

University of Pennsylvania
Teachers College, Columbia University
Harvard University
Stanford University
University of Michigan
University of Wisconsin-Madison
Northwestern University

Apples and Oranges: Comparing the Backgrounds and Academic Trajectories of International Baccalaureate (IB) Students to a Matched Comparison Group



Henry May
Awilda Rodriguez
Philip M. Sirinides
Laura W. Perna
April Yee
Tafaya Ransom

A COLLABORATIVE PUBLICATION BETWEEN

University of Pennsylvania
Teachers College, Columbia University
Harvard University
Stanford University
University of Michigan
University of Wisconsin-Madison
Northwestern University

Apples and Oranges: Comparing the Backgrounds and Academic Trajectories of International Baccalaureate (IB) Students to a Matched Comparison Group



Henry May
Awilda Rodriguez
Philip M. Sirinides
Laura W. Perna
April Yee
Tafaya Ransom

A COLLABORATIVE PUBLICATION BETWEEN

About the Consortium for Policy Research in Education (CPRE)

Since 1985, the Consortium for Policy Research in Education (CPRE) has brought together renowned experts from major research universities to improve elementary and secondary education by bridging the gap between educational policy and student learning. CPRE researchers employ a range of rigorous and innovative research methods to investigate pressing problems in education today.

Having earned an international reputation for quality research and evaluation, policy design and technical assistance, and dissemination and training, CPRE is a premier source of advice for education policymakers and practitioners. CPRE is known for its work in developing theory and evidence through studies of standards-based reform, education finance and resource allocation, educational leadership, assessment and data use, and instructional improvement initiatives. CPRE researchers have extensive experience conducting experimental studies, large-scale quasi-experimental research, qualitative studies, and multi-state policy surveys.

CPRE's member institutions are the University of Pennsylvania; Teachers College, Columbia University; Harvard University; Stanford University; University of Michigan; University of Wisconsin-Madison; and Northwestern University.

About the Center for Research in Education and Social Policy (CRESP)

The Center for Research in Education and Social Policy (CRESP) within the College of Education and Human Development at the University of Delaware conducts rigorous research to help policymakers and practitioners in education, health care, and human services determine which policies and programs are most promising for improving outcomes in children, youth, adults and families.

Although research in prevention sciences and health care have long used rigorous designs to assess the effectiveness of programs, it was not until the Education Sciences Reform Act of 2002 that we witnessed a dramatic increase in the quantity and quality of research to evaluate the effects of education programs and policies. The education community began to focus on research that could measure the impact of these programs through randomized experiments and other research designs that support causal conclusions and can determine whether, how well, for whom, and why new programs and interventions work.

CRESP specializes in experimental and quasi-experimental research that uses quantitative and mixed methods to evaluate how and how well programs and interventions work to improve educational, family, and health outcomes in schools and communities.

TABLE OF CONTENTS

5	Executive Summary
7	Introduction
9	The International Baccalaureate Diploma Programme
11	Literature Review and Conceptual Framework
11	Schools Choose to Offer IB, AP, and Dual Enrollment Programs
13	Schools Constrain and Enable Student Participation in Available Programs
14	Students Choose to Participate in Available Programs
16	Predictors of Student Participation, Recognizing Selection
18	A Conceptual Model of IB Participation
20	Research Methods
20	Population, Sample, and Data
21	Data Analysis
26	Results
26	Predictors of IB Participation
34	Multivariate Prediction of IB Participation
38	Comparison of Propensity Scores for IB and Non-IB Students
42	Reducing Selection Bias Through Propensity Score Stratification and Full Matching
50	Reducing (or Exacerbating) Selection Bias Through Propensity Pair Matching
57	Conclusions and Implications
63	References

The research reported here was supported [in whole or in part] by the Institute of Education Sciences, U.S. Department of Education, through Grant # R305E090049 and Grant #R305B090015 to the University of Pennsylvania and is based on data provided by the Florida K-20 Education Data Warehouse (EDW). The opinions expressed are those of the authors and do not represent the views of the Institute, the U.S. Department of Education, or the Florida EDW. This report has been internally and externally reviewed to meet CPRE and CRESPP quality assurance standards. All data presented, statements made, and views expressed are the responsibility of the authors and do not necessarily reflect the position of the Consortium for Policy Research in Education or the Center for Research in Education and Social Policy, or their institutional partners.

EXECUTIVE SUMMARY



This report presents findings from a retrospective study of the academic histories of International Baccalaureate (IB) students and other students in the state of Florida. The IB Diploma Program is an internationally recognized college-preparatory curriculum designed to provide students with a rigorous and comprehensive academic experience. IB has grown dramatically in recent years and is thought by many to be among the best college-preparatory programs in existence. As such, there is tremendous interest in the potential impacts of IB, but any attempts to examine those impacts must deal with selection bias that results from the voluntary participation of schools and students. Failure to do so makes it impossible to determine whether the performance of participating students was actually influenced by IB, or whether the outcomes for these students would have been just as good without IB.

As a critical step in understanding the impacts of IB, the analyses presented in this report examined the selection mechanisms behind IB participation across Florida, the state with the second highest representation of IB programs in the nation. We use longitudinal student and school-level data from 1995 through 2009 from the Florida K-20 Education Data Warehouse (EDW) to characterize individual students' educational histories from elementary school through high school and into college. To address issues of selection bias, we use propensity score methods (Rosenbaum & Rubin, 1983) to adjust for preexisting differences between IB and non-IB students. These analyses are designed to address the following research questions:

1. What are the student- and school-level predictors of participating in the IB Diploma Programme in Florida?
2. To what degree does propensity score stratification or matching reduce selection bias associated with key student and school-level factors?
3. What are the estimated differences in key postsecondary access indicators (i.e., SAT and ACT scores) and enrollment statistics (e.g., college selectivity) with and without different types of propensity score adjustments?

Results revealed that, when looking at the statewide population in Florida, the selection bias associated with voluntary participation in IB is very large, and that mechanisms for dealing with selection bias using propensity scores may not be sufficient. In other words, comparing IB and non-IB students in this statewide context is like comparing apples and oranges, and using propensity score methods to adjust for these differences require strong assumptions and extrapolation into regions with very thin data.

