

Running head: Selection Effects and First-Year Programs

Using Instrumental Variables to Account for Selection Effects in Research
on First-Year Programs

Gary R. Pike
Executive Director, Information Management & Institutional Research
Indiana University-Purdue University-Indianapolis
355 N. Lansing St., AO127
Indianapolis, IN 46202
pikeg@iupui.edu

Michele J. Hansen
Director of Assessment, University College
Indiana University-Purdue University-Indianapolis
355 N. Lansing St., AO127
Indianapolis, IN 46202
mjhansen@iupui.edu

Ching-Hui Lin
Graduate Student, Higher Education & Student Affairs
Indiana University Bloomington
2001 E. Lingelbach Ln, Apt. 331
Bloomington, IN 47408

Paper presented at the Annual Meeting of the Association for Institutional Research,
Chicago, IL, June 1, 2010

Using Instrumental Variables to Account for Selection Effects in Research on First-Year Programs

The widespread popularity of programs for first-year students is due, in large part, to studies showing that participation in first-year programs is significantly related to students' academic success. Because students choose to participate in first-year programs, self-selection effects prevent researchers from making causal claims about the outcomes of those programs. The present research examined the effects on first-semester grades of students participating in themed learning communities at a research university in the Midwest. Results indicated that membership in themed learning communities was positively associated with higher grade point averages, even after controlling for entering ability, motivation, gender, and first-generation/low-income status. However, when instrumental variables were introduced to account for self-selection, the effects of themed learning communities on grades were not statistically significant. The results have implications for campus leaders and assessment practitioners who are working to develop methods for understanding the effects of programs designed to enhance the undergraduate educational experiences on their campuses.

American colleges and universities have implemented a wide variety of programs for first-year students in an effort to improve the quality of undergraduate education and enhance student success. According to Upcraft, Gardner, and Barefoot (2005), nearly 75% of all colleges and universities offer programs specifically designed for first-year students. One reason for the popularity of these programs is that they appear to work. Studies show that participation in first-year programs is associated with a variety of positive educational outcomes, including a successful transition to college, higher grade point averages, and improved retention rates (Kuh, Kinzie, Schuh, Whitt, & Associates, 2005; Upcraft, Gardner, & Barefoot, 2005).

A common characteristic of most first-year programs is that students *choose* to participate in the programs, or are assigned to the programs based on major or an advisor's recommendation. The fact that students are not randomly assigned to first-year programs is not surprising given logistic, political, economic, and ethical concerns about the use of random assignment in higher education (Titus, 2006). Choosing to participate (or not participate) in first-year programs confounds research on program outcomes because the reasons students choose to participate in a first-year program are likely to be related to their subsequent success (DesJardins, McCall, Ahlburg, & Moye, 2002; Reynolds & DesJardins, 2009; Schneider et al., 2007). The present research examines the consequences of selection effects using data on themed learning community participation at a research university in the Midwest.

Background

Themed learning communities (TLCs) are an important element in efforts to enhance student success. At the institution of interest, a TLC consists of no more than 25 first-year students who are enrolled in three courses linked by a common theme. The theme may be major specific (e.g., biological sciences or business) or interdisciplinary (e.g., career exploration or the African American experience). In each TLC, one of the linked courses is a first-year seminar taught by an instructional team consisting of a faculty member, academic advisor, librarian, and student mentor. TLC participants also have a variety of opportunities for experiential learning through the co-curriculum and service learning (Williams, Chism, & Hansen, 2009). Evaluations of the program have shown that participating in a TLC is associated with significantly higher

grade point averages, particularly during the first semester, as well as higher persistence rates (Chism, Baker, Hansen, & Williams, 2008; Chism & Hansen, 2007; Hansen & Williams, 2005).

The positive outcomes associated with themed learning communities at the focus institution are not unique. Studies have consistently shown that learning community participation is positively related to a variety of beneficial educational experiences and outcomes. Past research has found that participating in a learning community facilitates the transition to college (Inkelas, Daver, Vogt, & Leonard, 2007; Inkelas & Weisman, 2003; Knight, 2003; Szelényi, Inkelas, Drechsler, & Kim, 2007) and is positively related to high levels of engagement during college (Inkelas et al., 2004; Inkelas & Weisman, 2003; Knight, 2003; Kuh, 2008; National Survey of Student Engagement, 2007; Pike, 1999, 2002; Stassen, 2003; Tinto & Goodsell, 1993; Zhao & Kuh, 2004). Membership in a learning community has also been linked to positive educational outcomes, including satisfaction with college (Baker & Pomerantz, 2000; Johnson & Romanoff, 1999; Zhao & Kuh, 2004), grades (Baker & Pomerantz, 2000; Knight, 2003; Pasque & Murphy, 2005; Pike, Schroder, & Berry, 1997; Purdie II & Rosser, 2007; Stassen, 2003), and persistence and graduation (Beckett & Rosser, 2007; Johnson & Romanoff, 1999; Knight, 2003; Pike, Schroder, & Berry, 1997; Purdie II & Rosser, 2007; Stassen, 2003).

The Self-Selection Problem

Rigorous experimental designs with random assignment of students to treatment conditions have become the “gold standard” for educational research and evaluation (Reynolds & DesJardins, 2009; Schneider et al., 2007). The value placed on the use of experimental/quasi-experimental methods and randomized field trials is a direct outgrowth of the Education Sciences Reform Act of 2002 which created the Institute of Education Sciences (IES) and charged it with providing “rigorous evidence on which to ground education practice and policy” (Institute of Education Sciences, no date, no page). The clear preference for randomized trials in program evaluations is evident in the U. S. Department of Education’s (2005) official notice in the *Federal Register*: “... the Secretary considers random assignment and quasi-experimental designs to be the most rigorous methods to address the question of project effectiveness” (p. 3586).

Despite the fact that random assignment and rigorous design are the preferred methods for evaluating education programs, they are the exception rather than the rule in research on programs for first-year students. In our review of the literature, we were not able to find any published research on learning communities that used random assignment to select participants. The Manpower Demonstration Research Corporation (2009) and the National Center for Postsecondary Research (2009) at Columbia University are currently conducting studies of learning communities in which students are randomly assigned to treatment conditions, but the results of this research are not currently available. The lack of random assignment to learning communities makes perfect sense. Most frequently, first-year students are allowed to choose whether they wish to join a learning community (and which learning community to join) because it is important that the theme of the learning community match students’ interests. In some cases, participation in a learning community is predicated on a student’s major or on an advisor’s recommendation. Here again, the assignment to a learning community is based on a choice and presumes that the students in the learning community are likely to benefit from participation in the program.

The absence of random assignment to learning communities creates serious problems for evaluating program outcomes. Absent random assignment, it may not be possible to make causal inferences about the effect of TLC participation. To understand why this is the case, it is helpful to review the logic underlying causal inference. For comprehensive reviews of this topic see Angrist and Pischke (2009), Cook and Campbell (1979), and Cook and Shadish (1994).

