). *Sociology of Education*, *43*, 223-254.

Astin, A. W. (1970b). The methodology of research on college impact (Part II). *Sociology of Education*, *43*, 437-450.

Astin, A. W. (1977). *Four critical years*. San Francisco: Jossey-Bass.

Astin, A. W. (1993). *What matters in college?* Four Critical Years *revisited*. San Francisco: Jossey-Bass.

Astin, A. W. & Panos, R. J. (1966). A national research bank for higher education. *Educational Record, 47*, 5-17.

Baldizan, L. & Frey, J.H. (1995). Athletics and moral development: Regulatory and ethical issues. *College Student Affairs Journal, 15*(1), 33-43.

Barber, J. P., King, P. M., & Baxter Magolda, M. B. (2013). Long strides on the journey toward self-authorship: Substantial developmental shifts in college students' meaning making. *Journal of Higher Education, 84*(6), 866-895.

Barrow, L. & Rouse, C. (2005). Causality, causality, causality: The view of education inputs and outputs from economics. *Federal Reserve Bank of Chicago Working Paper,* WP 2005-15, Chicago, IL: Federal Reserve Bank of Chicago.

Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development*. San Francisco: Jossey-Bass.

Baxter Magolda, M. B. (2001). *Making their own way: Narratives for transforming higher education to promote self-development*. Sterling, VA: Stylus.

Bebeau, M. J., & Thoma, S. J. (2003). *Guide for DIT-2: A guide for using the Defining Issues Test, Version 2 (DIT-2) and the scoring service of the Center for the Study of Ethical Development*. Minneapolis, MN: Center for the Study of Ethical Development.

Beckendorff, P., Ruhanen, L., & Scott, N. (2009). Deconstructing the student experience: A conceptual framework. *Journal of Hospitality and Tourism Management, 16,* 84-93.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.

Beller, J. M. & Stoll, S. K. (1992, Spring). A moral reasoning intervention program for student-athletes. *Academic Athletic Journal: The Official Publication of the National Association of Academic Advisors for Athletics,* 43-57.

Berger, J., & Milem, J. (2000). Organizational behavior in higher education and student outcomes. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. XV, pp. 268-338). New York: Agathon.

Bernacki, M. L. & Jaeger, E. (2008). Exploring the impact of service-learning on moral development and moral orientation. *Michigan Journal of Community Service Learning, 14*(2), 5-15.

Boss., J. A. (1994). The effect of community service work on the moral development of college ethics students. *Journal of Moral Development, 23*(2), 183-198.

Bowman, N. A. (2009). College diversity courses and cognitive development among students from privileged and marginalized groups. *Journal Diversity in Higher Education, 2*(3), 182-194.

Boyer, E. L. (1987). *College*. New York: Harper & Row.

Brand, J. E. & Yu, X. (2010). Who benefits from education? Evidence for negative selection in heterogenous economic returns to higher education. *American Sociological Review, 75*(2), 273-302.

Brendel, J. M., Kolbert, J. B., & Foster, V.A. (2002). Promoting student cognitive development. *Journal of Adult Development, 9*(3), 217-227.

Brim, O.G. (1966). Socialization through the life cycle. In O.G. Brim, Jr., and S. Wheeler, *Socialization After Childhood: Two Essays* (pp. 1-49). New York: Wiley.

Bronfenbrenner, U. (1993). The ecology of cognitive development: Research models and fugitive findings. In R. H. Wozniak & K. W. Fischer. (Eds.), *Development in context: Acting and thinking in specific environments* (pp. 3-44). Hillsdale, NJ: Erlbaum.

Brown-Liburd, H. L. & Porco, B. M. (2011). It's what's outside that counts: Do extracurricular experiences affect the cognitive moral development of undergraduate accounting students? *Issues in Accounting Education, 26*(2), 439-454.

Bunch, W. H. (2005). Changing moral judgment in divinity students. *Journal of Moral Education, 34*(3), 363-370.

Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119,* 197–253.

Cain, J. & Smith, D. (2009). Increasing moral reasoning skills through online discussions. *The Quarterly Review of Distance Education, 10*(2), 149-163.

Campus Compact. (2009). *2008 Service Statistics: Highlights and Trends of Campus Compact's Annual Membership Survey*. Boston, MA: Campus Compact.

Cano, C. R. & Sams, D. (2011). Advancing cognitive moral development: A field observation of college students. *Journal of Academic and Business Ethics, 3*. Retrieved from http://www.aabri.comwww.aabri.com/manuscripts/10556.pdf.

Casteen, L. D., Gibson, E. M., & Lampkin, P. M. (2007). Fostering moral reflection and perception in college. *Journal of College and Character, 9*(1), 1-8.