Key findings from our results are as follows:

- IB students in Florida are very, very different from non-IB students, and while school and student demographics are related to IB participation, the best predictors are indicators of prior academic performance.
 - IB students were more likely to be female, Asian or White, and identified as gifted/talented, while they were less likely to be English language learners, have a disability, or be eligible for free/reduced lunch.
 - Prior test scores, GPA, and course-taking indicators were by far the strongest predictors of IB participation, with 8th grade Algebra and advanced courses in 9th and 10th grade being the best predictors of IB participation overall.
 - A number of school-level variables (i.e., high average test scores, magnet status, racial composition) were predictive of IB participation, but these relationships were generally much weaker than student-level factors.
- There is very little overlap in the propensity scores for IB and non-IB students suggesting that decent causal inference is simply not possible for this statewide population of IB students.
- Any study using propensity score methods should include a comprehensive logic model of the selection mechanism in order to identify the degree to which the propensity score model does or does not include key elements influencing the selection of program participants.
- Future research on the impacts of IB should focus on contexts in which decent causal inference can be made. The most promising opportunities for this approach are situations where IB programs are over-enrolled and students must apply for admission through a lottery.

INTRODUCTION



With the goal of increasing students' academic readiness for college, high schools in the United States are increasingly offering "credit-based transition programs," including International Baccalaureate (IB), Advanced Placement (AP), and dual enrollment. In 2003, most public high schools in the nation offered at least one credit-based transition program, with 2% of high schools offering IB, 67% offering AP, and 71% offering dual enrollment (Waits, Stezer, & Lewis, 2005). Although not as prevalent as AP or dual enrollment, the IB Diploma Program may be the most rigorous credit-based transition program of the three. The IB Diploma Program is an internationally recognized college-preparatory curriculum designed to provide students with a rigorous and comprehensive academic experience. IB students are required to take advanced courses in all subjects, while students participating in AP, honors, or dual enrollment are typically permitted to choose which subjects they study at an advanced level, while selecting other courses from the standard high school curriculum. At the end of the 12th grade, IB students take an internationally standardized comprehensive examination that includes both oral and written components. Students who pass these assessments are granted an IB diploma.

Although some research points to the promise of IB, AP, and other credit-based transition programs for improving students' academic readiness for college (e.g., Duevel, 1999; Foust et al., 2009; Poelzer & Feldhusen, 1996; Moydell et al., 1991; Roderick, Nagoaka, Coca, & Moeller, 2009; Saavedra, 2011), conclusions about program effects are often limited by potential issues of selection bias. More specifically, most research is limited by the reality that (a) schools choose to offer these programs (either as whole school programs or as programs within schools), (b) schools enable and/or restrict access to these programs based on locally determined admissions processes, and (c) eligible students (and their families) choose to participate in available programs. Despite strong statistical controls and assumptions to address selection, such research may not be able to determine whether differences in outcomes are caused by program participation or are simply an artifact of the unmeasured characteristics of schools, students, and families that correlate with the decision to participate in these optional programs.

As the prevalence of IB and other credit-based transition programs continues to grow, and policies are implemented to increase students' access to these programs, it is important that research investigating program impacts carefully consider the selection of schools and students into the programs. For example, descriptive data collected from a survey of coordinators of Florida IB Diploma Programmes reveal that most programs require students to have a minimum grade point average, and about half also require minimum standardized test scores (Perna et al., 2013). Although these requirements are often minimal (e.g., a "B" average, or a passing score on the state test), and most coordinators admitted that these and other admissions requirements are not strictly enforced, coordinators also asserted the prestigious and academically-elite nature of their program and reported that only the most highly-motivated students volunteer to participate (Perna et al., 2013). Given the perceived and expected academic rigor of these programs, it is likely that a majority of the "best and brightest" students attending a high school will volunteer for IB. If so, then issues of selection threaten to dramatically bias results of any study comparing the outcomes of IB and non-IB students.

The most effective approach for eliminating selection bias is to randomly assign schools or students to program participation, thus creating probabilistically equivalent treatment and control groups. Yet efforts to randomly assign schools or students to credit-based transition programs are limited by many forces, including the current widespread availability of credit-based transition programs and the political issues involved with granting some schools and students access, while denying others. In situations where real-world challenges limit the random assignment of students into treatment and control groups, researchers have used quasi-experimental and statistical techniques that attempt to adjust for pre-existing differences between IB students and a comparison group of non-IB students. Unfortunately, most research to date on the impacts of IB has been limited by the scope of the sample studied (e.g., focusing on only one school or district) and the availability of relevant selection predictors. The variables used to adjust for selection are often selected simply because they are available, despite a lack of grounding in a comprehensive theory of how schools, students, and parents influence the selection of students into these programs.

To address this knowledge gap and inform future studies of the impacts of credit-based transition programs, this research report makes three contributions. First, a review of existing literature is used to produce an empirically-based conceptual model of selection into IB. Second, the conceptual model is used to identify the characteristics of students and schools that participate in the International Baccalaureate Diploma Programme using data from the National Center for Education Statistics and the Florida Education Data Warehouse. The conceptual model also allows us to identify key predictors for which there are no data available. Third, we test the ability of the available data to adjust for observed selection bias using propensity score methods (Rosenbaum & Rubin, 1983), with the degree of bias reduction reported for each predictor. If substantial selection bias persists after the adjustments, or if the adjustments impose dramatic extrapolations of the data (i.e., comparing apples and oranges), then we must question the utility and validity of propensity score analyses intended to estimate the causal impacts of this type of program on students' academic and college-related outcomes.

The International Baccalaureate Diploma Programme

IB Diploma Programme students are expected to enroll full-time in the two-year program in 11th and 12th grades and take courses in each of six subject groups (i.e., language, second language, individuals and societies, experimental sciences, mathematics and computer science, the arts). At least three of these courses must be taken at the higher level, while the other courses may be taken at the standard level. Higher-level courses represent approximately 240 teaching hours and standard level courses represent approximately 150 teaching hours. To earn an IB Diploma, candidates must pass the internationally standardized IB exam. Also, they must satisfy the three compulsory components of the IB Diploma Programme: Theory of Knowledge; Extended Essay; and Creativity, Action, Service. IB students who do not fulfill all of the requirements for an IB Diploma may earn an IB Certificate instead. Approximately 80% of participating students earn the IB Diploma (IB Americas, 2011).