When evaluating the effect of an education program on an individual, the effect of the program is defined as the difference in the outcomes for that individual depending on whether he or she did (Y_{1i}) or did not (Y_{0i}) participate in the program ($Y_{1i} - Y_{0i}$). For a group of students, the difference in average outcomes (or expectations) provides the best representation of program effects:

$$\rho = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]. \quad (1)$$

In the equation above, ρ represents the effect of the program, $E[Y_i|D_i = 1]$ is the expected (i.e., average) outcome for a group of students participating in a program ($D_i = 1$). $E[Y_i|D_i = 0]$ represents the expected outcome if the *same group of students* did not participate in the program ($D_i = 0$). Unfortunately, it is not possible for the same group of students to participate and not participate in a program. As a consequence, researchers must observe the outcome of two groups of students—those who did and those who did not participate in the program. The causal-inference model for comparing the outcomes of two groups of students is:

$$\rho = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \quad (2)$$

In equation (2), the term $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$ or $E[Y_{1i} - Y_{0i}|D_i = 1]$ represents the average causal effect of a first-year program on the students who participated in that program. The second term, $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$, represents the selection effect (i.e., the difference in the expected outcome of program participants and non participants, if they did not participate in the program). It is important to understand that outcomes presuming nonparticipation (i.e., $E[Y_{0i}]$) are not the same as a pretest. A pretest-posttest design *does not* control for selection effects in the absence of random assignment. Pretest-posttest designs can be effective in ruling out history, maturation, testing, instrumentation, and mortality as threats to internal validity, but students still have the opportunity to voluntarily participate in the intervention if random assignment protocols are not employed (Cook and Shadish, 1994). Moreover, the pretest measure is not likely to represent the selection process by which students ended up in different participant groups.

Equation (2) is a useful definition of causal effects, but it cannot be used to calculate causal effects because $E[Y_{0i}|D_i = 1]$ (i.e., the outcome associated with not participating in a program [Y_{0i}] for those who participated in the program [$D_i = 1$]) cannot be observed. Angrist and Pischke (2009) note that random assignment overcomes this unobserved/omitted variable problem. When students are randomly assigned to a first-year program, participation or nonparticipation (D_i) is independent of potential outcomes (Y_i), and it is possible to swap $E[Y_{0i}|D_i = 0]$ for $E[Y_{0i}|D_i = 1]$. Thus,

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]. \quad (3)$$

When students are not randomly assigned to first-year programs, program effects are, by definition, confounded by selection effects.

Accounting for Self Selection

Researchers have used a variety of approaches to account for the effects of self selection. These approaches include regression-discontinuity designs (Hahn, Todd, & Van der Klaauw, 2001; Thistlewaite & Campbell, 1960), econometric (statistical) methods of adjustment (Heckman, 2008), matching (Reynolds & DesJardins, 2009; Titus, 2006), and instrumental variables (Angrist, Imbens, & Rubin, 1996; Angrist & Pischke, 2009). This study focuses on using instrumental variables (IV) to account for self-selection bias. IV approaches offer several advantages over other statistical methods. For example, IV analysis does not presume that treatment effects are constant for all participants, and it allows for a relatively straightforward assessment of whether instrumental variables are appropriate for a given study (Angrist, Imbens, & Rubin, 1996). In addition, instrumental variables utilize two stage least squares (TSLS) regression, which can be used to address a variety of statistical problems, including the estimation of simultaneous equations, solving problems of bias due to including variables measured with error in regression models, and addressing issues of omitted variable bias (Angrist & Pischke, 2009).

To better understand how instrumental variables can account for the confounding effects of self selection, it is helpful to examine the IV approach in the context of a hypothetical first-year program. A characteristic of this particular program is that student participation is voluntary. A typical regression analysis of program effects would utilize a model similar to the one represented by the equation

$$Y_i = \alpha + \beta_j X_{ij} + \rho D_i + \eta_i. \quad (4)$$

In equation (4), Y_i is the outcome for a given student (i), and α is the intercept (i.e., the value of Y_i when the values of all other variables in the model are zero). Also in the equation, X is a vector of j exogenous covariates that are significantly related to the outcome measure. The covariates (X_{ij}) may or may not be related to program participation (D_i). As Angrist and Pischke (2009) note, inclusion of covariates that are significantly related to the outcome measure can substantially improve efficiency of estimation, even when the covariates are not significantly related to program participation. The vector β_j represents the effects for the covariates. In the model, D_i represents whether a student participated ($D_i = 1$) or did not participate ($D_i = 0$) in the first-year program, and ρ is the *unbiased* effect of participating in the first-year program. When evaluating first-year program outcomes using equation (4), η_i is the structural disturbance. Because of selection bias, the structural disturbance is composed of both a self-selection effect (γS_i) and an error term or residual (v_i):

$$\eta_i = \gamma S_i + v_i \quad (5)$$

If the self-selection effect could be measured, equation (4) could be combined with equation (5) to represent program effects:

$$Y_i = \alpha + \beta_j X_{ij} + \rho D_i + \gamma S_i + v_i. \quad (6)$$

Unfortunately, self-selection effects cannot be directly measured, and γS_i represents an omitted variable that biases the evaluation of the first-year program's effects. Most important γS_i is related to program participation and causes the effects of program participation to be overstated.

In the instrumental variables context, calculating program effects (ρ) when S_i is unobserved requires that a researcher have access to one or more instruments (Z_i) that are

strongly related to first-year program participation (D_i), but are uncorrelated with any other determinants of the outcome measure (i.e., the covariates or the error term). Assuming both of these conditions are met, it follows that

$$\rho = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)} = \frac{\text{cov}(Y_i, Z_i)/\text{var}(Z_i)}{\text{cov}(D_i, Z_i)/\text{var}(Z_i)} = \frac{\Pi_{11}}{\Pi_{21}}. \quad (7)$$

The effect of first-year program participation (ρ) is the ratio of the regression of the outcome on the instrument ($\Pi_{11} = \text{cov}(Y_i, Z_i)/\text{var}(Z_i)$) to the regression of program participation on the instrument ($\Pi_{21} = \text{cov}(D_i, Z_i)/\text{var}(Z_i)$) (Angrist, Imbens, & Rubin, 1996). In the evaluation of a first-year program, the regression coefficients can be estimated from the following equations:

$$Y_i = \alpha_1 + \beta_{1j}X_{ij} + \Pi_{11}Z_i + \varepsilon_{1i} \quad (8)$$

$$D_i = \alpha_2 + \beta_{2j}X_{ij} + \Pi_{21}Z_i + \varepsilon_{2i} \quad (9)$$

Equation (8) represents the reduced-form model, and Π_{11} is the numerator in equation (7). Equation (9) is the first-stage model, and Π_{21} is the denominator in equation (7).

