RESEARCH METHODS

Samples

Institutional Sample. The sample in the study consisted of incoming first-year students at 19 four-year and two-year colleges and universities located in 11 different states from 4 general regions of the United States: Northeast, Southeast, Midwest, and Pacific Coast. Institutions were selected from more than 60 colleges and universities responding to a national invitation to participate in the Wabash National Study of Liberal Arts Education (WNSLAE). Funded by the Center of Inquiry in the Liberal Arts at Wabash College, the WNSLAE is a large, longitudinal investigation of the effects of liberal arts colleges and liberal arts experiences on the cognitive and personal outcomes theoretically associated with a liberal arts education. The institutions were selected to represent differences in college and universities nationwide on a variety of characteristics including institutional type and control, size, location, and patterns of student residence. However, because the study was primarily concerned with the impacts of liberal arts colleges and liberal arts experiences, liberal arts colleges were purposefully over-represented.

Our selection technique produced a sample with a wide range of academic selectivity, from some of the most selective institutions in the country to some that were essentially open admissions. There was also substantial variability in undergraduate enrollment, from institutions with entering classes between 3,000 and 6,000, to institutions with entering classes between 250 and 500. According to the 2007 Carnegie Classification of Institutions, 3 of the participating institutions were considered research universities, 3 were regional universities that did not grant the doctorate, 2 were two-year community colleges, and 11 were liberal arts colleges.

Student Sample. The individuals in the sample were first-year, full-time undergraduate students participating in the WNSLAE at each of the 19 institutions in the study. The initial

sample was selected in either of two ways. First, for larger institutions, it was selected randomly from the incoming first-year class at each institution. The only exception to this was at the largest participating institution in the study, where the sample was selected randomly from the incoming class in the College of Arts and Sciences. Second, for a number of the smallest institutions in the study—all liberal arts colleges—the sample was the entire incoming first-year class. The students in the sample were invited to participate in a national longitudinal study examining how a college education affects students, with the goal of improving the undergraduate experience. They were informed that they would receive a monetary stipend for their participation in each data collection, and were also assured in writing that any information they provided would be kept in the strictest confidence and never become part of their institutional records.

Data Collection

Initial Data Collection. The initial data collection was conducted in the early fall of 2006 with 4,501 students from the 19 institutions. This first data collection lasted between 90-100 minutes and students were paid a stipend of $50 each for their participation. The data collected included a WNSLAE precollege survey that sought information on student demographic characteristics, family background, high school experiences, political orientation, educational degree plans, and the like. Students also completed a series of instruments that measured dimensions of intellectual and personal development theoretically associated with a liberal arts education. These are described in greater detail in the subsequent section on "WNSLAE Outcomes/Dependent Measures."

Follow-up Data Collection. The follow-up data collection was conducted in spring 2007. This data collection took about two hours and participating students were paid an additional

stipend of $50 each. Two types of data were collected. The first was based on questionnaire instruments that collected extensive information on students' experience of college. Two complementary instruments were used: the National Survey of Student Engagement (NSSE) (Kuh, 2001) and the WNSLAE Student Experiences Survey (WSES). These instruments were designed to capture student engagement in, or exposure to, empirically vetted good practices in undergraduate education. These good practices included such dimensions as: exposure to effective teaching, quality of nonclassroom interactions with faculty, active learning, integrative experiences, influential interactions with other students, high expectations, and the like (Pascarella, Cruce, Wolniak, & Blaich, 2004; Pascarella, Cruce, Umbach, Wolniak, Kuh, Carini, Hayek, Gonyea, & Zhao, 2006). They are described in greater detail in the subsequent section on "WNSLAE Measures of Good Practices/Liberal Arts Experiences." The second type of data collected consisted of follow-up (or posttest) measures of the instruments measuring dimensions of intellectual and personal development that were first completed in the initial data collection. All students completed the NSSE and WSES prior to completing the follow-up instruments assessing intellectual and personal development. Both the initial and follow-up data collections were administered and conducted by ACT.

Of the original sample of 4,501 students who participated in the fall 2006 testing, 3,081 participated in the spring 2007 follow-up data collection, for a response rate of 68.5%. These 3,081 students represented 16.2% of the total population of incoming first-year students at the 19 participating institutions. To provide at least some adjustment for potential response bias by sex, race, academic ability, and institution in the sample of students, a weighting algorithm was developed. Using information provided by each institution on sex, race, and ACT score (or appropriate SAT equivalent or COMPASS score equivalent for community college students),

follow-up participants were weighted up to each institution's first-year undergraduate population by sex (male or female), race (Caucasian, African American/Black, Hispanic/Latino, Asian/Pacific Islander, or other), and ACT (or equivalent score) quartile. While applying weights in this manner has the effect of making the overall sample more similar to the population from which it was drawn, it cannot adjust for nonresponse bias.

Conceptual Framework for Liberal Arts Outcomes

Since the WNSLAE was fundamentally concerned with understanding the conditions and experiences that constituted an influential liberal arts education, its first task was to conceptually define the desired intellectual and personal outcomes of such an education. Synthesizing much of the literature on liberal arts education, and building on the work of Jones and McEwen (2000), King, Kendall Brown, Lindsay, and VanHecke (2007) developed a comprehensive model of liberal arts outcomes that embraced seven general dimensions: effective reasoning and problem solving, well-being, inclination to inquire and lifelong learning, intercultural effectiveness, leadership, moral character, and integration of learning. Although such outcome dimensions appear central to the undergraduate mission of a large cross-section of American colleges and universities (see, for example, the outcome taxonomy employed by Pascarella & Terenzini, 1991, 2005, in organizing college impact outcomes), the distinctiveness of the liberal arts outcomes lies in the integrated connections that exist between outcomes and their holistic nature, which spans cognitive, interpersonal, and intrapersonal domains. Consequently, the WNSLAE was largely guided by this conceptual framework of liberal arts outcomes in selecting specific outcome measures. Indeed, with the single exception of integration of learning, the WNSLAE study was able to identify specific outcome or dependent measures representing six of the seven liberal arts outcomes.

WNSLAE Outcome/Dependent Measures

Effective Reasoning and Problem Solving. To tap this outcome, we used the critical thinking module from the Collegiate Assessment of Academic Proficiency (CAAP) developed by the American College Testing Program (ACT). The critical thinking test is a 40-minute, 32-item instrument designed to measure a student's ability to clarify, analyze, evaluate, and extend arguments. The test consists of four passages in a variety of formats (e.g., case studies, debates, dialogues, experimental results, statistical arguments, editorials). Each passage contains a series of arguments that support a general conclusion and a set of multiple-choice test items. The internal consistency reliabilities for the CAAP critical thinking test range between .81 and .82 (ACT, 1991). It correlates .75 with the Watson-Glaser Critical Thinking Appraisal (Pascarella, Bohr, Nora, & Terenzini, 1995).

Well Being. We operationalized this dimension of liberal arts outcomes with several individual measures. The first was the Ryff Scales of Psychological Well-Being (SPWB) (Ryff, 1989; Ryff & Keys, 1995). The SPWB is a 54-item, theoretically-grounded instrument that specifically focuses on measuring six dimensions of psychological well-being: positive evaluations of oneself (Self-Acceptance), sense of continued growth and development as a person (Personal Growth), belief in a purposeful and meaningful life (Purpose in Life), quality relations with others (Positive Relations with Others), capacity to effectively manage one's life and surrounding world (Environmental Mastery), and sense of self-determination (Autonomy) (Ryff & Keyes, 1995; Ryff, 1989; Keyes, Shmotkin, & Ryff, 2002). The six 9-item scales have internal consistency reliabilities ranging from .83 to .91 (C. Ryff, personal communication, August 2004). The six SPWB scales tend to have significant, positive associations with frequently used measures of happiness and satisfaction, and negative associations with

depression. However, the clearest evidence of predictive validity against these criteria is with the Self-Acceptance and Environmental Mastery Scales (Ryff & Keyes, 1995).

In addition to the SPWB, the WNSLAE also considered five additional measures of physical well-being: the use of tobacco, frequency and intensity of alcohol use, frequency of aerobic exercise, frequency of sleep deprivation, and overall assessment of personal health. We added these dimensions of health and health-related behaviors to the study because of the extensive evidence suggesting that they are significantly and perhaps causally, linked to exposure to postsecondary education (see Pascarella & Terenzini, 1991, 2005, for a summary of this research). The measure of tobacco use was a single item that asked students how many cigarettes they smoked a day. There were five response options ranging from "I don't smoke cigarettes" to "2 or more packs." Frequency of alcohol use was measured by a single item that asked students how often in a typical week during college they consumed alcoholic beverages. There were nine response options from "0 times per week" to "More than 7 times per week." Intensity of alcohol use was measured by responses to a single item that asked students how often in a typical week period during college did they have 5 or more drinks in one setting. A drink was defined as a 12-ounce can of beer, a 4-ounce glass of wine, 1 wine cooler, 1 shot of liquor, or 1 mixed drink. There were five response options from "0 times" to "5 or more times." Frequency of aerobic exercise was assessed with a single item that asked how frequently one engaged in aerobic exercise (e.g., running, walking, hiking, swimming). The five response options ranged from "I don't exercise regularly" to "More than 6 hours per week." The measure of sleep deprivation was a single item that asked students how often they felt sleep deprived. There were five response options ranging from "Almost always" to "Never." Finally, overall

self-reported health status was measured with a single question that asked "Overall, how would you rate your health?" The five response options ranged from "Excellent" to "Very Poor."