IB is less frequently offered than other credit-based transition programs. In the 2002-2003 academic year, of the 16,500 public high schools offering either dual credit, Advanced Placement, or IB, only 390 offered IB (Waits, Setzer, & Lewis, 2005). Although the number of IB Diploma schools in the

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

U.S. has nearly doubled since then, the IB program is still relatively uncommon. Nevertheless, the IB Diploma Programme may be a particularly effective mechanism for increasing academic readiness for college. In contrast to AP, honors, and dual-enrollment in which students may take courses “a la carte,” IB Diploma Programme students are typically required to take an entire curriculum of rigorous coursework. Since IB was first authorized in the United States in 1971, the program has been offered at a growing number of public and private schools and now includes offerings for elementary, middle, and high school years. In 2011, 1,302 IB schools were authorized in the United States: 286 offering the Primary Years Programme, 447 offering the Middle Years Programme, and 753 offering the Diploma Programme (IB Americas, 2011). Over the past decade in the U.S., both the number of schools offering IB and the number of students participating in IB have increased dramatically. Between 2000 and 2011, the number of schools offering the IB Diploma Programme increased 209%, from 360 to 753 (personal communication, J. Sanders, August 10, 2011). More than 60,000 students registered for exams (i.e., were IB Diploma Programme candidates) in 2010-11, up from 22,234 in 2000 (personal communication, J. Sanders, August 10, 2011). In Florida, the IB program began in 1983 in three school districts. Since then, and with both state and local support, the number of school districts in Florida offering the IB program has continued to grow. As of 2011, 68 public high schools in Florida offered the IB Diploma Program, with more than 7,000 students enrolled. This was the second largest IB enrollment among the 50 states.

LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK



Selection bias in participation in IB, AP, and dual enrollment programs may occur from three sources. First, schools choose to offer these programs. Second, schools have processes and practices that formally and informally determine which students have the opportunity to participate. Third, students (and their families) choose to participate in available programs. Research on each of these selection mechanisms is discussed in detail below.

Schools Choose to Offer IB, AP, and Dual Enrollment Programs

In order to participate in credit-based transition programs, students must attend schools where the programs are offered. Yet, descriptive analyses indicate that not all schools choose to offer these programs to their students. One national survey found that, in 2002-03, credit-based transition programs (i.e., dual enrollment, AP, or IB) were less common among public high schools with less than 500 students, rural locations, and high minority enrollments (Waits et al., 2005). The availability of credit-based transition programs also varied by geographic region, as dual enrollment programs were more prevalent in the Central region and less prevalent in the Northeast; AP was more common in the Northeast and less common in the Central region (Waits et al., 2005). Descriptive analyses also reveal differences in dual enrollment participation rates by county and region within the state of Florida, with participation rates ranging from 2.9% to 38.0% in 2006-07 (Estacion et al., 2011). Regional analyses such as this suggest important within-state variation in the availability of credit-based transition programs.

Another recent study uses data from the Florida Education Data Warehouse to identify school-level predictors of offering AP or IB at 407 high schools. Using a series of regression analyses, Iatarola et al. (2011) find a strong association between school size and the likelihood of offering either AP or IB. (The study does not disaggregate AP and IB.) Schools whose size is below the 20th percentile have less than a 60% chance of offering AP or IB courses in all subject areas, while nearly 100% of schools whose size was above the 50th percentile offered these courses. Teacher qualifications were not significantly related to whether a school offered AP or IB. The strongest predictor of offering AP or IB was the number of students with high prior achievement (measured by 8th grade FCAT scores). The authors surmise that schools need a “critical mass” of high-achieving students in order to offer advanced courses.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

The decision of a school to offer a credit-based transition program is likely influenced by several forces. Some schools and/or districts may be constrained from offering these programs because of insufficient human and financial resources. Although not required by the College Board, schools that offer AP classes often incur costs associated with specialized teacher professional development, additional instructional materials, and smaller class sizes (Lerner & Brand, 2008; Office of Program Policy Analysis & Government Accountability (OPPAGA), 2009). Unlike AP, IB courses cannot be taught in a school unless the school implements the entire Diploma Programme (Byrd, 2007). Offering the IB Diploma Programme requires a school to make an initial and continuing financial investment. Schools must submit a \$4,000 non-refundable application fee as well as an annual fee of \$9,500 during the pre-approval/application process. IB Americas (2010) then charges the school a participation fee of \$10,000 per year, as well as a fee of \$141 per student and \$96 per exam.

Representing one third of the nation's public schools and serving nearly 10 million students, rural schools face unique challenges in offering credit-based transition programs (Strange, Johnson, Showalter & Klein, 2012). As a consequence of their smaller size, some rural districts have found that offering AP courses is not only infeasible due to insufficient numbers of qualified teachers and interested, academically-prepared students, but also that offering AP for relatively few students can compromise the general education of the broader majority of students (Irvin, Hannum, Farmer, de la Varre, & Keane, 2009; Barbour & Mulcahy, 2006). Given their relative geographic isolation and lack of close proximity to higher education institutions, rural schools likely struggle to develop the partnerships that are required to offer dual enrollment programs.

On the other hand, while some schools may face resource constraints that limit the availability of credit-based transition programs, other schools may be encouraged to offer these programs because of support from the federal or state government. Since 2008, the federal Advanced Placement Test Fee Program has provided funding to states and educational agencies to subsidize AP and IB exam fees and IB registration fees for low-income students (U.S. Department of Education, 2012). Additionally, at least ten states provide schools and/or districts financial support for equipment, materials and instructional costs associated with offering AP, and three states financially reward schools and/or districts for the number students enrolled

in AP courses or passing AP exams (Education Commission of the States, 2012). In Florida, the state's AP funding program covers exam costs for all students and pays bonuses to teachers of students who pass the exams (Office of Program Policy Analysis & Government Accountability (OPPAGA), 2009). Several states also fund distance learning credit-based transition programs to improve access for students in rural schools (Lerner & Brand, 2008). In 2004, 38 states had legislation regulating various aspects of dual enrollment programs (Karp, Bailey, Hughes, & Fermin, 2004).

Schools Constrain and Enable Student Participation in Available Programs

Schools that choose to offer credit-based transition programs typically have substantial discretion over which students participate, as there are no universal admissions requirements or standards for enrollment in these programs. The College Board states only that it "strongly encourages educators to make equitable access a guiding principle for their AP programs by giving all willing and academically prepared students the opportunity to participate in AP" (The College Board, 2012).

The International Baccalaureate Organization (2010) specifies that admissions criteria for the IB programs are set at the school or district level. Data collected from a survey of IB program coordinators in Florida public high schools reveal differences in admissions criteria and processes (Perna et al., 2013). Whereas the majority of IB programs in Florida reported a minimum GPA requirement, only about half reported requiring prior advanced/honors coursework or a minimum score on a standardized test. A third require a writing sample or a letter of recommendation and a small number of programs require interviews as part of the admissions process (Perna et al., 2013).

Similar flexibility in admission requirements exists for AP and dual enrollment. Even when a state has established laws specifying the criteria to enroll in particular credit-based transition programs, school personnel typically have the ability to determine program participation locally. As an example, although stipulating that students who participate in dual enrollment courses for college credit must have a minimum 3.0 GPA, Florida state law also allows schools to make exceptions to this requirement and to create additional admissions criteria pertaining to grade-level or age of participation and/or to establish more stringent academic requirements

(Estacion et al., 2011; Florida Legislature, 2009). Moreover, while the Florida legislature codified eligibility requirements for dual enrollment, admission requirements for other credit-based transition programs are set locally.