Evaluating the assumptions of IV methods provides a clear indication of the appropriateness of using instrumental variables to account for the confounding effects of self selection. The first assumption focuses on the effect of the instrument (Z_i) on the endogenous program-participation variable (D_i) and evaluates the strength of the instrument. Bound, Jaeger, and Baker (1995) demonstrated that when the relationship between the instrument and the endogenous variable is weak, IV estimates will be biased in the same direction as ordinary least squares (OLS) regression estimates. The effect of weak instruments can also be seen in the fact that the denominator in equation (7) will be small and is likely to produce a ratio (ρ) that overstates program effects, just like OLS regression. The second assumption, termed the “exclusion restriction,” presumes that the instrument (Z_i) is not related to the outcome measure (Y_i), except through the endogenous variable (D_i). If the instrument is related to the outcome measure above and beyond the indirect effect acting through the endogenous variable, the numerator in equation (7) will be too large, and the effect for the endogenous variable will again be overstated.

Self Selection, TLCs, and Instrumental Variables

Although research on themed learning communities indicates that participation in the program is positively related to students’ grade point averages, the fact that students choose to participate in TLCs confounds the evaluation of program effects. A student typically chooses to participate in TLCs in a meeting with an academic advisor who provides the student with detailed descriptions of the learning communities’ unique themes, courses offered in the “blocks,” and educational experiences offered (e.g., service learning, co curricular activities, campus events). Subsequently, the student enrolls in the TLC depending on his/her preferences, schedule, major, etc. The current study assessed the effects of TLCs using a three-phase evaluation. The first phase of the evaluation examined the zero-order relationship between participation in a TLC and fall semester grade point average. The second phase of the evaluation examined the effect of TLC participation, net the effects of selected student characteristics. The final phase of the evaluation examined the effects of TLC participation after accounting for self-selection effects using instrumental variables.

Three research questions, representing the three phases of the study, guided this research:

1. Is participation in a themed learning community significantly and positively related to higher Fall grade point averages?
2. Is participation in a TLC significantly and positively related to higher Fall grade point averages, net the effects of selected student characteristics?
3. Is participation in a TLC significantly and positively related to higher Fall grade point averages after accounting for the confounding effects of self selection?

Research Methods

Participants

The population for this study consisted of 2552 first-time, full-time, degree-seeking freshmen who enrolled at the university during the Fall 2008 semester. Complete data were available for 2193 students. Table 1 displays descriptive statistics for the students in this study, as well as for the Fall 2008 entering cohort. Females comprised 60.6% of the students in the study and 59.1% of the cohort. White students represented 80.6% of the study sample. Approximately 8.9% of the participants were African American, 3.4% were Hispanic/Latino, 4.3% were Asian or Pacific Islander, 0.2% were American Indian or Alaskan Natives, and 2.5% were included in other classifications (including international). In the beginning freshman cohort, a smaller percentage of students were White (77.0%), and a higher percentage of students were included in the “other” classification (5.8%). The relatively small percentage of students classified as “other” in the study was attributable to international students being excluded because they did not have SAT scores and/or high school grade point averages.

Insert Table 1 about here

At the university, relatively few students are admitted directly to an academic program. As a result, the preponderance of students in the study (63.7%) and the entering cohort (65.6%) were enrolled in a university college. Other schools with noteworthy enrollments were science (11.0% and 10.3%, respectively), engineering and technology (7.2% and 7.6%, respectively), and art and design (4.5% and 4.0%, respectively). Entering qualifications for the students in the study and in the population were very similar. The average combined math and verbal SAT score was slightly over 1015 for the sample and slightly below 1015 for the population. The mean high school grade point average for students in the study was 3.26, compared to a mean of 3.25 for the beginning cohort. Other student characteristics were also remarkably similar for study participants and the entering freshman cohort.

Measures

Eight variables were included in the present research. The outcome variable of interest was students’ Fall semester grade point averages (Fall GPA). Although some scholars have decried what they believe to be an over emphasis on grades, Pascarella and Terenzini (1991) observed that first-year grades are the most revealing indicator of a student’s successful transition to college. In addition, grades are significantly related to persistence and graduation,

admission to graduate or professional school, and entry into high-level occupations (Baird, 1985; Pascarella & Terenzini, 1991, 2005; Tinto, 1975). The Fall GPA variable used in this study ranged from 0.00 to 4.00; the mean Fall GPA was 2.73; and the standard deviation was 0.95. The treatment variable, participation in a themed learning community (TLC Participation) was dichotomously scored to indicate participation (1) and nonparticipation (0). TLC participation was considered to be an endogenous variable because it was assumed to influence Fall GPA, but was itself influenced by an omitted variable—self selection.

Four variables were included in the study as exogenous covariates. Several other variables (e.g., credit hour load and race/ethnicity) were significantly related to Fall GPA, but were not included in the research because they did not improve the explanatory power of the model. The first covariate was students' predicted grade point averages (Pike & Saupe, 2002). Predicted GPAs were originally developed to assist in admission decisions at the university. The prediction formula was obtained by regressing actual grade point averages from previous cohorts on those students' combined SAT scores and high school grade point averages. The resulting formula was then used to calculate predicted grade point averages for subsequently freshman cohorts. The result was a single measure representing the best linear combination of students' academic qualifications for predicting first-year grades. Predicted GPA has been used in previous research on learning communities to account for differences in entering ability (Pike, Schroeder, & Berry, 1997). The formula used to calculate predicted grade point average was

$$\text{Predicted GPA} = -1.244 + 0.001 \times \text{SAT} + 0.944 \times \text{High School GPA.} \quad [10]$$

The resulting mean and standard deviation for predicted GPA were 2.85 and 0.50, respectively.

The second covariate included in the study was considered to be a proxy for students' commitment and motivation. Previous research has found that students' noncognitive characteristics, including their motivation and commitment to succeed academically, are significantly related to their grades in college (Pascarella & Terenzini, 1991, 2005; Williford, 1996). However, the inclusion of noncognitive characteristics that can serve as proxies for commitment and motivation frequently necessitates the use of self-report instruments such as entering student surveys. The disadvantages of relying on survey instruments include the potential for nonresponse error and socially desirable responses. Students' application dates were used to create a proxy for motivation because there were no missing values (i.e., every student has an application date). The use of application date was based on the assumption that more motivated and committed students would complete their applications to college earlier than less motivated and committed students. In this study, application date was used to calculate the number of weeks prior to the start of classes in the Fall that a student completed his or her application. Because the data were positively skewed, students who applied more than 50 weeks prior to the beginning of classes were assigned a value of 50. Scores for the motivation proxy ranged from 4 to 50; the mean was 35.99; and the standard deviation was 9.82.

Gender, the third covariate in the study, was dichotomously scored to indicate whether a student was female (1) or male (0). Gender was included as a covariate because research has shown that women tend to have higher grade point averages than men (Malin, Bray, Dougherty, & Skinner, 2005; Matthews, 1991; Pascarella & Terenzini, 1991, 2005; Pike, Schroeder, & Berry, 1997). The mean for this variable was 0.61, and the standard deviation was 0.49.