Inclination to Inquire and Lifelong Learning. This outcome was operationally represented with two scales. The primary measure was the 18-item Need for Cognition Scale (NCS). Need for cognition refers to an individual's "tendency to engage in and enjoy effortful cognitive activity" (Cacioppo, Petty, Feinstein, & Jarvis, 1996, p. 197). Those who have a high need for cognition "tend to seek, acquire, think about, reflect back on information to make sense of stimuli, relationships, and events in their world" (p. 198). In contrast, those with low need for cognition are more likely to rely on others, such as celebrities and experts, cognitive heuristics, or social comparison processes to provide or make sense of their world. The reliability of the NCS ranges from .83 to .91 in samples of undergraduate students (Cacioppo, et al.). With samples of undergraduates, the NCS has been positively associated with the tendency to generate complex attributions for human behavior, high levels of verbal ability, engagement in evaluative responding, one's desire to maximize information gained rather than maintain one's perceived reality (Cacioppo, et al.) and college grades (Elias & Loomis, 2002). The NCS is negatively linked with authoritarianism, need for closure, personal need for structure, the tendency to respond to information reception tasks with anxiety, and chronic concern regarding self-presentation (Cacioppo, et al.).

The second measure designed to tap continuing motivation for lifelong learning was a 6-item measure entitled the Positive Attitude Toward Literacy Scale (PATL). The PATL assesses students' enjoyment of such literacy activities as reading poetry and literature, reading scientific and historical material, and expressing ideas in writing, and has an internal consistency reliability of .71. The PATL score at entrance to college correlated .36 with three-year cumulative scores

during college on a measure of library use, .48 with the cumulative number of unassigned books read during three years of college, and .26 with a measure of reading comprehension administered after three years of college (Bray, Pascarella, & Pierson, 2004).

Intercultural Effectiveness. This outcome dimension was measured with two scales. The primary measure was the 15-item, short form of the Miville-Guzman Universality-Diversity Scale (M-GUDS). The M-GUDS measures an individual's universal-diverse orientation, which is defined as "an attitude of awareness and acceptance of both similarities and differences that exist among people" (Miville, Gelso, Pannu, Liu, Touradji, Holloway, & Fuertes 1999; Fuertes, Miville, Mohr, Sedlacek, & Gretchen, 2000). The instrument has a total scale score and three subscale scores: Diversity of Contact (interest and commitment to participating in diverse, intentionally focused social and cultural activities), Relativistic Appreciation (appreciation of both similarities and differences in people and the impact of these in one's self-understanding and personal growth), and Comfort with Differences (the degree of comfort with diverse individuals). The internal consistency reliability for the total M-GUDS score in the present study was .85, while reliabilities for the three subscales ranged from .77 to .78. The precollege total M-GUDS score correlated .47 with a measure of students' experiences and interactions with diverse others and diverse ideas during the first year of college.

The second instrument used to assess student growth in intercultural effectiveness was the seven-item Openness to Diversity/Challenge (ODC) scale. This scale measures one's openness to cultural and racial diversity as well as the extent to which one enjoys being challenged by different perspectives, values, and ideas (Pascarella, Edison, Nora, Hagedorn, & Terenzini, 1996). The ODC has internal consistence reliabilities in the present study ranging from .83 to .87. In previous research, precollege ODC scores have significantly predicted the

likelihood of participating in a racial/cultural workshop during the first year of college (Whitt, Edison, Pascarella, Terenzini, & Nora, 2001). In the present study, precollege ODC scores correlated .37 with a measure of students' experiences and interactions with diverse others and diverse ideas during the first year of college.

Leadership. This outcome dimension was assessed with the 68-item, revised version II of the Socially Responsible Leadership Scale (SRLS). The SRLS measures the eight dimensions of Astin's Social Change Model of leadership development (Astin, A., Astin, H., Boatsman, Bonous-Hammarth, Chambers, Goldberg, et al., 1996). According to this model, leadership is a collaborative group process directed toward promoting positive social change in an organization or community (Tyree, 1998). A person who demonstrates strong socially responsible leadership capabilities is self-aware, acts in accordance with personal values and beliefs, invests time and energy in activities that he or she believes are important, works with diverse others to accomplish common goals, has a sense of civic and social responsibility, and desires to make world a better place. The SRLS was developed specifically to measure leadership in college students. The instrument has eight scales corresponding to the eight dimensions of leadership specified in the Astin model (Astin, et al., 1996; Dugan, 2006). The eight scales are: Consciousness of Self (being aware of the values, emotions, attitudes, and beliefs that motivate one to take action); Congruence (thinking, feeling, and behaving with consistency, genuineness, authenticity, and honesty toward others); Commitment (intensity and duration in relation to a person, idea, or activity—the energy and passion that propels one to act); Collaboration (working with others in a common effort); Common Purpose (working with others within a shared set of aims and values); Controversy with Civility (recognizing two fundamental realities of any group effort, that (a) differences of viewpoint are inevitable and valuable, and (b) such differences must be aired

openly and with respect and courtesy; Citizenship (believing in a process whereby a person or group is responsibly connected to the environment and the community); and Change (adapting to continuously evolving environments and situations, while maintaining the primary functions of the group).

The internal consistency reliabilities for the eight subscales of the SRLS in the present study ranged from .77 to .88. The various scales of the SRLS have been shown to discriminate between involved and non-involved undergraduate students in community service, student organizational membership, formal leadership programs, and positional leadership roles (Dugan, 2006). Additional research by Rubin (2000) has demonstrated that undergraduates identified as "emerging student leaders" tend to score significantly higher on the SRLS congruency, collaboration, common purpose, citizenship, and change scales than a control group of students not identified as "emerging student leaders."

Moral Character. We assessed the outcome dimension of "Moral Character" with the Defining Issues Test 2 (DIT2). The DIT2 is a revised version of James Rest's original DIT from 1979 that measures one component of moral development, known as moral judgment or reasoning (Rest, Narvaez, Thoma, & Bebeau, 1999). The DIT2 presents several dilemmas about social problems, such as should a starving man steal food for his family from someone who is hoarding resources? After each, a series of 12 items representing different issues that might be raised by the problem are presented. For example, in the scenario described above, the items include such questions as: "Would stealing bring about more total good for everybody concerned or wouldn't it? Shouldn't the community's laws be upheld?" In response to the scenario and questions, respondents are asked to do three things:

1. make an action choice (i.e., yes, one should steal or no, one should not steal);

2. rate the series of 12 items in terms of their importance in making a decision about the scenario; and
3. rank the top four most important items.

The DIT2 produces two scores on which the study focused. The first is the P-score, which represents the degree to which an individual uses higher order (principled/post-conventional) moral reasoning in resolving moral issues presented in each scenario. The P-score is the proportion of items selected that appeal to moral ideas and/or theoretical frameworks for resolving complex moral issues—specifically, items that appeal to consensus building procedures, insistence on due process, safeguarding minimal basic rights, and organizing social arrangements in terms of appealing to ideals. The P-score has internal consistency reliabilities ranging from .74 to .77 (Rest, et al., 1999; University of Minnesota, n.d.). The second score from the DIT2 particularly salient to the WNSLAE was the N2-score. As with the P-score, the N2-score reflects the extent to which one exhibits high order moral reasoning. However, the relatively new N2-score also reflects the extent to which one rejects ideas because they are simplistic or biased (Bebeau & Thoma, 2003). The internal consistency reliability for the N2-score range from .77 to .81 (Rest, et al., 1999; University of Minnesota, n.d.).

An extensive body of evidence supports the validity of the DIT in predicting principled ethical behavior in a number of areas. These include: resistance to cheating, peer pressure, and unlawful or oppressive authority; whistle-blowing on corruption; the keeping of contractual promises; helping behavior; community involvement; ethical behavior in several professions; clinical performance in nursing students; and social/political activism (see Pascarella & Terenzini, 1991, 2005, for a synthesis of this body of evidence, including citations to original studies). The vast majority of the validity data on the DIT is based on the P-score. However,

correlations between the P-score and the N2-score ranged from .91 to .92, so it appears to be the case that they are measuring essentially the same construct.