Within schools, counselors play a primary role in determining which students participate in credit-based transition programs (Estacion et al., 2011; Godfrey, 2009; Hertberg-Davis and Callahan, 2008; Siskin et al., 2010). Drawing on data collected from interviews in nine school districts in Florida, one study found that school district and college administrators perceive the high school counselor as more effective at informing prospective students about dual enrollment programs than other sources of information, including printed materials, visits to the high school from college recruiters, group or individual meetings, and word of mouth (Estacion et al., 2011). In a qualitative study of the implementation of IB in four Title I high schools (i.e., schools that enroll a high proportion of low-income students) that ranged in size and student demographics, Siskin and colleagues (2010) concluded that high school counselors play a central gate-keeping role, as they may determine which students participate in the IB Diploma Programme. Using survey data collected from 613 Florida AP teachers and representing 44 school districts, Godfrey (2009) found that most teachers believed that they did not have enough input in selecting students for their AP classes; respondents reported that counselors or AP program coordinators determined students' placement without consulting teachers (Godfrey, 2009).

The discretionary dimension of placement processes may contribute to differences in program participation based on students' race/ethnicity, family income, and other characteristics. In a qualitative study involving approximately 200 teachers, 300 students, 25 building-level administrators and coordinators, and eight program coordinators at 23 schools, Hertberg-Davis and Callahan (2008) concluded that participants believe that the curriculum and instruction within AP and IB courses is not a good fit for all learners, particularly those from traditionally underserved populations.

Students Choose to Participate in Available Programs

Little is known about the processes that students use when deciding whether to participate in an available IB, AP, or dual enrollment program. Descriptive data reveal differences across groups in the characteristics of students who actually participate in credit-based transition programs (Bailey & Karp, 2003;

Chen, Wu, & Tasoff, 2010; College Board, 2011; Estacion et al., 2011; Perna et al., 2013). For instance, despite the growing availability of AP courses, African Americans continue to be underrepresented among AP test-takers relative to their representation among high school students (8.6% versus 14.6% in 2010, College Board, 2011). Latinos represented similar proportions of AP test-takers (16%) and high school students (16.8%) in 2010, largely because of the high participation of Hispanics/Latinos taking the Spanish language examination (Jaschik, 2011). IB programs tend to enroll high-achieving students from families who are aware of the program and its potential benefits to college readiness and admission (Bailey & Karp, 2003), as well as students from higher-income families and with better-educated parents (Chen, Wu, & Tasoff, 2010). These national patterns play out within states, likely reflecting school discretion in determining student participation as well as variations in the decisions and preferences of individual students (and their families). As an example, descriptive analyses show that, although dual enrollment programs are becoming increasingly available in Florida, African American and Latino students, students from low-income families, and English-language learners are underrepresented among participants in dual enrollment (Estacion et al., 2011).

The underrepresentation of African American, Latino, and low-income students in credit-based transition programs mirrors their patterns of representation in academically rigorous coursework (Perna, 2004). Data from the Education Longitudinal Study of 2002 show that African American, Latino students are less likely to take Calculus by their senior year than Whites or Asians (4.7% and 6.8% versus 16.0% and 33.4%). Differences in the share of students taking calculus are also large between the lowest and highest socioeconomic status quartiles (6.2% versus 26.4%, Planty, Bozick, & Ingels, 2006). Using data from the National Educational Longitudinal Study of 1988 (NELS:88), Attewell and Domina (2008) show that, even after accounting for prior achievement, students with higher socioeconomic status are more likely than students with lower socioeconomic status to participate in challenging curricular tracks (a derived variable that includes the number of AP courses taken). However, the opposite is shown for race/ethnicity. After controlling for socioeconomic status and prior academic achievement, African American, Hispanic/Latino and Asian students were more likely to enroll in a challenging curricular track than White students.

Predictors of Student Participation, Recognizing Selection

Despite the clear selection issues, few studies have attempted to statistically model the predictors of program participation taking into account selection at the school and student levels. Using data from Texas public high schools, Klopfenstein (2004) used logistic regression analyses to test an empirical model of racial/ethnic group differences in students' decision to enroll in at least one AP course. The analyses show that, for all students, participation rates are lower for those from low-income families and for those attending large schools. The analyses also reveal differences in the predictors of participation across racial/ethnic groups. For instance, being a recent immigrant reduces the likelihood of participating in AP only among Latinos. Attending a magnet school is associated with an increased likelihood of participation in AP for White and Latino students but a lower likelihood for African American students. Attending a school with a higher share of African American AP teachers is associated with greater likelihood of AP participation for African American males. Nonetheless, while pointing to potential predictors of AP participation, logistic regression alone is insufficient for accounting for school or student level selection issues.

Some studies try to account for selection bias using propensity score matching or stratification. Two recent studies use similar techniques and data from students attending Chicago public schools to examine the predictors of participating in IB (Saavedra, 2011) and the effects of participating in IB on students' college-related outcomes (Coca et al., 2012). Using propensity-score techniques, Saavedra (2011) finds that IB participation rates are higher for Asians than for Whites, lower for African Americans than for Whites, and lower for males than females. The likelihood of participating in IB also increases with students' seventh grade math and reading test scores. Using a longitudinal sample of 13,598 students who graduated from Chicago high schools between 2003 and 2007 and who were eligible to participate in pre-IB program in the 9th grade, Coca et al. (2012) find that participating in IB in the 11th grade is positively related to the likelihood of enrolling in college,

persisting in college, and attending a more selective college or university. Participating only in the pre-IB program was unrelated to these college outcomes. Of the students who enrolled in the pre-IB program, only 62% enrolled in the IB Diploma Programme in the 11th grade.

Nonetheless, the findings from both studies are limited by their consideration of only a small number of variables in their propensity-score analyses. Saavedra (2011) included only a handful of variables that “theoretically should predict enrollment” (p. 10), namely gender, race, family income, 7th grade reading and math test percentiles, and school and cohort fixed effects. Similarly, Coca et al. (2012) also used a fairly limited set of variables to estimate the propensity to participate in pre-IB and then IB. Coca et al. first used 8th grade achievement data to estimate the propensity to participate in the pre-IB cohort in the 9th grade. Then the authors used gender, race/ethnicity, neighborhood poverty, neighborhood socio-economic status, 8th grade percentile on the Illinois Test of Basic Skills, and the students’ elementary school test score average to estimate the propensity to participate in the IB program in the 11th grade. By including only a relatively limited set of predictors in the propensity score models, these approaches may fail to fully account for selection bias (Heckman et al., 1996).