Clopton, N. A. & Sorell, G. T. (1993). Gender differences in moral reasoning: Stable or situational? *Psychology of Women Quarterly, 17*, 85-101.

Colby, A., Ehrlich, T., Beaumont, E., and Stephens, J. (2003). *Educating citizens: Preparing America's undergraduates for lives of moral and civic responsibility.* San Francisco: Jossey-Bass, 2003.

Colby, A. & Kohlberg, L. (1987). *The measurement of moral judgment: Theoretical foundations and research validation (Vol. 1).* Cambridge: Cambridge University Press.

Cummings, R., Dyas, L., Maddux, C.D., & Kochman, A. (2001). Principled moral reasoning and behavior of preservice teacher education students. *American Education Research Journal, 38*, 143-158.

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist, 31*(3), 163-174.

Derryberry, W. P. & Thoma, S. J. (2005). Functional differences: Comparing moral judgment developmental phases of consolidation and transition. *Journal of Moral Education, 34*(1), 89-106.

Dey, E. L., Antonaros, M., Ott, M. C., Barnhardt, C. L., & Holsapple, M. H. (2010). *Developing a moral compass: What is the campus climate for ethics and academic integrity?* Washington, DC: Association of American Colleges and Universities.

Dey, E., Hurtado, S., Rhee, B., Inkelas, K. K., Wimsatt, L. A., & Guan, F. (1997). *Improving research on postsecondary outcomes: A review of the strengths and limitations of national data sources*. Palo Alto, CA: Stanford University, National Center for Postsecondary Improvement.

Dotger, B. H. (2010). "I had no idea": Developing dispositional awareness and sensitivity through a cross-sectional pedagogy. *Teaching and Teacher Education, 26*, 805-812.

Drake, M. J., Griffin, P. M., Kirkman, R., & Swann, J. L. (2005). Engineering ethical curricula: Assessments and comparison of two approaches. *Journal of Engineering Education, 94*(2), 223-231.

DuPont, A. M. & Craig, J. S. (1996). Does management experience change the ethical perceptions of retail professionals: A comparison of the ethical perceptions of current students with those of recent graduates. *Journal of Business Ethics, 15*, 815-826.

Eagen, M. K. & Jaeger, A. J. (2008). Closing the gate: Part-time faculty instruction in gatekeeper courses and first-year persistence. *New Directions for Teaching and Learning, 115*, 39-54.

Earley, C. E. & Kelly, P. T. (2004). A note on ethics educational interventions in an undergraduate auditing course: Is there an "Enron Effect"? *Issues in Accounting Education, 19*(1), 53-71.

Ehrlich, T. (2000). *Higher Education and Civic Responsibility.* Phoenix, Ariz.: Oryx Press.

Elm, D. R., Kennedy, E. J., & Lawton, L. (2001). Determinants of moral reasoning: Sex role orientation, gender, and academic factors. *Business Society, 40*, 241-265.

Endicott, L., Bock, T., & Narvaez, D. (2003). Moral reasoning, intercultural development, and multicultural experiences: Relations and cognitive underpinnings. *International Journal of Intercultural Relations, 27*, 403-419.

Evans, N. J. (2003). Psychological, cognitive, and typological perspectives on student development. In S. R. Komives & D. B. Woodward, Jr. (Eds.), *Student services: A handbook for the profession* (4th ed., pp. 179-202). San Francisco, CA: Jossey-Bass.

Evans, N. J., Forney, D. S., Guido-DiBrito, F., Patton, L. D., & Renn, K. A. (2010). *Student development in college: Theory, research, and practice (2nd Ed.).* San Francisco, CA: Jossey-Bass.

Evans, N. J. & Reason, R. D. (2001). Guiding principles: A review and analysis of student affairs philosophical statements. *Journal of College Student Development, 42*, 359-377.

Falls, M.D. (2009). *Psychological sense of community and retention: Rethinking the first-year experience of students in STEM* (Unpublished doctoral dissertation). University of Central Florida, Orlando, FL.

Feldman, K. (1971). Measuring college environments: Some uses of path analysis. *American Educational Research Journal, 8*, 51-70.

Feldman, L., & Newcomb, T. (1969). *The impact of college on students*. San Francisco: Jossey-Bass.

Finelli, C. J., Holsapple, M. A., Ra, E., Bielby, R. M., Burt, B. A., Carpenter, D. D., Harding, T. S., & Sutkus, J. A. (2012). An assessment of engineering students' curricular and co-curricular experiences and their ethical development. *Journal of Engineering Education, 101*(3), 469-494.