Additional WNSLAE Outcome/Dependent Measures

In addition to the outcome measures selected to tap the six dimensions of the King, et al. (2007) conceptual model of liberal arts outcomes, the WNSLAE also considered several additional measures thought to be germane to student development within the context of a liberal arts education. These included four scales designed to measure orientations toward life and career, a measure of political orientation, and a measure of academic motivation.

Orientations Toward Life and Career. To measure student development in orientations toward life and career, the WNSLAE received permission to use 21 items from the Cooperative Institutional Research Program (CIRP) Survey developed by the Higher Education Research Institute at the University of California at Los Angeles. A factor analysis of these 21 items yielded four meaningful scales:

- Contributions to the Arts Scale—a three-item scale in which respondents identify how important (ranging from not important to essential) it is for them to contribute to the arts. Contributing to the arts included "becoming accomplished in the performing arts," "writing original works," or "creating artistic work." The internal consistency reliability for the scale is .69.
- Political and Social Involvement—an eleven-item scale in which respondents identify how important (ranging from not important to essential) it is for them to be involved politically and socially in communities. Political and social involvement included such activities or goals as "influencing the political structure," "influencing social values,"

"becoming a community leader," and "volunteering in my community." The internal consistency reliability for the scale ranges from .80 to .83.

- Professional Success—a five-item scale which asks respondents to indicate how important (ranging from not important to essential) it is to them to be successful in a profession. Professional success includes such goals as "having administrative responsibility for the work of others," "becoming successful in a business of my own," and "working in a prestigious occupation." The internal consistency reliability for the scale ranges from .75 to .76.
- Contribution to Science—a two-item scale which asks respondents to indicate how important it is for them to contribute to advances in scientific fields. Contribution to science includes "making a theoretical contribution to science" and "working to find a cure for a disease or illness." The internal consistency reliability for the scale ranges from .70 to .76

Political Orientation. To measure political orientation, the WNSLAE used a single item that asked students "How would you characterize your political views?" There were five response options: "far left," "liberal," "middle-of-the-road," "conservative," and "far right."

Academic Motivation. To measure academic motivation, the WNSLAE used an eight-item scale in which respondents were asked to indicate the extent to which they agree or disagree (ranging from strongly agree to strongly disagree) with statements about their academic motivation. Academic motivation includes: a willingness to work hard to learn material even if it doesn't lead to a higher grade, the importance of getting good grades, reading more for a class than required because the material was interesting, enjoyment of academic challenge, and the importance of academic experiences in college. The internal consistency reliability for the scale

ranges from .69 to .74. Precollege Academic Motivation scores in the WNSLAE data had statistically significant, positive (though modest) correlations with end-of-first-year graduate degree plans (.19) and self-reported grades (.15). Net of such influences as precollege Need for Cognition (NFC) and precollege Positive Attitude Toward Literacy (PATL), a student's level of academic motivation at the beginning of college was a significant, positive predictor of end-of-first-year scores on both measures (NFC and PATL) of inclination to inquire and lifelong learning.

With two exceptions, each WNSLAE outcome/dependent measure discussed above was completed by all 3,081 participants during both the initial data collection in fall 2006 and the follow-up data collection in spring 2007. The two exceptions were the CAAP Critical Thinking Test and the Defining Issues Test. Each of these instruments took at least 40 minutes to complete, and because we were concerned with the amount of time required of students during each data collection, the CAAP Critical Thinking Test and the Defining Issues Test were not taken by all participants. Rather, during the first data collection the study participants were randomly divided into two approximately equal samples. The CAAP Critical Thinking Test was then taken by one random sample during both data collections and the Defining Issues Test taken during both data collections by the other random sample. Of the 3,081 students participating in both data collections, 1,485 had useable responses on the CAAP and 1,584 had useable responses on the DIT.

WNSLAE Measures of Good Practices/Liberal Arts Experiences

As previously indicated, the WNSLAE was primarily concerned with understanding the conditions and experiences that constituted an influential liberal arts education. One of these conditions was the type of institution attended. A major focus of WNSLAE was to estimate the

net or unique influence of attending a liberal arts college (vs. another type of postsecondary institution) on the outcome measures listed and described above. However, an equally important purpose of the study was to estimate the extent to which specific academic and nonacademic experiences during the first year of postsecondary education contributed to change or growth on each outcome.

In selecting academic and nonacademic experiences on which to focus, the WNSLAE cast a purposefully wide net. However, we were conceptually guided by a body of literature and evidence that specifies particular "good practices" in undergraduate education that are linked to personal and intellectual growth during college (Astin, 1993; Chickering & Reisser, 1993; Kuh, Schuh, Whitt, & Associates, 1991; Kuh, et al., 2005; Pascarella & Terenzini, 1991, 2005). To measure these "good practices," WNSLAE selected and adopted empirically vetted scales and items from the National Study of Student Learning (Cruce, Wolniak, Seifert, & Pascarella, 2006; Pascarella, Wolniak, Seifert, Cruce, & Blaich, 2005) and the National Survey of Student Engagement (Pascarella, et al., 2006). These scales and items are designed to tap a range of "good practices" that includes such dimensions as student-faculty interaction, active learning/time on task, quality of teaching, prompt feedback from faculty, cooperative learning, high academic expectations, diversity experiences, influential interactions with other students, and integrative experiences. Extensive evidence exists to indicate that, even in the presence of statistical controls for important confounding influences, the good practice dimensions measured by these dimensions are significantly linked to student cognitive and personal development during college (see Cruce, et al., 2006; Pascarella, et al., 2005; Pascarella, et al., 2006, for reviews of this body of evidence, including specific citations to original studies).

Interestingly, though perhaps not surprisingly, these "good practice" indicators are not only linked to student growth in college, they are also strongly linked to institutional type. Net of other influences, undergraduate students at liberal arts colleges appear to experience a wide range of good practices significantly and substantially more often than their counterparts at research universities or regional universities (Pascarella, Cruce, Wolniak, & Blaich, 2004; Pascarella, et al., 2005). Consequently, the WNSLAE uses the term "good practices/liberal arts experiences" to describe these salient influences on student development.

A recent study by Cruce, and colleagues (2006) indicates that the good practice/liberal arts experience dimensions described above tend to be substantially correlated with each other. When Cruce and his colleagues submitted the good practice scales from the National Study of Student Learning (NSSL) to a factor analysis, three clear underlying factors were yielded which were termed: "effective teaching and interaction with faculty," "interaction with peers," and "challenge/high expectations." Scales developed from this factor structure substantially reduced collinearity among the good practice dimensions, which, in turn, made their estimate net impact on change in measures of cognitive development, learning orientations, and graduate degree plans during the first year of college less ambiguous.

Consistent with the data-reduction procedures of Cruce, et al., (2006), the WNSLAE also submitted its numerous good practice/liberal arts experience scales and items to a similar factor analytic procedure. Because WNSLAE adapted scales from both the NSSL and NSSE, it is not particularly surprising that our underlying factor structure was somewhat more complex than that uncovered by Cruce and his colleagues. Six factors appeared to underlie the individual measures of good practices/liberal arts experiences in the WNSLAE. They were titled: "Good teaching and high quality interactions with faculty," "Academic challenge and high expectations," "Diversity

experiences," "Influential interactions with peers," "Frequency of interactions with faculty/professional staff," and "Cooperative learning." Scales were constructed for each factor by first standardizing each item from the individual good practice/liberal arts experience dimensions loading on the factor, and then computing the mean score. Only those respondents with 60% of items completed received a score across items.

Good teaching and high quality interactions with faculty was a 23-item scale that combined items from four subscales: Faculty interest in teaching and student development (e.g., the extent to which faculty are interested in helping students grow in more than just academic areas, the extent to which faculty are generally interested in teaching, and the extent to which faculty are willing to spend time outside of class to discuss issues of interest and importance to students); Prompt feedback (e.g., how often faculty informed students of level of performance in a timely manner, how often faculty checked to see if students had learned the material well before going on to new materials); Quality and impact of nonclassroom interactions with faculty (e.g., extent to which nonclassroom interactions with faculty have had an impact on: intellectual growth and interest in items; personal growth, values, and attitudes; and career goals and aspirations); and Overall exposure to clear and organized instruction (e.g., frequency that faculty give clear explanation, frequency that faculty make good use of examples and illustration to explain difficult points, frequency that class time was used effectively, frequency that course goals and requirements were clearly explained). The internal consistency reliability for the 23-item scale is .92.