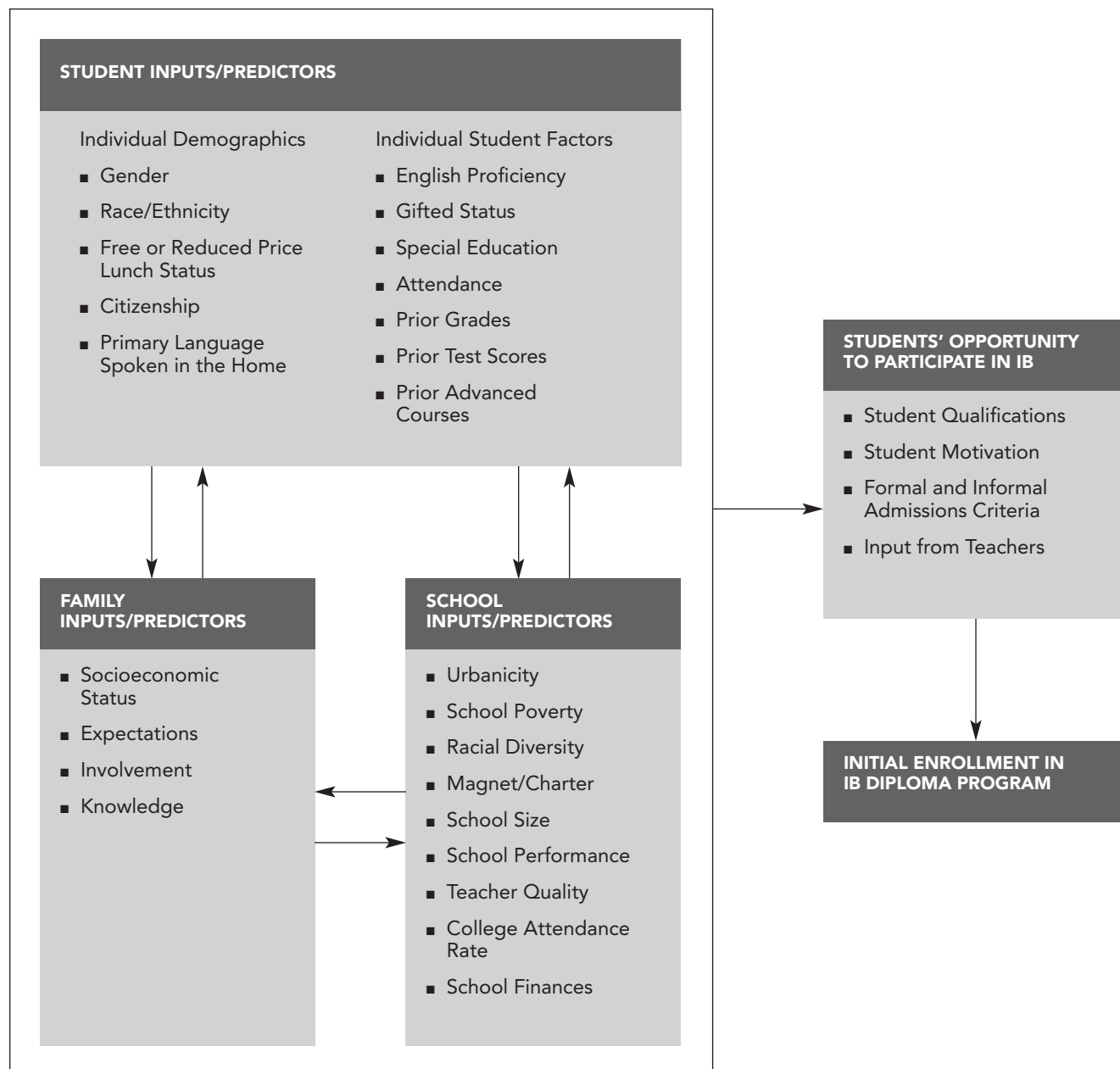
In summary, although recent research is clearly tackling the issue of selection bias in studying the impacts of IB, key factors in the selection process remain unmeasured and uncontrolled. Schools choose to offer these programs, schools have processes and practices that enable and/or restrict student participation, and students within these schools choose to participate in available programs. Moreover, the academic rigor and other unique characteristics of these programs create uncertainty about the extent to which it is possible to use statistical adjustments or create a matched control group of students who are similar to program participants in all measurable ways except for their program participation. Therefore, an essential question is, “What are the key selection factors we should be measuring, and to what extent are data actually obtainable for these constructs?”

A Conceptual Model of IB Participation

Figure 1 illustrates the conceptual model guiding the analyses in this study. This conceptual model is derived from the research reviewed in the previous sections and presumes that a student's decision to enroll in IB is influenced by characteristics of individual students, their families, and the schools they attend. At the student level, participation in IB is expected to correlate with demographic characteristics including gender, race/ethnicity, socioeconomic status, country of birth, and primary language spoken at home (Bailey & Karp, 2003; Chen, Wu, & Tasoff, 2010; Estacion et al., 2011; Klopfenstein, 2004; Perna et al., 2013; Perna, 2004; Saavedra, 2011). Additional family influences such as parents' education, expectations, involvement, and knowledge have been shown to play important roles in selection of IB students (Attewell & Domina, 2008; Bailey & Karp, 2003; Chen, Wu, & Tasoff, 2010; Perna et al., 2013). Research also confirms that participation in IB is related to students' academic characteristics including English-language proficiency, participation in gifted and talented programs, participation in special education, attendance rate, prior grades, prior test scores, and prior success in advanced courses (Bailey & Karp, 2003; Chen, Wu, & Tasoff, 2010; College Board, 2011; Estacion et al., 2011; Florida Legislature, 2009; Perna et al., 2013; Saavedra, 2011). At the school-level, such characteristics as urbanicity, poverty, racial diversity, magnet/charter status, school size, school performance, teacher characteristics, college attendance rate, and school finances are shown to predict IB-participation (Barbour & Mulcahy, 2006; Byrd, 2007; Coca et al., 2012; Irvin, Hannum, Farmer, de la Varre, & Keane, 2009; Karp, Bailey, Hughes, & Fermin, 2004; Iatarola et al., 2011; Lerner & Brand, 2008; OPPAGA, 2009; Strange, Johnson, Showalter & Klein, 2012; Waits et al., 2005). Lastly, through eligibility criteria and recruitment activities, student and school characteristics work together to influence a student's opportunity to participate in IB (Estacion et al., 2011; Godfrey, 2009; Hertberg-Davis and Callahan, 2008; Perna et al., 2013; Siskin et al., 2010).