Finger, W. W., Borduin, C. M., & Baumstark, K. E. (1992). Correlates of moral judgment in college students. *Journal of Genetic Psychology, 153*(2), 221-223.

Fitzgerald, J., Gottschalk, P., and Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources, 33*, 251–299.

Fleming, D. M., Romanus, R. N., Lightner, S. M. (2009). The effect of professional context on accounting students' moral reasoning. *Issues in Accounting Education, 24*(1), 13-30.

Fuertes, J. N., Miville, M. L., Mohr, J. J., Sedlacek, W. E., & Gretchen, D. (1999). *Factor structure of the Miville-Guzman Universality Diversity Scale* (Research Rep. No. 1-99). College Park: University of Maryland, Counseling Center.

Fuller, W.A. (2009). *Measurement error models*. John Wiley & Sons: New York City.

Gelman, A. & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics, 48*(2), 241-251.

Gilligan, C. (1977). In a different voice: Women's conceptions of self and morality. *Harvard Educational Review, 47*, 481-517.

Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.

Good, J.L., Cartwright, C. (1998). The development of moral judgment among undergraduate university students. *College Student Journal, 32*(2), 270-276.

Goodman, M., Duffy, J., & Heffernan, M. (1994). Service experience and the moral development of college students. *Religious Education, 89*(3), 422-431.

Green, L. L. (1981). Safety need resolution and cognitive ability as interwoven antecedents to moral judgment. Social Behavior and Personality 9(2): 139–145.

Groves, R. M., Fowler, F.J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2004). *Survey Methodology*, Hoboken, NJ: Wiley.

Grunwald, H. E. & Mayhew, M. J. (2008). Using propensity scores for estimating causal effects: A study in the development of moral reasoning. *Research in Higher Education, 49*, 758-775.

Guo, S. (2005). Analyzing grouped data with hierarchical linear modeling. *Children and Youth Services Review*, 27(6), 637-652.

Hannan, M. T., Freeman, J. H., & Meyer, J. W. (1976). Specification of models for organizational effectiveness. *American Sociological Review, 41*, 136-143.

Harding, T. S., Carpenter, D. D., Finelli, C. J., and Passow, H. J. (2004). Does academic dishonesty relate to unethical behavior in professional practice? An exploratory study. *Science and Engineering Ethics*, 10, 311–24.

Hare-Mustin, R. T. & Marcek, J. (1988). The meaning of difference: Gender theory, post-modernism, and psychology. *American Psychologist, 43*, 455-464.

Harper, S.R. (2009). Race-conscious student engagement practices and the equitable distribution of enriching educational experiences. *Liberal Education, 95*(4), 38-45.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 53-161.

Herington, C. & Weaven, S. (2007). Does marketing attract less ethical students? An assessment of the moral reasoning ability of undergraduate marketing students. *Journal of Marketing Education*, 29(2), 154-163.

Heinrich, C. J. & Lynn, L. E. (2001). Means and ends: A comparative study of empirical methods for investigating governance and performance. *Journal of Public Administration Research and Theory, 11*(1), 109-138.

Hersh, R.H. & Schneider, C.G. (2005). Fostering personal and social responsibility on college and university campuses. *Liberal Education, 91*(3), 6-13.

Holbrook, A., Cho, Y. I., & Johnson, T. (2006). The impact of questions and responded characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly, 70*(4), 565-595.

Holsapple, M. A., Sutkus, J., Carpenter, D., Finelli, C., Burt, B., Ra, E., Harding, T. & Bielby, R. (2011). We can't get no satisfaction!: The relationship between students' ethical reasoning and their satisfaction with engineering ethics education. Paper presented at the annual meeting of the American Society for Engineering Education.

Holsapple, M. A., Carpenter, D.D., Sutkus, J. A., Finelli, C.J., & Harding, T. S. (2012). Framing faculty and student discrepancies in engineering ethics education delivery. *Journal of Engineering Education, 101*(2), 169-186.

Humphreys, D., & Davenport, A. (2005). What really matters in college: How students view and value liberal education. *Liberal Education, 91*(3), 36–43.

Hurtado, S., Mayhew, M. J., & Engberg, M. E. (2012). Diversity courses and students' moral reasoning: A model of predispostiion and change. *Journal of Moral Education, 41*(2), 201-224.

Jaeger, A. J., & Hinz, D. (2008). The effects of part-time faculty on first semester freshman retention: A predictive model using logistic regression. *Journal of College Student Retention, 10*(3), 265-286.

Jagger, S. (2011). Moral judgment in computer undergraduates. *Journal of Information, Communication, & Ethics in Society, 9*(1), 20-33.