Academic challenge and high expectations was a 31-item scale that combined items from four subscales: Academic challenge and effort (e.g., how often one worked harder than one thought he or she could to meet an instructor's standards or expectations, number of hours a

week spent preparing for class, extent to which one's institution emphasizes spending significant amounts of time studying and on academic work, number of assigned textbooks, books, or book-length packs of course readings one read during current year); Frequency of higher-order exams and assignments (e.g., how often exams or assignments require students to: write essays, compare or contrast topics or ideas from a course, argue for or against a particular point of view and defend an argument); Challenging classes and high faculty expectations (e.g., how often faculty: asked challenging questions in class; challenged students' ideas in class; asked students to argue for or against a particular point of view; asked students to point out any fallacies in basic ideas, principles, or points of view presented in the course); and Integrating ideas, information, and experiences (e.g., extent to which one agrees that courses have helped him or her understand the historical, political, and social connections of past events; how often one has worked on a paper or project that required integrating ideas or information from various sources; how often one has put together ideas or concepts from different courses when completing assignments or during class discussions). The internal consistency reliability for the 31-item scale is .88.

Diversity experiences was a 9-item scale that combined items from two subscales: Diversity experiences (e.g., extent to which one's institution encourages contact among students from different economic, social, and racial or ethnic backgrounds; how often one had serious conversations with students of a different race or ethnicity than one's own; how often one participated in a racial or cultural awareness workshop during the academic year); and Meaningful discussions with diverse peers (e.g., how often one had meaningful and honest discussions about issues related to social justice with diverse students, how often one had discussions regarding intergroup relations with diverse students). The internal consistency reliability for the 9-item scale is .80.

Influential interactions with peers was a 9-item scale that combined items from two subscales: Positive peer interactions (e.g., the student friendships one has developed at the institution have been personally satisfying; interpersonal relationships with other students have had a positive influence on one's intellectual growth and interest in ideas; interpersonal relationships with other students have had a positive influence on one's personal growth, attitudes, and values) and Co-curricular involvement (number of hours per week spent in co-curricular activities). The 9-item scale has an internal consistency reliability of .85.

Frequency of interactions with faculty/professional staff was a 9-item scale that combined items from two subscales: Frequency of interactions with faculty (e.g., how often one discussed grades or assignments with an instructor, how often one worked with faculty members on activities other than coursework such as committees, orientation, student life activities) and Frequency of interactions with student affairs staff (e.g., how often one discussed a personal problem or concern with student affairs professions; how often one worked on out-of-class activities, such as committees, orientation, student life activities with student affairs professions; how often one discussed grades or assignments with student affairs professionals). The internal consistency reliability for the 9-item scale is .83.

Cooperative learning was a 4-item scale (e.g., in classes, students taught each other in addition to faculty teaching; participation in one or more study groups outside of class; how often one worked with other students on projects outside of class). The scale had an internal consistency reliability of .70.

Detailed operational definitions of all good practice/liberal arts experience variables (including subscales, specific items, and response options) are shown in Appendix A.

WNSLAE Control Variables

A particular methodological strength of the WNSLAE was its longitudinal design. This permitted us to introduce a wide range of statistical controls, not only for student background and precollege traits and experiences, but also for other experiences during the first year of college. Our control variables included the following:

Student background/precollege traits:

- Sex
- Race/Ethnicity (African-American, Hispanic, American Indian, Asian/Pacific Islander, Caucasian, Resident Alien, Race not reported)
- Age
- English as a second language (English is one's second language vs. English is one's first language ; did not respond)
- Average parental education (a ten-point scale from "did not finish high school" to "doctorate")
- Tested academic preparation (ACT score, SAT equivalent score, or COMPASS equivalent score for community college students; score provided by each institution)
- Educational degree plans (graduate degree or higher vs. less than a graduate degree; did not respond)
- Institution attended was first choice vs. not first choice
- Academic motivation (An eight-item scale previously described as an outcome measure. We employed the precollege measure of academic motivation as a control in all analyses. The internal consistency reliability for the scale is .69.)
- Have dependent children vs. do not have dependent children

High school experiences:

- High school racial composition (All White or Mostly All White vs. Other)
- High school involvement (A seven-item scale that measured involvement during high school in such areas as: studying with a friend, socializing with friends, community service/volunteer work, talking with teachers outside of class, and extracurricular activities. The internal consistency reliability for the scale is .58.)

Other college experiences:

- Live in campus housing vs. off campus or a fraternity/sorority
- Full-time enrollment vs. part-time enrollment
- Number of hours per week during college one worked on campus (eight response options from "zero" to "more than 30 hours")
- Number of hours per week during college one worked off campus (eight response options from "zero" to "more than 30 hours")
- Was a member of a fraternity or sorority vs. was not Greek-affiliated
- Participated in an intercollegiate sport vs. did not participate in an intercollegiate sport
- Number of courses during the current academic year taken in each of nine general areas: "Fine Arts, Humanities, and Languages" (e.g., art, music, philosophy, religion, history); "Mathematics/Statistics/Computer Science"; Natural Sciences" (e.g., chemistry, physics); "Social Science" (e.g., anthropology, economics, psychology, political science, sociology); "Allied Health" (e.g., nursing, physical therapy); "Business"; "Education"; "Engineering"; "Other Pre-professional" (e.g., architecture, agriculture, journalism). Response options ranged from "0 Courses" to "5 or More Courses."

Detailed operational definitions of all control variables (including specific items and response options) are shown in Appendix B.

Data Analyses

The WNSLAE employed various forms of multiple regression analysis, configured in general regression models that were consistent with the basic questions or purposes of the study. The first purpose of the study was to determine the extent to which the WNSLAE data replicated previous findings indicating that, net of important confounding influences, students attending liberal arts colleges were significantly more likely to experience good practices/liberal arts experiences than students at other types of institutions (Pascarella, et al., 2004, 2005). To accomplish this, we estimated the total and direct effects of attending a liberal arts college (vs. attending another type of institution) on student exposure to good practices/liberal arts experiences. The independent variable in these analyses was institutional type, operationally defined as three dummy (1, 0) variables representing attendance at a research university, a regional institution, or a community college—with the comparison group always being attendance at a liberal arts college. The dependent variables were the six factorially derived WNSLAE good practice/liberal arts experience variables previously described (i.e., Good teaching and high quality interactions with faculty, Academic challenge and high expectations, Diversity experiences, Influential interactions with peers, Frequency of interactions with faculty/professional staff, and Cooperative learning).

These analyses were conducted in two stages. In the first stage, we employed reduced-form equations (Alwin & Hauser, 1975) to estimate the total effect of attending a liberal arts college on exposure to good practices/liberal arts experiences. Figure 1 visually portrays the conceptual model guiding these analyses. As the figure indicates, each of the six good

practice/liberal arts experience scales was regressed on the three dummy variables representing institutional type plus all student background/precollege traits and high school experiences specified in the section on control variables. In the second stage, we estimated the direct (or unmediated) effects of attending a liberal arts college on exposure to good practices/liberal arts experiences. This is visually portrayed in Figure 2. As the figure illustrates, the direct effects estimates of attendance at a liberal arts college are obtained by simply adding the battery of other college experiences to the total effects equations. The results from this set of equations also permit one to estimate the indirect effects of attendance at a liberal arts college on exposure to good practices/liberal arts experiences (mediated through other college experiences) by subtracting the estimated direct effect from the total effect (Alwin & Hauser, 1975).

Figures 1 and 2 about here

The second purpose of the WNSLAE study was to estimate the net total and direct effects of institutional type, and the net direct effects of exposure to good practices/liberal arts experiences, on each liberal arts outcome. These analyses were also carried out in two stages. In the first stage, we estimated the total effects of attendance at a liberal arts college on each WNSLAE outcome measure, consistent with the conceptual model illustrated in Figure 3. As the figure shows, we regressed each specific WNSLAE outcome on the dummy variables representing institutional type plus all student background/precollege traits, high school experiences, and the appropriate pretest score for each specific outcome. The conceptual model guiding the second stage of the analyses is shown in Figure 4. As the figure indicates, other college experiences and scores on the good practices/liberal arts experiences scales were added to the total effects equations. Solutions to these equations not only provided the net direct, or unmediated, effects of attendance at a liberal arts college on each outcome, it also permitted

estimation of the indirect effects of attendance at liberal arts colleges by subtracting the direct effect from the total effect. Similarly, these equations also yielded the net direct effects of the good practices/liberal arts experience scales on each WNSLAE outcome measure. We hypothesized only direct, unmediated effects of good practices/liberal arts experiences on each outcome.

Figures 3 and 4 about here

A final purpose of the WNSLAE study was to determine if the net effects or each outcome of attending a liberal arts college and exposure to good practices/liberal arts experiences differed in magnitude for different kinds of students. That is, are the effects of liberal arts colleges and good practices/liberal arts experiences conditional rather than general? To test for the presence of conditional effects, we considered four student characteristics: sex, race/ethnicity, tested academic preparation at entrance to college (i.e., ACT/SAT/COMPASS score), and the pretest score on each outcome measure. We formed cross-product terms between these four student characteristics on the one hand, and both institutional type and the good practice/liberal arts experience variables on the other. These cross-product terms were then added to the direct effects equations for each outcome. (The conceptual model for this set of equations was previously illustrated in Figure 4.) A statistically significant increase in the explained variance (R^2) would indicate the presence of conditional effects. This being the case, the sample could then be disaggregated by the appropriate student characteristic (e.g., men vs. women, high vs. low pretest score) and the direct effects equations rerun for each subsample to determine the nature of the conditional effect.