The final covariate used in the study was an indicator of whether the student was generally first generation *and* low income. Previous research has clearly shown that both first-

generation status and low socioeconomic status are negatively related to academic success (DesJardins et al., 2002; Ishitani, 2003; Terenzini et al., 1994; Terenzini et al., 1996). Preliminary analyses indicated that both first-generation status and low-income status were negatively related to Fall grade point average at the university; however, the combination of first-generation and low-income status was most strongly related to Fall grades (Pike, 2009). In Indiana, the 21st Century Scholars program was created to increase access and affordability for prospective low income students (Indiana Commission for Higher Education, 2009). Eligibility for the 21st Century Scholars “Gear Up” scholarship was used as the fourth covariate in this research. Although some of the scholarship-eligible students were not the first in their families to attend college, most were first generation. Most important, both first-generation and low-income status were not significantly related to grades when 21st Century Scholarship eligibility was included in the analysis. The variable was dichotomously scored so that a value of “1” indicated that a student was eligible for the scholarship, whereas a score of “0” indicated as student was not eligible for the scholarship. The mean for the first-generation, low-income indicator was 0.12, and the standard deviation was 0.32.

Two variables were used as instruments in the IV analysis. The first variable was a measure of whether a student had participated in the university’s summer bridge (i.e., transition) program. Some students who participated in the bridge program were expected to participate in themed learning communities, whereas other bridge participants were strongly encouraged to participate in TLCs. Nevertheless, some bridge students did not participate in TLCs and some TLC participants did not participate in the bridge program. Of the 363 summer bridge participants, 199 (54.8%) also participated in a themed learning community. In addition, 359 (19.6%) of the students who did not participate in the summer bridge program did participate in a TLC. Examination of the correlations among summer-bridge participation, TLC participation, and Fall grades strongly suggested that the assumptions required of an IV analysis would be satisfied. Summer-bridge and TLC participation were significantly correlated (0.30), whereas summer-bridge participation and Fall GPA were weakly correlated (0.05). Correlations between summer bridge participation and the exogenous covariates were also low.

The second instrumental variable included in the study was whether a student had decided on a major. It was expected that students who had decided on a major would be more likely to join a themed learning community because most of the TLCs were discipline specific. The relationship between having decided on a major and TLC participation was relatively weak. Of the 2,017 students who had decided on a major, 524 (26.0%) participated in a TLC. At the same time, 34 (19.3%) of the 176 students who were still deciding on a major joined a themed learning community. The correlation between deciding on a major and TLC participation (0.06) was statistically significant, but weak, whereas the correlation between deciding on a major and fall grades was nonsignificant (-0.01). Again, having decided on a major was weakly correlated with the exogenous covariates. By itself, the measure of whether a student had decided on a major would be a weak instrument that could not adequately account for self-selection effects. However, whether a student had decided on a major was largely uncorrelated with participating in a summer bridge program (0.01). As a consequence, deciding on a major was included in the IV analysis, along with participation in the summer bridge program because it was thought it would increase the explanatory power of the instruments.

Data Analysis

The data analysis was carried out in three phases corresponding to the three research questions in the study. All of the analyses were conducted using the Stata 10 computer program (Stata Corp., 2007). For the first phase of the data analysis, students' Fall grade point averages were regressed on the TLC participation variable. Preliminary regression diagnostics revealed that errors were not distributed uniformly ($\chi^2 = 43.21$; $df = 1$; $p < 0.05$). That is, the assumption of homoscedasticity was not met (Breusch & Pagan, 1979; Cook & Weisberg, 1982). As a consequence, robust standard errors that were appropriate under conditions of heteroscedasticity were utilized (Davidson & MacKinnon, 1993).

During the second phase of the data analysis, predicted GPA, motivation, gender (being female), and first-generation/low-income status were included in the regression model as exogenous covariates. Variables were entered one at a time and estimates of explained variance were calculated at each step in order to identify the proportion of the variance in Fall grades that was uniquely attributable to each covariate. Robust standard errors were utilized and variance inflation factors (VIF) were calculated to determine if collinearity was a serious issue in the regression analysis. In addition, Ramsey's (1969) RESET test was performed to determine if the final model was misspecified and variables representing exponents of the fitted variables should be included in the model.

In the final phase of the data analysis, an instrumental variables regression was performed using two stage least squares. The model included the same variables as the final model from the second phase of the data analysis. In other words, Fall GPA served as the dependent variable, and TLC participation was the endogenous variable. Predicted GPA, motivation, gender, and first-generation/low-income status were the exogenous covariates. The first-stage model for the IV analysis included TLC participation as the dependent variable. Gender, predicted GPA, first-generation/low-income status, and student motivation were the exogenous covariates, and both summer-bridge participation and having decided on a major served as independent variables in the first-stage model. Robust standard errors appropriate for conditions of heteroscedasticity were utilized.

Several statistics were calculated to test the assumptions of IV regression and thereby evaluate the appropriateness of the instruments. A partial R^2 statistic representing the squared correlation between TLC participation, net the effects of the covariates, and the two instruments was calculated to assess the strength of the instruments. A F statistic representing the joint significance of the relationship between the two instruments and TLC participation was also calculated. Following the recommendation of Stock and Yogo (2005), a minimum F value of 10.00 was set as the standard for strong and reliable instruments.

A test of overidentification restrictions was also conducted to verify that the instruments were not correlated with the error term and that no other instruments needed to be included in the model (Davidson & MacKinnon, 1993). The test of overidentification restrictions was particularly relevant given the relatively weak correlation between TLC participation and whether the student had decided on a major. Wooldridge's (1995) score test of overidentification restrictions was used in this study because it is robust with respect to heteroscedasticity.

Results

The OLS regression analyses revealed that there was a statistically significant relationship between participating in a themed learning community and Fall semester grade point averages ($F = 61.06$; $df = 1, 2191$; $p < 0.05$). This relationship accounted for approximately 2% of the variance in students' Fall grade point averages. Table 2 displays the results for all of the OLS regression models. The columns included under the heading "Model 1" present the regression coefficients and robust standard errors for the initial model. An examination of the regression coefficients and standard errors for the first model revealed that both the constant (i.e., the intercept) and the effect for TLC participation were statistically significant. Because it was a dichotomous variable, the coefficient for TLC participation indicated that participation in a themed learning community (TLC = 1) was associated, on average, with a 0.32 higher grade point average than nonparticipation (TLC = 0).

Insert Table 2 about here

Including all four exogenous covariates in the analysis produced a statistically significant result ($F = 181.99$; $df = 5, 2187$; $p < 0.05$). Slightly less than 27% of the variance in students' Fall semester grade point averages was explained by the model. Results of the RESET test indicated that the model was not misspecified ($F = 0.60$; $df = 3, 2184$; $p > 0.05$). In addition, examination of the variance inflation factors (VIF) revealed that collinearity was not an issue. The mean VIF was 1.02 and the largest VIF coefficient was 1.04. An examination of the coefficients for the "Model 5" in Table 2 indicated that all four covariates were significantly related to Fall-term GPA, as was participation in a themed learning community. TLC participation was associated with a 0.28 increase in Fall GPA, and accounted for approximately 2% of the variance in Fall grades. Predicted GPA, the proxy for student motivation, and gender (being female) were all positively related to Fall GPA. The partial R^2 coefficients for the three variables were 0.22, 0.02, and 0.01, respectively. First-generation/low-income status was negatively related to Fall GPA and accounted for slightly less than 1% of the variance in the outcome measure.