Deriving a conceptual model of selection into IB allows us to not only recognize the important factors that differentiate IB students and schools from non-IB students and schools, but also to evaluate the extent to which the data available address or ignore aspects of the selection process. In the methods section that follows, we describe the data from the Florida EDW used as indicators for each part of the conceptual model. Use of a conceptual model also allows us to point out which selection factors remain as potential sources of bias, given that no data are available to model them. Lastly, it is important to point out that our conceptual model is probably incomplete. In other words, other selection factors certainly exist that we have yet to recognize.

FIGURE 1.
Conceptual Model of the Inputs/Predictors of Students' Participation in IB



RESEARCH METHODS



Our analyses examine the selection mechanisms behind IB participation across Florida, the state with the second highest representation of IB programs in the nation. Our analyses utilize longitudinal student and school-level data to address the following research questions:

1. What are the student- and school-level predictors of participating in the IB Diploma Programme in Florida?
2. To what degree does propensity score stratification or matching reduce selection bias associated with key student and school-level factors?
3. What are the estimated differences in key postsecondary access indicators (i.e., SAT and ACT scores) and enrollment statistics (e.g., college selectivity) with and without different types of propensity score adjustments?

Population, Sample, and Data

The data used in this study come from the Florida K-20 Education Data Warehouse (FL-EDW) and the U.S. Department of Education's Common Core of Data (CCD). Our subset of data from FL-EDW has student-level records for 20,373 students who participated in an IB Diploma Programme¹ and graduated between 2002 and 2007, and student-level records for 86,008 randomly sampled students who did not participate in an IB Diploma Programme and graduated over the same time period. These records include information from elementary school through high school on student demographics, participation in school programs (e.g., special education, gifted, free/reduced lunch), attendance, promotion/retention, grade point average, state achievement test scores (i.e., FCAT scores), course-taking patterns in high school, SAT and ACT scores, and postsecondary enrollment data. A total of 635 different high schools are represented by one or more students in this sample. The school-level data from the CCD include school type (e.g., regular, alternative, magnet, charter), locale, Title I eligibility, pupil/teacher ratio, student demographics (i.e., by race and free/reduced lunch eligibility), and school size.

¹ All students who participated in IB are included, regardless of whether they earned an IB Diploma.

Comparing our available data to the conceptual model for IB selection (see Figure 1), it is clear that we have numerous indicators representing the majority of student and school factors. Where we lack data are factors that are largely intangible and difficult to measure such as family expectations, involvement, and influence; student motivation; a school's informal admissions criteria; and influence of teachers and counselors. In addition, we are missing information on schools' college attendance rates, but this variable is likely to be highly correlated with school performance data (i.e., state test scores) and may be endogenous with one of our key outcomes (i.e., postsecondary enrollment). We are also missing information on school finances that may influence schools' abilities to offer IB courses and pay program costs.

Thus, although our data represent what may be the most complete set of predictors of IB participation to date, some of the most important factors revealed in our review of the literature on IB participation are not captured. Even though we may be well-positioned to address many aspects of the selection bias that makes comparisons of IB and non-IB students problematic, there are very likely other predictors that are at least as strong as the variables we do have. But even with a relatively complete set of predictors, there is yet another danger—the variables included in our analyses might reveal that IB and non-IB students are like apples and oranges, and that any attempts to adjust for selection bias will be dependent on heroic assumptions and extrapolations of statistical models (Rubin, 2004). In other words, although the models may suggest that IB students tend to have certain combinations of characteristics, it is possible that similar students simply do not exist in the population of non-IB students.

Data Analysis

The procedures for addressing the research questions involved five stages. In the first stage, we used multiple imputation to address missing data problems. In the second stage, we estimated bivariate relationships between IB participation and individual student and school-level variables. In the third stage, we estimated a hierarchical multivariate logistic regression predicting IB participation based on the full set of available student and school-level

variables. The fourth stage used the predicted values from the third stage as propensity scores and assessed comparability of IB students to other students in the state on measures taken prior to 11th grade. These analyses assessed the ability of propensity score stratification and matching to reduce selection bias. The fifth stage used the propensity scores and matching results to estimate adjusted differences in postsecondary outcomes between IB and non-IB students. The data and methods for each of these five stages are described in more detail below.

First, multiple imputation (Rubin, 1987) was used to address missing data among our predictor variables. The multiple imputation process was carried out separately for each cohort of students. Across the six cohorts, most variables in the analyses had little to no missing data, with nearly all variables missing less than 5% of their data. In one exception, “highest math course through 10th grade” was missing between 7% and 9% of the data across the 2002 through 2007 graduating cohorts.

Missing data was a larger challenge for state-administered test score data. FCAT scores from grades 3-8 were generally unavailable for students graduating before 2005 (because they completed the 8th grade prior to the roll-out of the current FCAT assessment in 2001). The rates of missing data for elementary/middle grade average FCAT scores were 26%, 19%, and 17% in 2005, 2006, and 2007 respectively. The rates of missing data for average 9th/10th grades FCAT scores were 14% in 2003, and ranged from 2% to 4% from 2004 through 2007. As such, analyses of pre-2005 cohorts did not include elementary or middle grades FCAT data. Likewise, because 9th and 10th grade FCAT scores are not available for students who graduated in 2002, this variable was not included in analyses for that year.

Although no data were missing for the school-level variables or the IB participation indicator, these variables were included in the imputation process to improve precision and accuracy of the results (Allison, 2001). PROC MI in SAS 9.3 was used to create the imputed data sets. Markov Chain Monte Carlo (MCMC) via Gibbs Sampling (Geman & Geman, 1984) was used², with starting values obtained based on the covariance matrix estimated via the Expectation Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). Categorical variables were recoded into dummy indicators, and imputed values were rounded to the nearest category value

² The MCMC chain included a 500 iteration burn-in period to allow the Gibbs sampler to converge, which was followed by 30 periods of 50 iterations (i.e., 1,500 iterations total), with plausible values drawn after the burn-in and at the end of each 50 iteration period.

(Allison, 2001). Trace and autocorrelation plots were used to assess MCMC convergence and independence of plausible value imputations (Enders, 2010). The MCMC imputation exhibited rapid convergence (in fewer than 200 iterations) for all variables. Autocorrelation plots suggested that imputations for all variables are independent after 20 iterations or less. These results from the multiple imputation process suggest that the MI process successfully imputed plausible values.

Thirty plausible values were drawn instead of the typical five in order to average across plausible values and produce a single-imputation dataset for estimation of individual-level propensity scores (Little & Rubin, 2002). Using multiple plausible values for the production of individual propensity scores is unnecessary since we seek only the maximum likelihood point estimate for each propensity score, and not the standard errors. For our other analyses in which standard errors and p-values were produced (i.e., those analyses which focus on the significance of individual predictors), the increase in variance due to imputation was important to capture, but 30 plausible values was far more than is needed for such an analysis; therefore, for those models, we used a more traditional subset of five plausible values evenly spaced throughout the full set of 30 plausible values (i.e., the 6th, 12th, 18th, 24th, and 30th plausible values).