LIMITATIONS

As with nearly all multi-institutional studies, the WNSLAE data have several limitations that should be kept in mind when interpreting the findings. First, although the overall sample included a broad range of different kinds of postsecondary institutions from 11 different states, the inclusion of only 19 institutions and the fact that institutions were not selected randomly means that one cannot necessarily generalize the results to the population of all two-year and four-year institutions in the United States. Indeed, because a major purpose of the WNSLAE was to estimate the impacts of liberal arts colleges and liberal arts education, liberal arts colleges were purposefully over-sampled in the study. In turn, due to lower numbers of students of color in private liberal arts colleges, our overall sample, while weighted to be representative of the actual institutional populations in the study, had a larger representation of Caucasian students (85%) than is found across the population of American postsecondary students.

A second limitation is the fact that not all students who participated in the first (precollege) data collection participated in the second (follow-up) data collection. The 68.5% persistence rate in the WNSLAE from the first to second data collections is quite consistent with other large longitudinal studies requiring a substantial amount of participation in terms of time and intellectual effort (see for example, the National Study of Student Learning, Pascarella, Edison, Nora, Hagedorn, & Terenzini, 1998). However, attrition from the first to second data collections is a major, if perhaps unavoidable, limitation of the study. Our weighting procedures adjusted the final sample for respondent bias by sex, race/ethnicity, tested precollege academics, and institution; but this in no way guarantees that those students who dropped out of study after the first data collection would have responded in the same way as their counterparts who persisted in the study from the first to second data collections.

Finally, although our selection of outcome measures was guided by a persuasive conceptual model of the purposes of liberal arts education, the specific outcomes we measured are certainly not the only conceivable outcomes that could be assessed. Consequently, our findings should be regarded as specific to the outcomes considered. We cannot necessarily generalize our findings with regard to important influences to other measurable outcomes of liberal arts education.

References

Alwin, D., & Hauser, R. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*, 37-47.

American College Testing Program (ACT). (1991). *CAAP technical handbook*. Iowa City, IA: Author.

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

Astin, A., Astin, H., Boatsman, K., Bonous-Hammarth, M., Chambers, T., Goldberg, S., et al. (1996). *A social change model of leadership development: Guidebook (Version III)*. Los Angeles: University of California at Los Angeles, Higher Education Research Institute.

Bebeau, M., & Thoma, S. (2003). *Guide for the DIT-2*. Minneapolis, MN: Center for the Study of Ethical Development, University of Minnesota.

Bray, G., Pascarella, E., & Pierson, C. (2004). Postsecondary education and some dimensions of literacy development: An exploration of longitudinal evidence. *Reading Research Quarterly, 39*, 306-330.

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

Chickering, A., & Reisser, L. (1993). *Education and identity*. San Francisco: Jossey-Bass.

Cruce, T., Wolniak, G., Seifert, T., & Pascarella, E. (2006). Impacts of good practices on cognitive development, learning orientations, and graduate degree plans during the first year of college. *Journal of College Student Development*, *47*, 365-383.

Dugan, J. (2006). Explorations using the social change model: Leadership development among college men and women. *Journal of College Student Development, 47,* 217-225.

Elias, S., & Loomis, R. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance. *Journal of Applied Social Psychology*, *32*, 1687-1717.

Fuertes, J., Miville, M., Mohr, J., Sedlacek, W., & Gretchen, D. (2000). Factor structure and short form of the Miville-Guzman Universality-Diversity Scale. *Measurement and Evaluation in Counseling and Development*, *33*, 157-169.

Jones, S., & McEwen, M. (2000). A conceptual model of multiple dimensions of identity. *Journal of College Student Development*, *41*, 405-414.

Keyes, C., Shmotkin, D., & Ryff, C. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality and Social Psychology, 82,* 1007-1022.

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

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

Kuh, G., Kinzie, J., Schuh, J., Whitt, E., & Associates. (2005). *Student success in college: Creating conditions that matter*. San Francisco: Jossey-Bass.

Kuh, G., Schuh, J., Whitt, E., & Associates. (1991). *Involving colleges: Successful approaches to fostering student learning and development outside the classroom*. San Francisco: Jossey-Bass.

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

Pascarella, E., Bohr, L., Nora, A., & Terenzini, P. (1995). Cognitive effects of 2-year and 4-year colleges: New evidence. *Educational Evaluation and Policy Analysis, 17,* 83-96.

Pascarella, E., Cruce, T., Umbach, P., Wolniak, G., Kuh, G., Carini, R., Hayek, J., Gonyea, R., & Zhao, C. (2006). Institutional selectivity and good practices in undergraduate education: How strong is the link? *Journal of Higher Education, 77,* 251-285.

Pascarella, E., Cruce, T., Wolniak, G., & Blaich, C. (2004). Do liberal arts colleges really foster good practice in undergraduate education? *Journal of College Student Development, 45,* 57-74.

Pascarella, E., 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., Edison, M., Nora, A., Hagedorn, L., & Terenzini, P. (1998). Does work inhibit cognitive development during college? *Educational Evaluation and Policy Analysis, 20,* 75-93.

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

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

Pascarella, E., Wolniak, G., Seifert, T., Cruce, T., & Blaich, C. (2005). *Liberal arts colleges and liberal arts education: New evidence on impacts*. San Francisco: Jossey-Bass/ASHE.

Rest, J., Narvaez, D., Thoma, S., & Bebeau, M. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91,* 644-659.

Rubin, J. (2000). *The emerging leaders: An evaluation of the social change model of leadership*. Unpublished doctoral dissertation, The Union Institute.

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.

Tyree, T. (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.

University of Minnesota, Center for the Study of Ethical Development (n.d.). *New index (N2).* Retrieved January 14, 2008 from http://www.centerforthestudyofethicaldevelopment.net/New%20Index.htm.

Whitt, E., Edison, M., Pascarella, E., Terenzini, P., & Nora, A. (2001). Influences on students' openness to diversity in the second and third years of college. *Journal of Higher Education, 72,* 172-204.

Appendix A – Operational Definitions of Good Practice/Liberal Arts Experience Scales

<u>*Good teaching and high quality interactions with faculty*</u> is a 23-item scale with an alpha (internal consistency) reliability of .92 that combines items from four subscales. The scales and constituent items are:

- *Faculty interest in teaching and student development*
    - Most faculty with whom the respondent had contact are genuinely interested in students.
    - Most faculty with whom the respondent had contact are interested in helping students grow in more than just academic areas.
    - Most faculty with whom the respondent had contact are outstanding teachers.
    - Most faculty with whom the respondent had contact are genuinely interested in teaching.
    - Most faculty with whom the respondent had contact are willing to spend time outside of class to discuss issues of interest and importance to students.

  For each item, the response options are "strongly agree," "agree," "neutral," "disagree," or "strongly disagree." The alpha reliability of this scale is .85.
- *Prompt feedback*
    - How often faculty informed the respondent of their level of performance in a timely manner.*
    - How often faculty checked to see if the respondent had learned the material well before going on to new material.*
    - During the current school year, how often the respondent received prompt written or oral feedback from faculty regarding their academic performance.**

*The response options for this item are “very often,” “often,” “sometimes,” “rarely,” or “never.”

**The response options for this item are “very often,” “often,” sometimes,” or “never.” The alpha reliability of this scale is .68.

- *Quality of non-classroom interactions with faculty*
    - Non-classroom interactions with faculty have had a positive influence on personal growth, values, and attitudes.
    - Non-classroom interactions with faculty have had a positive influence on intellectual growth and interest in ideas.
    - Non-classroom interactions with faculty have had a positive influence on career goals and aspirations.
    - Since coming to this institution, the respondent has developed a close, personal relationship with at least one faculty member
    - The respondent is satisfied with the opportunities to meet and interact informally with faculty members.

  For each item, the response options are “strongly agree,” “agree,” “neutral,” “disagree,” or “strongly disagree.” The alpha reliability of this scale is .85.

- *Overall exposure to clear and organize instruction*
    - Faculty gave clear explanations.
    - Faculty made good use of examples and illustrations to explain difficult points.
    - Faculty effectively reviewed and summarized the material.
    - Faculty interpreted abstract ideas and theories clearly.
    - Faculty gave assignments that helped in learning the course material.

- The presentation of material was well organized.
- Faculty were well prepared for class.
- Class time was used effectively.
- Course goals and requirements were clearly explained.
- Faculty had a good command of what they were teaching.