The omnibus result for the two-stage least squares instrumental variables model was statistically significant (Wald $\chi^2 = 841.84$; $df = 5$; $p < 0.05$) and explained slightly more than 25% of the variance in students' Fall-semester grades. Regression coefficients and robust standard errors for the variables included in the analysis are presented in Table 3. The results in the top half of Table 3 tell a different story than the results for the OLS regression analyses. In line with the OLS results, predicted GPA (0.82), the student-motivation proxy (0.01), and gender (0.21) were positively related to Fall GPA. First-generation/low-income status was negatively related to Fall grades (-0.20). In contrast to the results of the OLS regression analyses, TLC participation *was not* significantly related to Fall GPA in the IV analysis.

Insert Table 3 about here

The bottom half of Table 3 displays the regression results with learning community participation as the dependent variable. An examination of the coefficients in the bottom half of the table reveal that both gender (being female) and first-generation/low-income status were significantly, but weakly related to learning community participation. The coefficient for gender was positive, indicating that TLC participants were more likely than nonparticipants to be female. The negative coefficient for first-generation/low-income status indicated that TLC participants were less likely than nonparticipants to be first-generation and low-income students. Neither predicted GPA nor student motivation were significantly related to participating in a themed learning community. Both the coefficient for participating in the summer bridge program (0.36) and having decided on a major (0.06) were significantly and positively related to TLC participation. These findings were consistent with the expectation that students who participated in the summer bridge program would be more likely to participate in a themed learning community and that students who had decided on a major would be more likely to join a TLC.

Tests of the strength of the relationship between the instruments and the endogenous variable supported the reliability of the instruments. The partial R^2 for the instruments in the first-stage model was 0.09, and the robust F statistic was 81.01 ($df = 2, 2186; p < 0.05$). This F value was well above the threshold of 10.00 recommended by Stock and Yogo (2005). The result of Wooldridge's (1995) score test of overidentifying restrictions was not statistically significant ($\chi^2 = 1.92; df = 1; p > 0.05$), indicating that the model was appropriately specified and the instruments were not correlated with the error term.

Discussion

Care should be taken not to overgeneralize the findings of this study. The results are limited to a single first-year program at a single institution, and they are not representative of possible results for all types of first-year programs or all types of colleges and universities. In addition, the participants in this research were drawn from a single freshman cohort. Although the students in the study were very similar to students in the cohort, and the students in the cohort were generally similar to students in previous freshman cohorts, it is possible that research using a sample from a different cohort would product different results. Moreover, the data analysis included a limited number of exogenous variables as statistical controls. Including a different array of variables in the research could have influenced the results in unknown ways. Finally, the results of this study are limited to a single outcome—Fall semester grade point average. As previously noted, grades are an important outcome of college (Baird, 1985; Pascarella & Terenzini, 1991, 2005). However, there are many other important outcomes of college, and themed learning communities may be positively related to many of these outcomes (e.g., critical thinking, integrative and interdisciplinary learning, civic engagement, written communication skills, persistence, degree completion, etc.), even after accounting for self selection

Despite these limitations, the results of the current research have important implications for theory, research, and practice. First and foremost, the findings of this study suggest that self-selection effects can confound research on first-year programs, such as themed learning communities. In this study, a direct assessment of the relationship between TLC participation and grades using OLS regression found that participating in a themed learning community was positively associated with students' Fall semester grade point averages. Including exogenous covariates in the analyses reduced the strength of the relationship between TLC participation and

grades only slightly. TLC participants still had significantly higher Fall grade point averages than nonparticipants. These findings are consistent with the results of previous research on learning communities at IUPUI and elsewhere (Baker & Pomerantz, 2000; Chism et al., 2008; Chism & Hansen, 2007; Hansen & Williams, 2005; Knight, 2003; Pasque & Murphy, 2005; Pike, Schroeder, & Berry, 1997; Purdie II & Rosser, 2007; Stassen, 2003).

When instrumental variables were used to account for omitted variable bias due to self selection, the effect of TLC participation on Fall semester grades was not statistically significant. In fact, the direct causal effect of TLC participation on Fall GPA was trivial, accounting of 0.01 on a four-point (i.e., 0.00 to 4.00) scale. This is not to say that TLC participants did not have higher grade point averages than nonparticipants. The observed difference between the Fall semester grade point averages of TLC participants and nonparticipants was nearly one-third of a letter grade (0.32). What the results of the IV analysis suggest is that the *causal effect* of TLC participation on grades was negligible. It appears that students who participated in themed learning communities would have had significantly higher grades than nonparticipants, irrespective of whether or not they participated in a themed learning community.

Given that participation in most first-year programs is voluntary, accounting for self selection is critically important in research on the freshman year experience. Since most studies in higher education use volunteer participants, the likelihood of selection effects may be substantial. These selection effects may result in the over-representation of participants who feel relatively comfortable in settings where interaction with other students and faculty members is paramount; they may not fear taking active steps to participate in co-curricular activities and experiential learning opportunities, and look forward to the self-reflection that is necessary for full participation in most first-year academic support programs. Additionally, factors influencing the students' selection in first-year programs may vary depending on the setting and educational context. Concerns about self-selection effects extend well beyond the first year of college. Opportunities for students to choose (e.g., institutions, majors, courses, and co-curricular activities) abound in higher education. The findings of this study suggest that researchers interested in the experiences and outcomes of college students should carefully consider how self selection may confound their findings and take steps to account for self-selection bias.

The confounding effects of self selection are critically important for program evaluation as well. External pressures to demonstrate the positive effects of education programs on students' academic achievement, levels of satisfaction, and persistence and graduation rates have been growing steadily (Reynolds & DesJardins, 2009; Starke, Harth, & Sirianni, 2001). If, as the results of this study suggest, traditional evaluation methods can overstate (either positively or negatively) the magnitude of program effects in the face of self selection, then evaluation research may be providing decision makers with inaccurate information. In addition to providing an incomplete accounting for external audiences, inaccurate information about program effectiveness can lead to the misallocation of scarce institutional resources.

The Education Sciences Reform Act of 2002 and the creation of the Institute of Education Sciences is a clear indication that the U. S. Department of Education prefers that educational researchers and evaluation professionals utilize randomized field trials to eliminate the confounding effects of self selection (Schneider et al., 2007). However, there are many instances where higher education research and program evaluation do not lend themselves to random assignment of participants and experimental or quasi-experimental designs. Furthermore, there may be excellent ethical, economic, and political reasons for not randomly assigning

students to education programs (Titus, 2006). In these instances, statistical control may be most appropriate. Approaches such as IV and the use of propensity scores offer a number of viable quasi-experimental research designs that provide reasonably credible and inexpensive alternatives for random assignment.