After imputing missing data, our second stage of analyses focused on estimating bivariate relationships between IB participation and individual student and school-level predictor variables. Because many of the predictor variables in the dataset were highly correlated, the confounding and multicollinearity between them was expected to cause parameter estimates to behave strangely in a multiple regression model. For example, while FCAT math scores may be positively related to IB participation, estimating this relationship in a multiple regression model that also includes FCAT reading scores may cause the coefficient for math scores to become insignificant or even negative. This finding might suggest that, among two students with the same reading scores, the student with the lower math score is more likely to participate in IB. On the other hand, this change may be an artifact of collinearity and instability in the estimate, thereby complicating the interpretation of the coefficient. Therefore, to reduce confusion about the significance, direction, and magnitude of student- and school-level predictors of IB participation, we first estimated the relationship between each predictor variable and IB participation without including any control variables in the

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

model. We included random effects for schools in this analysis to reflect the multilevel nature of the data and to produce accurate standard errors for the school-level predictors. Each multilevel logistic regression model was estimated using PROC GLIMMIX in SAS 9.3. Separate models were estimated for each of the five imputed datasets described above, with results combined using PROC MIANALYZE. The resultant logistic regression parameters were converted to odds ratios, with those for categorical indicators reflecting the difference in odds of IB participation relative to a reference category, and with the estimates for continuous variables reflecting the difference in odds of IB participation associated with a one standard deviation increase in the predictor variable.

The third stage of analyses involved estimating a multiple logistic regression model predicting IB participation based on all available student and school characteristics. The primary function of this model was to produce propensity score estimates (Rosenbaum & Rubin, 1983) that reflected the probability of each student enrolling in an IB program, conditional on all measured characteristics of that student and his/her school. Since the selection mechanisms involve both school and student-level processes, a multilevel model with school random effects is the preferred method for estimating propensity scores (Steiner, 2011). Unlike fixed effects approaches, which support only within-school matching and would require many whole-school IB programs be excluded from our analyses, our multilevel propensity model allows students to be stratified or matched both within and across schools, with proper recognition that schools with certain characteristics are more likely to offer IB programs. Once again, our multilevel logistic regression models were estimated using PROC GLIMMIX in SAS 9.3, but now with all predictors entered simultaneously. The best linear unbiased predictors (BLUPs) from these models were used as estimates of the individual propensity scores (Steiner, 2011). The individual propensity scores are based on the averaged imputed dataset, while the standard errors for propensity score model coefficients are based on analyses involving the subset of five plausible values.

In the fourth stage of analyses, the propensity scores were used to assess and correct observed selection bias in measured student and school characteristics. The estimated propensity scores were compared for IB and non-IB students through visual inspection of density plots. Next, three alternative approaches were undertaken to evaluate the utility of the

propensity score method in reducing selection bias. The first involves propensity score stratification, in which the propensity score is used to create groups of IB and non-IB students with similar propensity scores. In this case, we divided the propensity scores into five evenly spaced strata (i.e., 0-.20, .20-.40, .40-.60, .60-.80, .80-1.00). Comparability of IB and non-IB students after propensity score stratification was then assessed using multilevel linear and logistic regression to estimate differences on student and school-level variables. Combined models pooling data across the six cohorts were estimated, with random intercepts for each cohort within each school. Fixed effects were included for the stratification variable, and the raw propensity scores were also included as a continuous control variable (as suggested by Rubin, 2004).

The second and third approaches to evaluating the utility of the propensity score method in reducing selection bias involved using the propensity score to create matched groups of IB and non-IB students with nearly identical propensity scores. The matching process was implemented using the `fullmatch` and `pairmatch` algorithms from the `Optmatch` library (Hansen & Fredrickson, 2012, version 0.7-3) as implemented in R x64 (version 2.15.0). Optimal full matching links each IB student to at least one non-IB student and also allows each non-IB student to be matched to multiple IB students, although each student appears in only one matched group. This also allows use of the full sample instead of matching only a subset of students. Optimal pair matching links each IB student to no more than one control student, with unmatchable IB students dropped from the dataset in subsequent analyses. Rosenbaum (2010) shows that optimal full matching typically produces the best results of any matching method. When full matching is performed using the complete sample (as in our study), it is similar to stratification with a potentially infinite number of strata (i.e., the matching algorithm determines the optimal number of strata). Pair matching can result in substantial reductions in sample size when estimated propensity scores have limited overlap between the two groups.

Once the matching process was completed, fixed effects for the matched groups were included in subsequent analyses to adjust selection bias. Comparability of IB and non-IB students after propensity score matching was again assessed using the same multilevel modeling strategy described in the previous paragraph, with the addition of the matched group fixed effects under both full matching and pair matching.

RESULTS

Predictors of IB Participation



The results from the bivariate analyses of student- and school-level predictors of IB participation are shown in Tables 1 through 4. Several student background characteristics, as shown in Table 1, are related to participation in IB. The majority of estimates are remarkably stable over time. Across the six cohorts, male students are 19% less likely than female students to participate in IB (i.e., $100\% - 81\% = 19\%$). Compared with White students, Asian students are about 3.1 times more likely to participate, while African American students are more than 70% less likely and Latino students are about 40% less likely to participate. Some estimates suggest that Native American and multiracial students are less likely to participate, but the relationships are not consistent across years.

Students who are U.S. citizens are 41% more likely than non-citizens to participate, while non-resident aliens (e.g., students whose parents have a valid visa to work and reside in the U.S.) are 2.1 times as likely to participate in IB. Students who speak English as the primary language in their homes are 25% more likely to participate in IB, and students whose parents speak English are 19% more likely to participate in IB; however, these trends appear to diminish or even disappear in more recent years. Compared with other students, students identified as having limited English proficiency are more than 85% less likely to participate in IB, while special education students are 58% less likely to participate. Students who are eligible for free or reduced lunch are 70% less likely to participate in the IB Diploma Programme, whereas students identified as gifted are 700% more likely (i.e., 7 times more likely) than other students to participate.