For each item, the response options are "very often," "often," "sometimes," "rarely," or "never." The alpha reliability of this scale is .89.

Academic challenge and high expectations is a 31-item scale with an alpha reliability of .88 that combines items from four subscales. The scales and constituent items are:

- *Academic challenge and effort*
  - During current year, the number of assigned textbooks, books, or book-length packs of course readings that the respondent read. The response options for this item "none," "between 1 and 4," "between 5 and 10," "between 11 and 20," or "more than 20."
  - During current year, the number of written papers or reports between 5 and 19 pages that the respondent wrote. The response options for this item "none," "between 1 and 4," "between 5 and 10," "between 11 and 20," or "more than 20."
  - In a typical week, the number of problem sets that take the respondent more than an hour to complete. The response options for this item are "none," "1-2," "3-4," "5-6," or "more than 6."
  - The extent to which examinations during the current school year challenged the respondent to do his/her best work. There were seven response options for this item ranging from "very little" to "very much."

- Number of hours per week the respondent spends preparing for class. The response options for this item are “0 hours,” “1-5 hours,” “6-10 hours,” “11-15 hours,” “16-20 hours,” “21-25 hours,” “26-30 hours,” or “more than 30 hours.”
- Extent to which the respondent's institution emphasizes spending significant amounts of time studying and on academic work. The response options for this item are “very little,” “some,” “quite a bit,” or “very much.”
- During current school year, how often the respondent asked questions in class or contributed to class discussions.*
- During current school year, how often the respondent made a class presentation.*
- During current school year, how often the respondent prepared two or more drafts of a paper or assignment before turning it in.*
- During current school year, how often the respondent came to class without completing readings or assignments (reverse-coded).*
- During current school year, how often the respondent worked harder than he/she thought he/she could to meet an instructor's standards or expectations.*

*The response options for this item are “very often,” “often,” “sometimes,” or “never.”

The alpha reliability of this scale is .65.

- *Frequency of higher-order exams and assignments*
  - Exams or assignments required the respondent to write essays.
  - Exams or assignments required the respondent to use course content to address a problem not presented in the course.

- Exams or assignments required the respondent to compare or contrast topics or ideas from a course.
- Exams or assignments required the respondent to point out the strengths and weaknesses of a particular argument or point of view.
- Exams or assignments required the respondent to argue for or against a particular point of view and defend an argument.

For each item, the response options are "very often," "often," "sometimes," "rarely," or "never." The alpha reliability of this scale is .76.

- *Challenging classes and high faculty expectations*
  - Faculty asked challenging questions in class.
  - Faculty asked the respondent to show how a particular course concept could be applied to an actual problem or situation.
  - Faculty asked the respondent to point out any fallacies in basic ideas, principles, or points of view presented in the course.
  - Faculty asked respondent to argue for or against a particular point of view.
  - Faculty challenged the respondent's ideas in class.
  - Students challenged each others ideas in class.

  For each item, the response options are "very often," "often," "sometimes," "rarely," or "never." The alpha reliability of this scale is .82.
- *Integrating ideas, information, and experiences*
  - The extent the respondent agrees that courses have helped them understand the historical, political, and social connections of past events.*

- The extent the respondent agrees that courses have helped them see the connections between intended career and how it affects society.*
- The extent the respondent agrees that out-of-class experiences have helped them connect what was learned in the classroom with life events.*
- The extent the respondent agrees that out-of-class experiences have helped them translate knowledge and understanding from the classroom into action.*
- During current school year, how often the respondent worked on a paper or project that required integrating ideas or information from various sources.**
- During current school year, how often the respondent put together ideas or concepts from different courses when completing assignments or during class discussions.**
- During current school year, how often the respondent discussed ideas from readings or classes with others outside of class (students, family members, co-workers, etc.).**
- During current year, the time the respondent spent synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships.***
- During current year, the time the respondent spent making judgments about the value of information, arguments, or methods, such as examining how others gathered and interpreted data and assessing the soundness of their conclusions.***

*The response options for this item are “strongly agree,” “agree,” “neutral,” “disagree,” or “strongly disagree.”

**The response options for this item are “very often,” “often,” “sometimes,” or “never.”

***The response options for this item are “very little,” “some,” “quite a bit,” or “very much.”

The alpha reliability of this scale is .76.

Diversity experiences is a 9-item scale with an alpha reliability of .80 that combines items from two subscales. The scales and constituent items are:

- *Diversity experiences*
    - How often the respondent attended a debate or lecture on a current political/social issue during this academic year.*
    - How often the respondent had serious discussions with student affairs professionals whose political, social, or religious opinions were different from own.*
    - Extent to which the respondent's institution encourages contact among students from different economic, social, and racial or ethnic backgrounds.**
    - During current school year, how often the respondent had serious conversations with students from a different race or ethnicity.***
    - During current school year, how often the respondent had serious conversations with students who are very different from the respondent in religious beliefs, political opinions, or personal values.***
    - How often the respondent participated in a racial or cultural awareness workshop during this academic year.*

*The response options for this item are “very often,” “often,” “sometimes,” “rarely,” or “never.”

**The response options for this item are “very little,” “some,” “quite a bit,” or “very much.”

***The response options for this item are “very often,” “often,” “sometimes,” or “never.”

The alpha reliability of this scale is .65.

- *Meaningful discussions with diverse peers*
    - How often the respondent had discussions regarding inter-group relations with diverse students while attending this college.
    - How often the respondent had meaningful and honest discussions about issues related to social justice with diverse students while attending this college.
    - How often the respondent shared personal feelings and problems with diverse students while attending this college.

    For each item, the response options are “very often,” “often,” sometimes,” “rarely,” or “never.”

    The alpha reliability of this scale is .82.

Influential interactions with peers is a 9-item scale with an alpha reliability of .85 that combines items from two subscales. The scales and constituent items are:

- *Co-curricular involvement (single item)*
    - The number of hours per week respondent spends participating in co-curricular activities. Response options for this item are “0 hours,” “1-5 hours,” “6-10 hours,” “11-15 hours,” “16-20 hours,” “21-25 hours,” “26-30 hours,” or “more than 30 hours.”

- *Positive peer interactions*
    - The respondent has developed close personal relationships with other students.*
    - The student friendships the respondent has developed at this institution have been personally satisfying.*
    - Interpersonal relationships with other students have had a positive influence on the respondent's personal growth, attitudes, and values.*
    - Interpersonal relationships with other students have had a positive influence on the respondent's intellectual growth and interest in ideas.*
    - It has been difficult for the respondent to meet and make friends with other students (reverse-coded).*
    - Few of the students the respondent knows would be willing to listen to and help the respondent with a personal problem (reverse-coded).*
    - Most students at this institution have values and attitudes different from the respondent (reverse-coded).*
    - The respondent's quality of relationships with other students.**

    *The response options for this item are "strongly agree," "agree," "neutral," "disagree," or "strongly disagree."

    **The response options for this item are "unfriendly, unsupportive, sense of alienation" or "friendly, supportive, sense of belonging."

    The alpha reliability of this scale is .87.

Frequency of interactions with faculty/professional staff is a 9-item scale with an alpha reliability of .83 that combines items from two subscales. The scales and constituent items are:

- *Frequency of interactions with faculty*

- Discussed grades or assignments with an instructor.
- Talked about career plans with a faculty member or advisor.
- Discussed ideas from readings or classes with faculty members outside of class.
- Worked with faculty members on activities other than coursework (committees, orientation, student life activities, etc.).

For each item, the response options are "very often," "often," "sometimes," or "never." The alpha reliability of this scale is .70.

- *Frequency of interactions with student affairs staff*
  - How often the respondent discussed a personal problem or concern with student affairs professionals.
  - How often the respondent worked on out-of-class activities (e.g., committees, orientation, student life activities) with student affairs professionals.
  - How often the respondent talked about career plans with student affairs professionals.
  - How often the respondent discussed ideas from readings or classes with student affairs professionals.
  - How often the respondent discussed grades or assignments with student affairs professionals.

  For each item, the response options are "very often," "often," "sometimes," "rarely," or "never." The alpha reliability of this scale is .84.

Cooperative learning is a 4-item scale with an alpha reliability of .70. Items included in this scale ask the student how frequently:

- In respondent's classes, students taught each other in addition to faculty teaching.*

- Faculty encouraged the respondent to participate in study groups outside of class.*
- The respondent participated in one or more study group(s) outside of class.*
- During current school year, how often the respondent worked with other students on projects outside of class.**

*The response options for this item are “very often,” “often,” sometimes,” “rarely,” or “never.”