Jagger, S. & Strain, J. (2007). Assessing students' ethical development in computing with the Defining Issues Test: Challenges to be addressed. *Journal of Information, Communication & Ethics in Society, 5*(1), 33-42.

Jeffrey, C. (1993). Ethical development of accounting students, non-accounting business students, and liberal arts students. *Issues in Accounting Education, 8*(1), 86-96.

Jones, D. A. (2009). A novel approach to business ethics training: Improving moral reasoning in just a few weeks. *Journal of Business Ethics, 88*, 367-379.

Kaplan, L. E. (2006). Moral reasoning of MSW social workers and the influence of education. *Journal of Social Work Education, 42*(3), 507-522.

Kegan, R. (1982). *The evolving self: Problem and process in human development*. Cambridge, MA: Harvard University Press.

Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.

King, P. M. (2009). Principles of development and developmental change underlying theories of cognitive and moral development. *Journal of College Student Development*, 50(6), 597-620.

King, P. M. (2010). The role of the cognitive dimension of self-authorship: An equal partner or the strong partner? In M. B. Baxter Magolda, E. G. Creamer & P. S. Meszaros (Eds.), *Development and assessment of self-authorship: Exploring the concept across cultures,* pp. 167-185). Sterling, VA: Stylus.

King, P. M. (2014). Enriching the student learning experience: Linking student development and organizational perspectives. *About Campus: Enriching the Student Learning Experience*, 19(1), 7-13.

King, P. M., Baxter Magolda, M. B., Barber, J. P., Kendall Brown, M., & Lindsay, N. K. (2009). Developmentally effective experiences for promoting self-authorship. *Mind, Brain, and Education*, *3*(1), 108-118.

King, P. M., Baxter Magolda, M. B., & Massé, J. (2011). Maximizing learning from engaging across difference: The role of anxiety and meaning-making. *Equity & Excellence in Education*, *44*(4), 1-20.

King, P. M., Brown, M. K., Lindsay, N. K., & Vanhecke, J. R. (2007). Liberal arts student learning outcomes: An integrated approach. *About Campus, 12*(4), 2-9.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.

King, P. M. & Kitchener, K. S. (in press). Cognitive development in the emerging adult: The emergence of complex cognitive skills. In J. Arnett, (Ed.), *Handbook of Emerging Adulthood*. Oxford University Press.

King, P. M. & Mayhew, M. J. (2002). Moral judgment development in higher education: Insights from the Defining Issues Test. *Journal of Moral Education, 31*(3), 247-270.

King, P. M. & Mayhew, M. J. (2004). Theory and research on the development of moral reasoning among college students. *Higher Education: Handbook of Theory and Research, Vol. XIX*, 375-440.

Klimek, J. & Wenell, K. (2011). Ethics in accounting: An indispensible course? *Academy of Educational Leadership Journal, 15*(4), 107-118.

Kohlberg, L. (1972). A cognitive-developmental approach to moral education, *Humanist, 6*, 13-16.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research, and social issues* (pp. 31-53). New York, NY: Holt, Reinhart, and Winston.

Krawczyk, R.M. (1997). Teaching ethics: Effect on moral development. *Nursing Ethics*, *4*(1), 57-64.

Kreft, I., & de Leeuw, J. (1999). *Multilevel analysis.* Thousand Oaks, CA: Sage.

Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the National Study of Student Engagement. *Change, 33*(3), 10-17, 66.

Kuh, G. D. (2008). *High-impact educational practices: What are they, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities.

Lacy, W. (1978). Interpersonal relationships as mediators of structural effects: college student socialization in a traditional and an experimental university environment. *Sociology of Education, 51*, 201-211.

Latiff, D. A. (2000). The relationship between ethical dilemma discussion and moral development. *American Journal of Pharmaceutical Education, 64*(2), 126-133.

Lee. V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35*(2), 125-141.

Lies, J. M., Bock, T., Brandenberger, J., & Trozzolo, T. A. (2012). The effects of off-campus service-learning on the moral reasoning of college students. *Journal of Moral Education, 41*(2), 189-199.

Lincoln, T. D. (2009). *The seminary experience: Conceptual worlds of first-career and second-career seminarians* (Unpublished doctoral dissertation). University of Texas, Austin, TX.

Livingstone, G., Derryberry, W. P., King, A., & Vendetti, M. (2006). Moral development consistency? Investigating differences and relationships among academic majors. *Ethics & Behavior, 16*(3), 265-287.