The present research demonstrates that instrumental variables (IV) analysis is one method that can be used to account for the confounding effects of self selection. However, IV methods are not without their limitations. As Cellini (2008) observed, finding instruments that satisfy IV assumptions can be extremely difficult, and debates about the validity of instruments are always possible. The present research is a case in point. The instruments used in this study clearly satisfy the statistical assumptions for instrumental variables analysis, but they are also the products of choices students made. For this reason, some researchers and policy makers may question the validity of these instruments. Depending on the situation, alternative approaches such as regression discontinuity or matching may be more appropriate than instrumental variables. What is clearly not appropriate is failing to account for selection effects of when attempting to make causal claims about education programs or college experiences.

The findings of this study also have implications for theory and research related to college grades and grade point averages. Specifically, the results of the present research indicate that students' entering characteristics are significantly related to their grades in college. For example, this study found that females had significantly higher first semester grade point averages than males, a result that has been reported in other studies of college students (Malin et al., 2005; Matthews, 1991; Pascarella & Terenzini, 1991, 2005; Pike, Schroeder, & Berry, 1997). Indicators of students' entering ability levels (e.g., SAT scores and high school GPA) were also significantly related to grades during the first semester of college. An important contribution of the present research is that these indicators of entering academic ability can be summarized in a predicted grade point average. This finding supports Pike and Saupe's (2002) recommendation that predicted grade point averages be used for admission decision and program evaluation.

Previous research has shown that students who are the first in their families to attend college and students from lower socioeconomic backgrounds are at a serious disadvantage in terms of academic performance (i.e., grades) (DesJardins et al., 2002; Ishitani, 2003; Terenzini et al., 1994; Terenzini et al., 1996). Although preliminary analyses found that both first-generation status and low-income status were negatively related to grades, the combination of first-generation and low-income status was most strongly related to lower grades (Pike, 2009). That finding was generally confirmed in the present research. In Indiana, the 21st Century Scholars program seeks to make college accessible and affordable to low-income students, many of whom are the first in their families to go to college. However, increasing accessibility and affordability does not necessarily translate into student success. The present research suggests that the combination of first-generation and low-income status negatively impacts academic achievement above and beyond any academic skills deficits that may be associated with being a first-generation, low-income student. Additional research is needed to understand how the combination of being the first in one's family to attend college and low socioeconomic status adversely affects academic achievement. Once these effects are understood, colleges and universities will need to design educational interventions to offset the negative effects of first-generation/low-income status.

A serendipitous outcome of the present research is the identification of a readily available measure that may serve as a proxy for student motivation. Scholars and practitioners have long

believed that noncognitive variables, including student motivation and commitment, can have a profound effect on student success (Sedlacek, 2004). The problem has been identifying appropriate and easy-to-use measures of motivation and commitment. Self-report items are easy to administer and score, but they may produce socially desirable responses that are not valid indicators of motivation. Standardized measures, such as the Non-Cognitive Questionnaire (NCQ), have been shown to tap a variety of domains that are related to motivation, and these measures can be used to accurately predict student success (Thomas, Kuncel, & Credé, 2007). Unfortunately, these measures may not be readily available, and they are not appropriate for *ex post facto* studies.

Previous research on the relationship between noncognitive variables and academic achievement indicated that statistically significant effects of motivation tend to disappear when cognitive predictors of academic success are included in explanatory models (Pascarella & Terenzini, 1991). That finding was not replicated in the current research. This study found that application date was significantly related to semester grades, even after controlling for differences in entering academic ability. To the extent that how early a student completes his or her application is an indicator of motivation and commitment to succeed in college, application date may serve as a valuable proxy for student motivation. In addition to serving as a statistical control in evaluations of education programs, application date may be a useful method of identifying at-risk students. More research is needed to understand how and why application date is related to achievement and how educational programs can offset the negative effect of an apparent lack of motivation and commitment to academic success.

The results of this study also support the argument made by Angrist and Pischke (2009) that covariates that are not related to program participation can be useful in evaluating program outcomes, as long as the covariates are related to the outcome measure. Angrist and Pischke's (2009) claim is based on their understanding that including covariates that are related to the outcome of interest increases the explained variance for the model and decreases the root mean square error for the model. Because the standard errors for the regression coefficients are a function of root mean square error, tests of program effects will be more powerful in models that explain more of the variance in the outcome measure. In other words, the model will be a more efficient estimator of program effects. This phenomenon is evident in the results of this study. The estimate of explained variance for the model that contained only the TLC participation variable (Model 1) was 0.02 and the root mean square error was 0.94. In contrast, the estimate of explained variance for the model that contained the TLC participation variable, predicted GPA, and the student motivation proxy (Model 3) was slightly more than 0.25 and the root mean square error for the model was approximately 0.83. The standard error for the effect of TLC participation was 0.040 in Model 1 and 0.036 in Model 3. What is significant about Model 3 is that predicted GPA and student motivation *were not* significantly related to TLC participation.

Although small, the 10% reduction in the magnitude of the standard error for TLC participation is noteworthy. As Pascarella and Terenzini (1991, 2005) observed, the effects of educational interventions are usually quite modest. Moreover, program evaluators often must use relatively small samples in their research. As a consequence, it is possible that the combination of small effects and small sample size will combine to make program effects appear to be nonsignificant. Including in a research or evaluation design theoretically and empirically justified covariates that are meaningfully related to outcomes can improve the power of statistical tests and increase the likelihood of identifying meaningful program effects.

This study offers numerous implications for the practice of conducting research and program evaluations in higher educational settings. It is important to note that research methods other than randomized or matched experiments can also be of great value. For example, correlational and descriptive research is essential in theory building and in the exploration of variables worthy of inclusion in program evaluation studies and research. Additionally, correlational and descriptive studies can be useful in exploring factors that are outside the scope of studies of educational interventions. In many educational settings and policy contexts, controlled experiments are not feasible and well-designed correlational or descriptive studies may be adequate to meet stakeholders' information needs and requests. (Slavin, 2002). Higher education policy makers, administrators, program evaluators, and assessment practitioners should carefully consider the type of information required to make critical policy decisions when selecting research designs and approaches. The controlled experiment may be the most appropriate design for studies that seek to make causal conclusions based on costly educational interventions and where settings allow for the implementation of carefully controlled experiments. In other cases, statistical control may be the most viable approach. In such cases, it is vital to take into account (1) the unique local threats to internal validity, (2) the possible selection processes by which students end up in different treatment groups, (3) the causes of the intended outcomes that are related to participation and not related to participation, and (4) the program theory or how and why the program intends to affect key outcomes.