**The response options for this item are “very often,” “often,” “sometimes,” or “never.”

Appendix B: Operational Definitions of Control Variables

Student background/pre-college traits:

- Sex (coded as 1 = male, 2 = female, 9 = not reported)
- Race/Ethnicity (coded as 1 = Nonresident alien, 2 = Black, non-Hispanic, 3 = American Indian/Alaska Native, 4 = Asian/Pacific Islander, 5 = Hispanic, 6 = White, non-Hispanic, 7 = Race/ethnicity unknown, 9 = Race not reported)
- Age (as of September 1, 2006)
- English as a second language (coded as 0 = English is one's second language, 1 = English is one's first language)
- Average parental education.
    - Computed as the average of the respondent's parents' education provided that the student provided a response for at least one parent. The item asked "What is the highest level of education each of your parents/guardians completed?" The response options are: 1 = Did not finish high school, 2 = High school graduate/GED, 3 = Attended college but no degree, 4 = Vocational/technical certificate or diploma, 5 = Associate or other 2-year degree, 6 = Bachelors or other 4-year degree, 7 = Masters, 8 = Law, 9 = Doctorate).
- Tested academic preparation. This was an imputed variable, created in several steps.
    - First, for those who have school file ACT data, give them a score.
    - For those who didn't have school file ACT data but who have school file SAT data give them their score for the SAT conversion to ACT.
    - For those who didn't have school file ACT data and who didn't have school file SAT data, give them their score for the COMPASS conversion to ACT.

  - For those who didn't have any school file data (i.e., no ACT, no SAT, and no COMPASS scores), give them their score for the self-reported ACT. Some student reported a score of '8' this is outside of the range and will be marked as 'missing'.
  - 18 students self-reported scores that don't match the SAT scoring matrix (i.e., do not end in zero). They are retained in the recodes because they're scores do not fall outside of the range of possible scores and thus, can be recoded using the ACT conversion scale.
  - 4 of the 18 students were removed because they're self-reported scores were outside of the conversion. For example, if a student self-reported a 1085, they would be dropped from the analysis.
  - For those who didn't have any school file data (i.e., no ACT, no SAT, and no COMPASS scores) and no self-reported ACT, give them their score for the self-reported SAT total, converted to ACT.
- Educational degree plans (1 = graduate degree or higher, 0 = less than a graduate degree)
- Institution attended was first choice (coded as 1 = first choice, 0 = not first choice)
- Academic motivation. An eight-item scale previously described as an outcome measure. We employed the precollege measure of academic motivation as a control in all analyses. The internal consistency reliability for the scale is .69. Constituent items included are:
  - I am willing to work hard in a course to learn the material even if it won’t lead to a higher grade.

- When I do well on a test, it is usually because I am well-prepared; not because the test is easy.
- In high school, I frequently did more reading in a class than was required simply because it interested me.
- In high school, I frequently talked to my teachers outside of class about ideas presented during class.
- Getting the best grades I can is very important to me.
- I enjoy the challenge of learning complicated new material.
- My academic experiences will be the most important part of college.
- My academic experiences will be the most enjoyable part of college.

For each item the response options are "strongly agree," "agree," "neutral," "disagree," or "strongly disagree." The responses we reverse coded and converted to means before scaling. 1 = low motivation for academic pursuits; 5 = high motivation for academic pursuits).

- Dependent children (coded as 1 = no dependents, 0 = has dependents)

<u>High school experiences:</u>

- High school racial composition (1 = All White or Mostly White, 0 = Other)
- High school involvement. A seven-item scale with an internal consistency reliability of .58 that measured involvement during high school. Constituent items included are:
  - During your last year in high school, how often did you study with a friend?
  - During your last year in high school, how often did you socialize with friends?
  - During your last year in high school, how often did you talk with teachers outside of class?

- During your last year in high school, how often did you participate in community service or volunteer activities?
- During you last year in high school, how often did you participate in extra-curricular activities?
- During your last year in high school, how often did you use the internet for homework or research?
- During your last year of high school, how often did you exercise?

For each item the response options are "very often," "often," "occasionally," "rarely," or "never."

Other college experiences:

- Live in dormitory or other campus housing (coded as 1 = yes, 0 = no)
- Full-time enrollment (coded as 1 = less that full-time, 2 = full-time)
- Number of hours per week during college one worked on campus (coded as 1 = 0 hours, 2 = 1-5 hours, 3 = 6-10 hours, 4 = 11-15 hours, 5 = 16-20 hours, 6 = 21-25 hours, 7 = 26-30 hours, 8 = More than 30 hours)
- Number of hours per week during college one worked off campus (coded as 1 = 0 hours, 2 = 1-5 hours, 3 = 6-10 hours, 4 = 11-15 hours, 5 = 16-20 hours, 6 = 21-25 hours, 7 = 26-30 hours, 8 = More than 30 hours)
- Was a member of a fraternity or sorority ("Are you a member of a social fraternity of sorority?" coded as 1 = no, 2 = yes).
- Participated in an intercollegiate sport ("Are you a student-athlete on a team sponsored by your institution's athletic department?" coded as 1 = no, 2 = yes).

- Number of courses during the current academic year taken in each of nine general areas: “Fine Arts, Humanities, and Languages” (e.g., art, music, philosophy, religion, history); “Mathematics/Statistics/Computer Science”; Natural Sciences” (e.g., chemistry, physics); “Social Science” (e.g., anthropology, economics, psychology, political science, sociology); “Allied Health” (e.g., nursing, physical therapy); “Business”; “Education”; “Engineering”; “Other” (e.g., architecture, agriculture, journalism). Response options were coded as 1 = 0 courses, 2 = 1 course, 3 = 2 courses, 4 = 3 courses, 5 = 4 courses, and 6 = 5 courses.

Figure Captions

Figure 1. Conceptual Model Guiding Estimation of the Total Effects of Attending a Liberal Arts College on Exposure to Good Practices/Liberal Arts Experiences.

Figure 2. Conceptual Model Guiding Estimation of the Direct Effects of Attending a Liberal Arts College on Exposure to Good Practices/Liberal Arts Experiences.

Figure 3. Conceptual Model Guiding Estimation of the Total Effects of Attending a Liberal Arts College on WNSLAE Outcome Measures.

Figure 4. Conceptual Model Guiding Estimation of the Direct Effects of Attending a Liberal Arts College and Exposure to Good Practices/Liberal Arts Experiences on WNSLAE Outcome Measures.

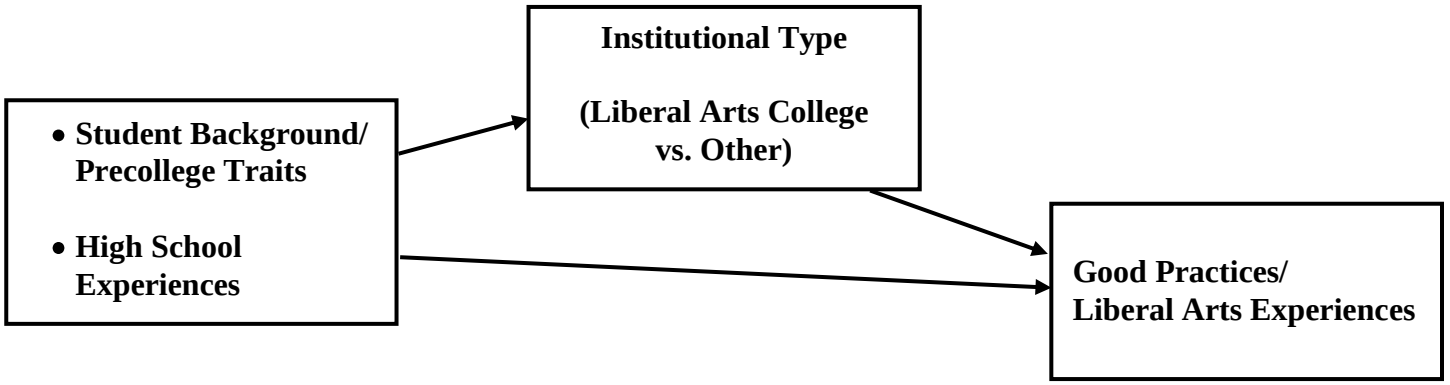
Institutional Type
(Liberal Arts College vs. Other)
• Student Background/ Precollege Traits
• High School Experiences
Good Practices/ Liberal Arts Experiences