Loe, T. W. & Weeks, W. A. (2000). An experimental investigation of efforts to improve sales students' moral reasoning. *Journal of Personal Selling and Sales Management, 20*(4), 243-251.

Loui, M. C. (2006). Assessment of an engineering ethics video: Incident at Morales. *Journal of Engineering Education, 95*(1), 85-91.

Lovik, E. G. (2011). The impact of organizational features and student experiences on spiritual development during the first year of college. *Journal of College and Character, 12*(3), 1-10.

Lovik, E. G. & Volkwein, J. F. (2010). The impact of organizational features and student experiences on spiritual development during the first year of college. Paper presented at the annual conference of the Association for the Study of Higher Education.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oakes: Sage Publications.

Lyons, N. P. (1987). Ways of knowing, learning, and making moral choices. *Journal of Moral Education, 16,* 226-239.

Maeda, Y., Thoma, S. J., & Bebeau, M. J. (2009). Understanding the relationship between moral judgmnet and development and individual characteristics: The role of educational contexts. *Journal of Educational Psychology, 101* (1), 233-247.

Martin, G. L., Hevel, M. S., Asel, A. M., & Pascarella, E.T . (2011). New evidence on the effects of fraternity and sorority affiliation during the first year of college. *Journal of College Student Development, 52*(5), 543-559.

Mayhew, M. J. (2012). A multilevel examination of the influence of institutional type on the moral reasoning development of first-year students. *Journal of College Student Development, 83*(3), 367-388.

Mayhew, M. J. & Engberg, M. E. (2010). Diversity and moral reasoning: How negative diverse peer interactions affect the development of moral reasoning in undergraduate students. *Journal of Higher Education, 81*(4), 459-488.

Mayhew, M. J. & Fernandez, S. D. (2007). Pedagogical practices that contribute to social justice outcomes. *Review of Higher Education, 31*(1), 55-80.

Mayhew, M. J. & King, P. M. (2008). How curricular content and pedagogical strategies affect moral reasoning development in college students. *Journal of Moral Education, 37*(1), 17-40.

Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2010). A multi-institutional assessment of moral reasoning development among first-year students. *Review of Higher Education, 33*(3), 357-390.

Mayhew, M. J., Seifert, T. A., Pascarella, E. T., Nelson Laird, T. F., & Blaich, C. F. (2012). Going deep into mechanisms for moral reasoning growth: How deep learning approaches affect moral reasoning development for first-year students. *Research in Higher Education, 53(1),* 26-46.

Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2012). How the first year of college influences moral reasoning development for students in moral consolidation and moral transition. *Journal of College Student Development, 53*(1), 19-40.

McCabe, D. L. (1992). The influence of situational ethics on cheating among college students. *Sociological Inquiry, 62,* 365–374.

McCabe, D. L. & Treviño, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *Journal of Higher Education, 64,* 522–538.

McCoach, D.B. & Adelson, J. L. (2010). Dealing with dependence (Part 1): Understanding the effects of clustered data. *Gifted Child Quarterly, 54*(2), 152-155.

McEwen, M.K. (2003). The nature and use of theory. In S. R. Komives & D. B. Woodward, Jr. (Eds.), *Student services: A handbook for the profession* (4th ed., pp. 153-178). San Francisco, CA: Jossey-Bass.

McNeel, S. P., Abou-Zeid, B., Essenburg, T., Smith, R., Danforth, D., & Weaver, R. T. (1996). Business students: Are they really different? *Research in Christian Higher Education, 3,* 33-56.

Mednick, M. T. (1989). On the politics of psychological constructs: Stop the bandwagon, I want to get off. *American Psychologist, 44*, 1118-1123.

Mennuti, R. B. & Creamer, D. G. (1991). Role of orientation, gender, and dilemma content in moral reasoning. *Journal of College Student Development, 32*, 241-248.

Mentkowski, M., and Strait, M. . (1983). *A Longitudinal Study of Student Change in Cognitive Development, Learning Styles, and Generic Abilities in an Outcome-Centered Liberal Arts Curriculum* (Final Report to the National Institute of Education No. 6). Milwaukee, WI: Alverno College, Office of Research and Evaluation.

Mezirow, J. (1997). Transformative learning: Theory to practice. In P. Cranton (Ed.), *Transformative learning in action: Insights from practice* (New Directions for Adult and ContinuingEducation No. 74, pp. 5-12). San Francisco: Jossey-Bass.

Milville, M. L., Gelso, C. J., Pannu, R., Liu, W., Touradji, P., Holloway, P., & Fuertes, J. (1999). Appreciating similarities and valuing differences: The Milville-Guzman Universality-Diversity Scale. *Journal of Counseling Psychology*, 46(3), 291-307.