Conclusion

As Upcraft, Gardner, and Barefoot (2005) noted, first-year programs that are designed to bolster student success have become a prominent feature of the higher education landscape. In large part, the popularity of these first-year programs rests on numerous studies showing the positive educational outcomes associated with program participation. At many institutions, learning communities have become the poster child for successful first-year programs because of seemingly unequivocal evidence that participating in a learning community is associated with a variety of positive educational outcomes. However, evidence of a significant positive relationship between learning community participation and educational outcomes is not equivalent to showing that learning community participation is responsible for those positive outcomes, as the current research amply demonstrates. In this study, participating in a themed learning community was associated with significantly higher Fall semester grades, but when statistical controls for self selection were introduced the effects of themed learning communities were not significant. At least at the focus institution, the grade-related benefits of themed learning communities appear to be a result of the kinds of students who are attracted to themed learning communities, rather than the educational experiences they offer. Research on college students and evaluations of education programs must be mindful of the effects of self selection and take appropriate steps to ensure that self-selection bias does not confound research results.

References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444-455.
- Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Baird, L. L. (1985). Do grades and tests predict adult accomplishment? *Research in Higher Education*, *21*, 261-278.
- Baker, S., & Pomerantz, N. (2000). Impact of learning communities on retention at a metropolitan university. *Journal of College Student Retention*, *2*, 115-126.
- Beckett, A. K., & Rosser, V. J. (2007, November). *Relationship between learning community participation and degree attainment*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Louisville, KY.
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*, 443-450.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, *47*, 1287-1294.
- Cellini, S. R. (2008). Causal inference and omitted variable bias in financial aid research: Assessing solutions. *Review of Higher Education*, *31*, 329-354.
- Chism, L. P., Baker, S. S., Hansen, M. J., & Williams, G. (2008). Implementation of first-year seminars, the summer academy bridge program, and themed learning communities. *Metropolitan Universities*, *19*(2), 8-17.
- Chism, L.P. & Hansen, M. J. (July, 2007). *Proving and improving: Purposeful themed learning communities assessment*. Paper presented at the International Conference on First-Year Year Experience, Honolulu, HI.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroskedasticity in regression. *Biometrika*, *70*, 1-10.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Cook, T.D., & Shadish, W.R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology*, *45*, 545-580.
- Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.
- DesJardins, S. L., McCall, B. P., Ahlburg, D. A., & Moye, M. J. (2002). Adding a timing light to the "Tool Box." *Research in Higher Education*, *43*, 83-114.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*, 201-209.

- Hansen, M. J., & Williams, G. A. (2005). Assessing learning communities at Indiana University-Purdue University in Indianapolis: Comprehensive approaches, leveraging results, lessons learned, and remaining challenges. *Metropolitan Universities Journal*, 16, 69-90.
- Heckman, J. J. (2008). *Econometric Causality* (Working Paper No. 13934). Cambridge, MA: National Bureau of Economic Research. Accessed July 3, 2009 from <http://www.nber.org/papers/w13934>.
- Indiana Commission for Higher Education (2009). *21st Century Scholars*. Accessed July 29, 2009 from <http://www.learnmoreindiana.org/college/paying/TwentyFirstCenturyScholars>.
- Inkelas, K. K., Brower, A. M., Crawford, S., Hummel, M., Pope, D., & Zeller, W. J. (2004, November). *National Study of Living-Learning Programs: 2004 report of findings*. College Park, MD: University of Maryland & Association of College and University Housing Officers International. Retrieved August 15, 2008 from <http://www.livelearnstudy.net/additionalresources/reports.html>.
- Inkelas, K. K., Daver, Z. E., Vogt, K. E., & Leonard, J. B. (2007). Living-learning programs and first-generation college students' academic and social transition to college. *Research in Higher Education*, 48, 403-434.
- Inkelas, K. K., & Weisman, J. L. (2003). Different by design: An examination of student outcomes among participants in three types of living-learning programs. *Journal of College Student Development*, 44, 335-368.
- Institute of Education Sciences (no date). *About IES*. Accessed July 7, 2009 from <http://www.ies.ed.gov/director>.
- Ishitani, T. T. (2003). A longitudinal approach to assessing attrition behavior among first-generation students: Time-varying effects of pre-college characteristics. *Research in Higher Education*, 44, 433-449.
- Johnson, J. L., & Romanoff, S. J. (1999). Higher education residential learning communities: What are the implications for student success? *College Student Journal*, 33, 385-399.
- Knight, W. E. (2003). Learning communities and first-year programs: Lessons for planners. *Planning for Higher Education*, 31(4), 5-12.
- Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities.
- Kuh, G.D., Kinzie, J., Schuh, J.H., Whitt, E.J., & Associates (2005). *Student success in college: Creating conditions that matter*. San Francisco: Jossey-Bass.
- Malin, J. T., Bray, J. H., Dougherty, T. W., & Skinner, W. K. (1980). Factors affecting the performance and satisfaction of adult men and women attending college. *Research in Higher Education*, 13, 115-130.
- Manpower Development Research Corporation (2009). *Project page: Learning communities demonstration*. New York: Author. Accessed August 28, 2009 from http://www.mdrc.org/project_31_76.html.

- Matthews, D. B. (1991). The effects of learning style on grades of first-year college students. *Research in Higher Education*, 32, 253-268.
- National Center for Postsecondary Research (2009). *Learning communities*. New York: NCPR, Teachers College, Columbia University. Accessed September 3, 2009 from <http://www.postsecondaryresearch.org/index.html?Id=Research&Info=Learning+Communities>.
- National Survey of Student Engagement (2007). *Experiences that matter: Enhancing student learning and success*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- 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: A third decade of research* (Volume 2). San Francisco: Jossey-Bass.
- Pasque, P. A., & Murphy, R. (2005). The intersections of living-learning programs and social identity as factors of academic achievement and intellectual engagement. *Journal of College Student Development*, 46, 429-441.
- Pike, G. R. (1999). The effects of residential learning communities and traditional residential living arrangements on educational gains during the first year of college. *Journal of College Student Development*, 40, 269-284.
- Pike, G. R. (2002). The differential effects of on- and off-campus living arrangements on students' openness to diversity. *NASPA Journal*, 39, 283-299.
- Pike, G. R. (2009, March). *An Analysis of First-Time Freshmen Dismissed from University College after Fall 2008*. Unpublished research report, Office of Information Management and Institutional Research, Indiana University-Purdue University-Indianapolis, Indianapolis, IN.
- Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education*, 43, 187-207.
- Pike, G. R., Schroeder, C. C., & Berry, T. R. (1997). Enhancing the educational impact of residence halls: The relationship between residential learning communities and first-year college experiences and persistence. *Journal of College Student Development*, 38, 609-621.
- Purdie, J. R. II, & Rosser, V. J. (2007, November). *Examining the academic performance and retention of four-year students in academic themed floors, freshman interest groups and first-year experience courses*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Louisville, KY.
- Ramsey, J. B. (1969). Tests of specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B (Methodology)*, 31, 350-371.
- Reynolds, C. L., & DesJardins, S. L. (2009). The use of matching methods in higher education research: Answering whether attendance at a 2-year institution results in differences in educational attainment. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Volume XXIV, pp. 47-104). Dordrecht, The Netherlands: Springer.

- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, D. C.: American Educational Research Association.
- Sedlacek, W. E. (2004). *Beyond the bug test: Noncognitive assessment in higher education*. San Francisco: Jossey-Bass.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21
- Stassen, M. L. A. (2003). Student outcomes: The impact of varying living-learning community models. *Research in Higher Education*, 44, 581-613.
- Starke, M. C., Harth, M., & Sirianni, F. (2001). Retention, bonding, and academic achievement Success of a first-year seminar. *Journal of the First-Year Experience & Students in Transition*, 13, 7-35.
- StataCorp LP. (2007). *Stata users guide: Release 10*. College Station, TX: Author.
- Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. Cambridge, UK: Cambridge University Press.
- Szelényi, K., Inkelas, K. K., Drechsler, M. J., & Kim, Y. C. (2007, November). *Exploring social capital in the transition to college of students in living-learning programs from differing socioeconomic backgrounds*. Paper presented at the annual meeting of the Association for the Study of Higher Education, Louisville, KY.
- Terenzini, P. T., Rendon, L. I., Upcraft, M. L., Millar, S. B., Allison, K. W., Gregg, P. L., Jalomo, R. (1994). The transition to college: Diverse students, diverse stories. *Research in Higher Education*, 35, 57-73.
- Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education*, 37, 1-22.
- Thistlewaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post factor experiment. *Journal of Educational Psychology*, 51, 309-317.
- Thomas, L. L., Kuncel, N. R., & Credé, M. (2007). Noncognitive variables in college admissions: The case of the Non-Cognitive Questionnaire. *Educational and Psychological Measurement*, 67, 635-657.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.
- Tinto, V. (2000). What have we learned about the impact of learning communities on students? *Assessment Update: Progress, Trends, and practices in Higher Education*, 12(2), 1-2, 12.
- Titus, M. A. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: An application to the private returns to a Master's degree. *Research in Higher Education*, 48, 487-521.

- United States Department of Education (2005, January 25). Scientifically based evaluation methods. *Federal Register*, 70(15), 3586-3589.
- Upcraft, M. L., Gardner, J. N., & Barefoot, B. O. (2004). *Challenging and supporting the first-year student: A handbook for improving the first year of college*. San Francisco: Jossey-Bass.
- Williams, G., Chism, L. P., & Hansen, M. J. (2009, May). *Linking a summer bridge program to themed learning communities*. Paper presented at the annual meeting of the Educational Policy Institute Retention Conference, New Orleans, LA.
- Williford, L. E. (1996, October). The freshman year: How do personal factors influence academic success and persistence? *1996 SAIR annual report: Charting the course in a changing environment*. Mobile, AL: Southern Association for Institutional Research.
- Wooldridge, J. M. (1995). Score diagnostics for linear models estimated by two stage least squares. In G. S. Maddala, P. C. B. Phillips, & T. N. Srinivasan (Eds.), *Advances in econometrics and quantitative economics: Essays in honor of Professor C. R. Rao*. Cambridge, MA: Blackwell.
- Zhao, C., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, 45, 115-138.

Table 1

Descriptive Statistics for the Study Participants and the Fall 2008 Freshman Cohort

Variable		Participants (N=2,193)	2008 Cohort (N=2,552)
Gender:	Female	60.6%	59.1%
	Male	39.4%	40.9%
Race/Ethnicity:	Black/African American	8.9%	9.3%
	Hispanic/Latino	3.4%	3.3%
	Asian/Pacific Islander	4.3%	4.3%
	American Indian/Alaska Native	0.2%	0.3%
	White	80.6%	77.0%
	Other	2.5%	5.8%
	Academic Unit:	Art	4.5%
	Business	1.8%	1.6%
	Dentistry	0.1%	0.1%
	Education	3.4%	3.0%
	Engineering & Technology	7.2%	7.6%
	Informatics	1.0%	0.9%
	Journalism	1.1%	1.0%
	Liberal Arts	3.1%	3.0%
	Medicine	0.1%	0.1%
	Physical Education & Tourism	2.5%	2.2%
	Public & Environmental Affairs	0.6%	0.6%
	Science	11.0%	10.3%
	University College	63.7%	65.6%

Table 1 Continued

Variable	Participants (N=2,193)	2008 Cohort (N=2,552)
Fall Semester GPA	2.73 <i>0.95</i>	2.69 <i>1.02</i>
SAT (Math & Verbal Combined)	1015.40 <i>146.42</i>	1014.99 <i>147.44</i>
High School GPA	3.26 <i>0.44</i>	3.25 <i>0.45</i>
Predicted GPA	2.85 <i>0.50</i>	2.84 <i>0.50</i>
Fall Credit Hours Attempted	13.75 <i>1.35</i>	13.71 <i>1.35</i>
First-Generation/Low-Income Status (21 st Century Scholar)	0.12 <i>0.32</i>	0.12 <i>0.32</i>
Motivation (Application Week)	35.99 <i>9.82</i>	34.68 <i>10.74</i>
Participated in Themed Learning Community	0.25 <i>0.44</i>	0.24 <i>0.43</i>
Participated in Summer Bridge Program	0.17 <i>0.37</i>	0.16 <i>0.37</i>
Decided on a Major	0.92 <i>0.27</i>	0.92 <i>0.27</i>

Standard deviations are presented in italics.

Table 2

Results of the OLS Regression Analyses

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Constant	2.65*	0.023	0.14	0.101	-0.21	0.112	-0.22	0.111	-0.21	0.111
TLC Participation	0.32*	0.040	0.31*	0.036	0.29*	0.036	0.28*	0.036	0.28*	0.036
Predicted GPA			0.88*	0.033	0.84*	0.033	0.83*	0.033	0.82*	0.033
Student Motivation					0.01*	0.002	0.01*	0.002	0.01*	0.002
Gender							0.19*	0.037	0.19*	0.037
First-Gen./Low-Inc.									-0.21*	0.061
R ²	0.021		0.236		0.253		0.262		0.267	
Root MSE	0.944		0.834		0.825		0.825		0.818	

* $p < 0.05$

Table 3

Two Stage Least Squares Results for the Instrumental Variables Model

Dependent Variable: Fall GPA		
	Coefficient	Robust Std. Err.
Constant	-0.18	0.113
Participated in Themed Learning Community	0.01	0.132
Predicted GPA	0.82*	0.033
Student Motivation	0.01*	0.002
Gender	0.21*	0.038
First-Generation/Low-Income	-0.20*	0.061
Dependent Variable: TLC Participation		
	Coefficient	Robust Std. Err.
Constant	0.13*	0.063
Predicted GPA	-0.02	0.017
Student Motivation	0.00	0.001
Gender	0.04*	0.018
First-Generation/Low-Income	-0.06*	0.029
Participated in Summer Bridge Program	0.36*	0.028
Decided on a Major	0.06*	0.032

* $p < 0.05$