Figure 1

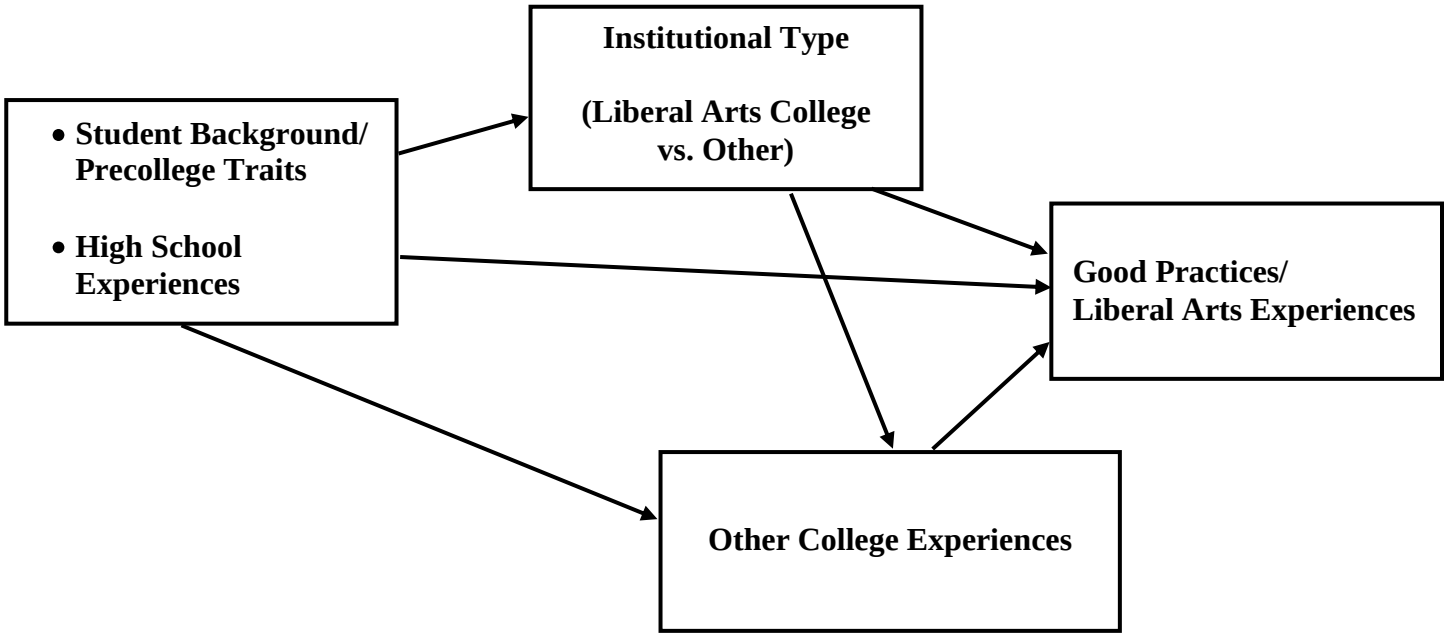
Institutional Type
(Liberal Arts College vs. Other)
• Student Background/ Precollege Traits
• High School Experiences
Good Practices/ Liberal Arts Experiences
Other College Experiences

Figure 2.

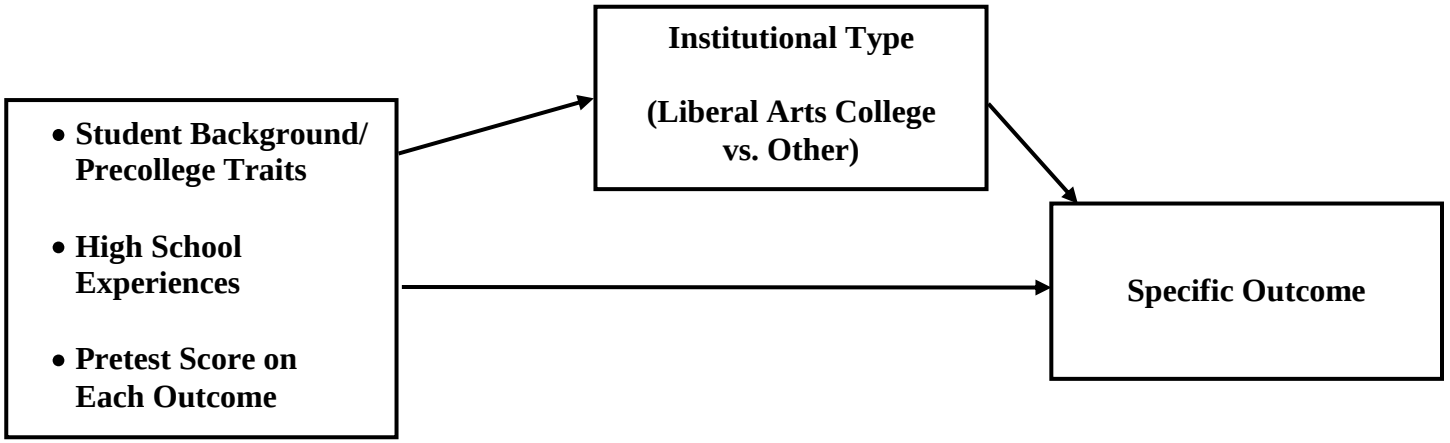
Institutional Type
(Liberal Arts College vs. Other)
• Student Background/ Precollege Traits
• High School Experiences
• Pretest Score on Each Outcome
Specific Outcome

Figure 3.

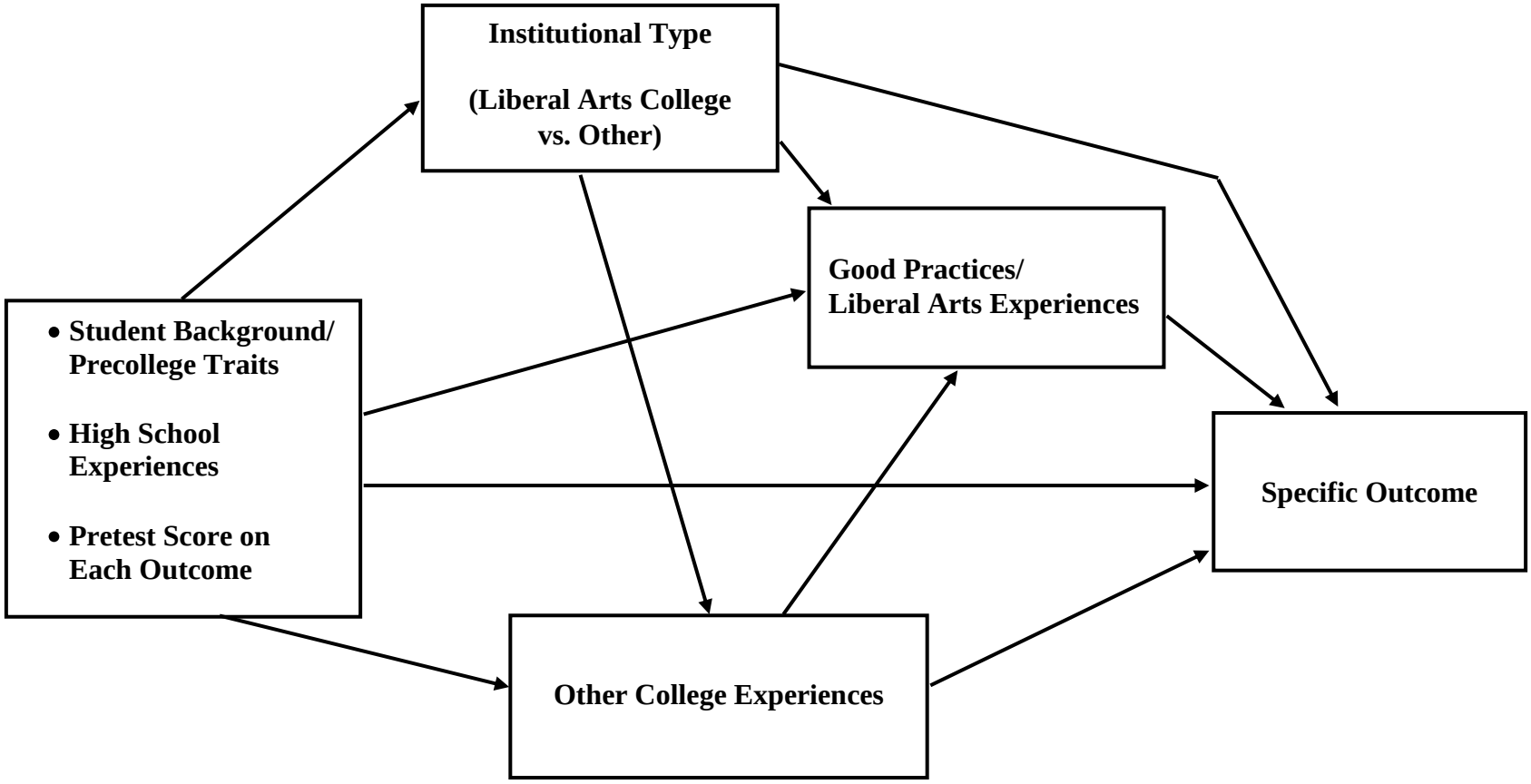
Institutional Type
(Liberal Arts College vs. Other)
Good Practices/ Liberal Arts Experiences
• Student Background/ Precollege Traits
• High School Experiences
• Pretest Score on Each Outcome
Specific Outcome
Other College Experiences

Figure 4.