Mortimer, J. T. & Simmons, R.G. (1978). Adult socialization. *Annual Review of Sociology, 4*, 421-454.

Murphy, C. (2010). *Residential learning outcomes: Analysis using the college student experiences questionnaire at a large public research university* (Unpublished doctoral dissertation). University of South Florida, Tampa, FL.

Narvaez, D. & Bock, T. (2002). Moral schemas and tacit judgment or How the Defining Issues Test is supported by cognitive science. *Journal of Moral Education, 31*(3), 297-314.

Kellogg Commission on the Future of State and Land-Grant Universities. (1997). *Returning to our Roots: The Student Experience.* Washington, DC: National Association of State Universities and Land-Grant Colleges.

Nonis, S. and Swift, C.O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for Business, 77*(2), 69–77.

Ogilby, S. M., (1995). The ethics of academic behavior: Will it affect professional behavior? *Journal of Education for Business, 71*(2), 92-97.

O'Muircheartaigh, C. A., Krosnick, J. A., & Helic A. (1999). Middle alternatives, acquiescence, and the quality of questionnaire data. Presented at annual meeting of the American Association of Public Opinion Research: Fort Lauderdale, FL.

Padgett, R. D., Johnson, M. P., & Pascarella, E.T. (2012). First-generation undergraduate students and the first year of college: Additional evidence. *Journal of College Student Development, 53*(2), 243-266.

Pascarella, E. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research*, *50*, 545-494.

Pascarella, E. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. I). New York: Agathon.

Pascarella, E. T., & Colleagues (2007). *Methodological report for the Wabash National Study of Liberal Arts Education.* Retrieved from http://www.liberalarts.wabash.edu/study-research/.

Pascarella, E. T., Edison, M., Nora, A., Hagedorn, L., & Terenzini, P. (1996). Influences on students' openness to diversity and challenge in the first year of college. *Journal of Higher Education, 67*, 174-195.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students (Vol. 2): A third decade of research*. San Francisco: Jossey-Bass.

Perry, W. G., Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, and Winston.

Pfeffer, J. (1982) *Organizations and organization theory*. New York: Oxford University.

Piaget, J. (1948). *The moral judgment of the child*. Glencoe, IL: Free Press.

Piaget, J. (1970). *Genetic epistemology* (E. Duckworth, Trans.). New York: Columbia University Press.

Piaget, J. (1973). *Psychology of intelligence*. Totowa, NJ: Littlefield, Adams and Co.

Pizzoloto, J. E., Chaudhari, P., Murrell, E. D., Podobnik, S., & Schaeffer, Z. (2008). Ethnic identity, epistemological development, and academic achievement in under-represented students. *Journal of College Student Development, 49*(4), 301-318.

Ponemon, L. A. (1993). Can ethics be taught in accounting? *Journal of Accounting Education, 11*, 185-199.

Porter, S. R. (2005). What can multilevel models add to institutional research? In M. A. Coughlin (Ed.), *Applications of intermediate/advanced statistics in institutional research* (pp. 110–131). Resources in Institutional Research, No. 16. Tallahassee, FL: Association for Institutional Research.

Power, C., and Kohlberg, L. (1986). Moral development: Transforming the hidden curriculum. *Curriculum Review, 26(5),* 14-17.

Quarry, E. B. (1997). *The defining issues test as a measure of moral reasoning of undergraduate students at a Christian liberal arts institution in southern California* (Unpublished doctoral dissertation). Biola University, La Mirada, CA.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Reason, R. D. (2009). Understanding the development of the whole person. *Journal of College Student Development, 50,* 573-575.

Reason, R.D., Cox, B.E., Quaye, B.R.L., & Terenzini, P.T. (2010). Faculty and institutional factors that promote student encounters with difference in first-year courses. *Review of Higher Education*, *33*(3), 391-414.

Reason, R. D., Terenzini, P.T., & Domingo, R. J. (2005). Developing social and personal competence in the first year of college. Paper presented at the meeting of the Association for the Study of Higher Education, Philadelphia, PA.

Reason, R. D., Terenzini, P. T., & Domingo, R. J. (2006). First things first: Developing academic competence in the first year of college. *Research in Higher Education, 47*(2), 149-175.

Reason, R. D., Terenzini, P. T., & Domingo, R. J. (2007). Developing social and personal competence in the first year of college. *Review of Higher Education, 30,* 271-299.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.

Renn, K. A., & Arnold, K. D. (2003). Reconceptualizing research on peer culture. *Journal of Higher Education, 74,* 261-291.

Rest, J. R. (1979a). *Development in Judging Moral Issues*. Minneapolis, MN: University of Minnesota.

Rest, J. R. (1979b). *The impact of higher education on moral judgment development*. Moral Research Projects, Technical Report no. 5. Minneapolis, MN.

Rest, J. R. (1983) Morality. In P. H. Mussen (Series Ed.), J. Flavell & E. Markman (Vol. Eds.) *Handbook of child psychology: Cognitive development: Vol. 3.* (556–629). New York: Wiley.

Rest, J. R. (1986). *Moral development: Advances in research and theory.* New York: Praeger.

Rest, J. R., Narvaez, D., Bebeau, M., & Thoma, S. (1999a). A Neo-Kohlbergian approach: The DIT and schema theory. *Educational Psychology Review, 11*(4), 291-324.

Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. (1999b). *Postconventional thinking: A neo-Kohlbergian approach*. Hillsdale, NJ: Erlbaum.

Rest, J. R., Narvaez, D., Thoma, S., Bebeau, M. (2000). A Neo-Kohlbergian approach to morality research. *Journal of Moral Education, 29*(4), 381-395.

Rest, J. R., & Thoma, S.J. (1985). Relation of moral judgment development to formal education. *Developmental Psychology,* 21(4): 709–714.

Rest, J. R., Thoma, S. J., Narvaez, D., Bebeau, M. J. (1997). Alchemy and beyond: Indexing the Defining Issues Test. *Journal of Educational Psychology, 89*(3), 498-507.

Roberts, J. K. (2007). Group dependency in the presence of small intraclass correlation coefficients: An argument in favor of not interpreting the ICC. Presented at the annual meeting of the American Educational Research Association: Chicago, IL.

Rogers, G. (2002). Rethinking moral growth in college and beyond. *Journal of Moral Education, 31*(3), 325–338.

Russell, T. R. (2004). Assessing outcomes of character-building programs: Problems and prospects. *New Directions for Institutional Research, Summer*. 37-54.

Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069–1081.

Ryff, C., & Keyes, C. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69,* 719–727.

Saad, L. (2011). Fewer Americans down on U.S. moral values. Retrieved from http://www.gallup.com/poll/147794/fewer-americans-down-moral-values.aspx.

Sarin, S. & McDermott, C. (2003). The effects of team leader characteristics on learning, knowledge application, and performance of cross-functional new product development teams. *Decisions Sciences, 34*(4), 707-739.

Sax, L. J. (2004). Citizenship development and the American college student. *New Directions for Institutional Research, Summer,* 37-54.

Scott, W. R. (1976). Effectiveness of organizational effectiveness. In P.S. Goodman and J.M. Pennings (Eds.) New perspectives on organizational effectiveness. San Francisco, CA: Jossey Bass.

Seifert, T., Goodman, K., King, P. M. & Magolda, M. (2010). Using mixed methods to study first-year college impact on liberal arts learning outcomes. *Journal of Mixed Methods Research, 4* (3), 248-267.

Self, D. J., Baldwin, D. C., & Olivarez, M. (1993). Teaching medical ethics to first-year students by using film discussion to develop their moral reasoning. *Academic Medicine, 68*(5), 383-385.

Sheppard, S., K. Macatangay, A. Colby, W. Sullivan. (2009). *Educating engineers*. Jossey-Bass: San Francisco, CA.

Sheppard, J. P. & Young, M. (2007). The routes of moral development and the impact of exposure to the Milgram obedience study. *Journal of Business Ethics, 75*, 315-333.

Sims, R. L. (1993). The Relationship between Academic Dishonesty and Unethical Business Practices. *Journal of Education for Business 68*, 207-211.

Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models. *Journal of Educational and Behavioral Statistics, 24*, 323–55.

Smith, C. A., & Bunting, C. J. (1999). The moral reasoning of two groups of college students. *Research in Education, 62*, 72-74.

Smith, C. A., Strand, S. E., & Bunting, C. J. (2002). The influence of challenge course participation on moral and ethical reasoning. *Journal of Experiential Education, 25*(2), 278-280.

Snijders, T. A. B. & Bosker, R. J. (1994). Modeled variance in two-level model. *Sociological Methods and Researchers, 22*, 342-363.

Snodgrass, J. & Behling, R. (1996). Differences in moral reasoning between college and university business majors and non-business majors. *Business & Professional Ethics Journal, 15*(1), 79-84.

Stepp, J. O. (2002). *The relationship of moral development to critical thinking and family structure of matriculating college students* (Unpublished doctoral dissertation). The Southern Baptist Theological Seminary, Louisville, KY.

Stiller, N. J. & Forrest, L. (1990). An extension of Gilligan's and Lyons' investigation of morality: Gender differences in college students. *Journal of College Student Development, 31*, 54-63.

Strange, C. (1994). Student development: The evolution and status of an essential idea. *Journal of College Student Development, 35*(6), 399-412.

Strange, C.C. & Banning, J. (2001). *Educating by design: Creating campus learning environments that work.* San Francisco: Jossey-Bass.

Subar, A. F., Ziegler, R. G., Thompson, F. E., Johnson, C. C., Weissfeld, J. L., Reding, D., Kavounis, K. H., & Hayes, R. B. (2000). Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. *American Journal of Epidemiology, 153*(4), 404-409.

Sweat, J., Jones, G., Han, S., & Wolfgram, S. M. (2013). How does high-impact practice predict student engagement? A comparison of White and minority students. *International Journal for the Scholarship of Teaching and Learning, 7*(2), 1-24.

Terenzini, P. T. & Reason, R D. (2005). Parsing the first year of college: A conceptual framework for studying college impacts. Paper presented at the annual meeting of the Association of the Study of Higher Education: Philadelphia, PA.

Terenzini, P. T. & Reason, R. D. (2010). Toward a more comprehensive understanding of college effects on student learning. Paper presented at the annual meeting of the Consortium of Higher Education Researchers: Oslo, Norway.

Terenzini, P., Springer, L., Pascarella, E., & Nora, A. (1995). Academic and out-of class influences affecting the development of students' intellectual orientations. *Review of Higher Education, 19*, 23–44.

Terman, L. M. (1973). *Concept Mastery Test manual*. New York: Psychological Corporation.

Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. *Developmental Review, 24,* 128-137.

Thoma, S. J. (2002). An overview of the Minnesota approach to research on moral development. *Journal of Moral Education, 31*(3), 225-245.

Thoma, S. J. (2006). Research on the Defining Issues Test. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 67–91).Mahwah, NJ: Erlbaum

Thoma, S. J. & Rest, J. R. (1999). The relationship between moral decision making and patterns of consolidation and transition in moral judgment development. *Developmental Psychology, 35*(2), 323-334.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Torres, V. (2011). Using student development theories to explain student outcomes. *Higher Education: Handbook of Theory and Research, 26*, 425-448.

Traiser, S. & Eighmy, M. A. (2011). Moral development and narcissism of private and public university business students. *Journal of Business Ethics, 99*, 325-224.

Tyree, T. M. (1998). *Designing an instrument to measure socially responsible leadership using the social change model of leadership development* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

Walberg, H. A. (1982). A psychological theory of educational productivity. In F. Farley and N. Gordon (eds.), *Psychology and Education*, Berkeley, CA: McCutchan.

Walberg, H. A., Pascarella, E., Haertel, G., & Junker, L. (1982). Probing a model of educational productivity in high school science with national assessment samples. *Journal of Educational Psychology, 74*, 295-307.

Weick, K. E. (1969). *The social psychology of organizing*. Reading, MA: Addison-Wesley.

Weidman, J. C. (1984). Impacts of campus experiences and parental socialization on undergraduates' career choices. *Research in Higher Education, 20*, 445-476.

Weidman, J. (1989). Undergraduate socialization: A conceptual approach. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. V). New York: Agathon.

Weidman, J. C. & Freidmann, R. R. (1984). The school-to-work transition for high school dropouts. *Urban Review, 16,* 25-42.

Weidman, J. C. & White, R. N. (1985). Postsecondary "high-tech" training for women on welfare: Correlates of program completion. Journal of Higher Education, *56*, 555-568/.

Wenger, D. M. & Wheatley, I. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480-492.

Wilhelm, W. J. & Czyzewski, A. B. (2012). Ethical reasoning instruction in undergraduate cost accounting: A non-intrusive approach. *Academy of Educational Leadership Journal, 16*(2), 131-142.

Wilson, K. L., Rest, J. R., Boldizar, J. P. & Deemer, D. K. (1992). Moral judgment development: The effects of education and occupation. *Social Justice Research, 5*(1), 31-48.

Wingspread Group on Higher Education (1993). *An American imperative: Higher expectations for higher education.* Racine, WI: Johnson Foundation.

Yuan, Y.C. (2000). Multiple imputation for missing data: Concepts and new development. *In Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference* (Paper No. 267). Cary, NC: SAS Institute.

Young, R.B . (2003). Philosophies and values guiding the student affairs profession. In S. R. Komives & D.B. Woodward, Jr. (Eds.), *Student services: A handbook for the profession* (4th ed., pp. 89-106). San Francisco, CA: Jossey-Bass.