TABLE 1.
Bivariate Odds Ratios for Student Demographic Predictors of Participation
in the International Baccalaureate (IB) Diploma Programme

	YEAR OF HIGH SCHOOL GRADUATION						
	2002	2003	2004	2005	2006	2007	2002-07
Number of IB Students	2,927	3,000	3,223	3,507	3,754	3,962	20,373
Number of Non-IB Students	13,108	13,937	14,215	14,247	14,888	15,613	86,008
PREDICTOR VARIABLE							
Male	0.81**	0.81**	0.85*	0.77***	0.77***	0.82**	0.81***
Race/Ethnicity (Caucasian reference)							
Asian	2.95***	3.06***	2.96***	3.02***	2.88***	3.63***	3.09***
African American	0.28***	0.26***	0.26***	0.27***	0.28***	0.24***	0.27***
Hispanic/Latino/ Latina	0.62***	0.51***	0.57***	0.66***	0.62***	0.56***	0.59***
Native American	0.61	1.02	0.62	0.73	0.96	0.39*	0.68 ^o
Multiracial	0.24 [~]	0.17*	0.11**	1.41	1.21	0.46	0.51**
US Residency Status							
Nonresident Alien	1.20	1.13	7.81**	1.34	1.63	3.13	2.13**
US Citizen	1.16	1.58***	1.31*	1.47***	1.37**	1.59***	1.41***
Born outside the US	0.95	0.85 [~]	1.09	1.01	1.07	1.04	1.00
Family Language							
English	1.33**	1.50***	1.30**	1.13	1.26**	1.10	1.25***
Parent Speaks English	1.32**	1.43***	1.23*	1.05	1.15	1.10	1.19***
School Program Participation							
Limited English Proficiency	0.16***	0.18***	0.13***	0.14***	0.11***	0.11***	0.14***
Special Education Student	0.53***	0.38***	0.37***	0.40***	0.43***	0.47***	0.42***
Free/Reduced Lunch Eligible	0.31***	0.30***	0.28***	0.35***	0.30***	0.29***	0.30***
Gifted Student	7.35***	7.30***	9.05***	6.06***	5.95***	6.80***	6.97***

Note [~]p<.10, *p<.05, **p<.01, ***p<.001

Odds ratios in this table are based on bivariate multilevel models (students within schools) with no control variables.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

As shown in Table 2, indicators of prior student academic performance are highly predictive of participation in IB, also with very stable estimates across the six cohorts. Attendance is moderately related to IB participation—a one standard deviation increase in attendance rate is associated with a 95% increase in the odds of participating in IB. Having ever been retained in grade is a very strong predictor, with retention associated with an 87% drop in the odds of IB participation. Grade point averages in the 9th and 10th grades are also highly predictive of IB participation, although the positive relationship is greater for weighted GPA than unweighted GPA; a one standard deviation increase in weighted 9th grade GPA is associated with as much as a 518% increase in the odds (i.e., 5.18 times) of participating in IB. Prior FCAT scores in reading and math are also highly predictive of participation in IB. A one standard deviation increase in FCAT math scores while in the elementary grades is associated with a 742% increase in the odds of participating in IB, while a one standard deviation increase in FCAT math scores in 9th and 10th grades is associated with a 749% increase in the odds of participating in IB.

TABLE 2.
Bivariate Odds Ratios for Student Performance Indicators as Predictors
of Participation in the International Baccalaureate (IB) Diploma Programme

	YEAR OF HIGH SCHOOL GRADUATION						
	2002	2003	2004	2005	2006	2007	2002-07
Number of IB Students	2,927	3,000	3,223	3,507	3,754	3,962	20,373
Number of Non-IB Students	13,108	13,937	14,215	14,247	14,888	15,613	86,008
PREDICTOR VARIABLE							
Average Attendance Rate ^a	1.80***	2.00***	1.88***	1.97***	1.90***	2.12***	1.95***
Retained in Grade at Least Once	0.14***	0.34***	0.10***	0.09***	0.07***	0.07***	0.13***
Prior Grade Point Average^a							
Unweighted 9th Grade GPA	3.26***	3.30***	3.41***	3.69***	3.55***	3.88***	3.52***
Unweighted 10th Grade GPA	2.68***	3.03***	2.66***	2.99***	2.70***	3.03***	2.85***
Weighted 9th Grade GPA	4.60***	4.86***	5.21***	5.53***	5.31***	5.62***	5.18***
Weighted 10th Grade GPA	3.70***	4.36***	3.97***	4.36***	3.92***	4.26***	4.09***
Prior FCAT Test Scores^a							
Mean FCAT Math Score in Grades 3-8				6.34***	7.98***	8.14***	7.42***
Mean FCAT Reading Score in Grades 3-8				4.24***	5.92***	6.20***	5.37***
Mean FCAT Math Score in Grades 9-10		6.99***	8.08***	7.59***	7.07***	7.84***	7.49***
Mean FCAT Reading Score in Grades 9-10		5.28***	7.07***	6.62***	6.11***	6.02***	6.17***

Note ~p<.10, *p<.05, **p<.01, ***p<.001

Odds ratios in this table are based on bivariate multilevel models (students within schools) with no control variables.

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Course-taking indicators, shown in Table 3, are also very highly predictive of participation in IB. Students who fail to reach Algebra II by the 10th grade (i.e., the standard course for the college prep track in Florida) are 97% less likely to participate in IB, while students who reach Trigonometry or Pre-Calculus by the 10th grade are 8.2 times more likely to participate in IB. There are also positive estimates for reaching Calculus or above by the 10th grade, but these estimates vary greatly across years due to the very small number of students taking these advanced math classes by the 10th grade. A key gatekeeper, taking Algebra I before or after the 9th grade is one of the strongest predictors of participation in IB. While those students who take Algebra I late (i.e., after the 9th grade) are 96% less likely to participate in IB, those students who take Algebra I early (i.e., in 8th grade or before) are 23 times more likely (i.e., 2,300% more likely) to participate in IB. Lastly, the number of advanced credits (e.g., honors, AP courses) taken in 9th and 10th grade is also very highly predictive of IB participation. For example, a one standard deviation increase in the number of advanced courses taken in 10th grade is associated with a 4,300% increase in the odds of participating in IB.

TABLE 3.
Bivariate Odds Ratios for Early High School Course-Taking Indicators as Predictors
of Participation in the International Baccalaureate (IB) Diploma Programme

	YEAR OF HIGH SCHOOL GRADUATION						
	2002	2003	2004	2005	2006	2007	2002-07
Number of IB Students	2,927	3,000	3,223	3,507	3,754	3,962	20,373
Number of Non-IB Students	13,108	13,937	14,215	14,247	14,888	15,613	86,008
PREDICTOR VARIABLE							
Highest Math Through 10th Grade (reference: Algebra II)							
Basic Math	0.07***	0.02***	0.02***	0.02***	0.01***	0.04***	0.03***
Algebra I	0.02***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
Geometry	0.07***	0.04***	0.02***	0.02***	0.02***	0.03***	0.03***
Trigonometry/ Precalculus	11.27***	8.56***	8.12***	7.53***	8.17***	6.92***	8.20***
Calculus or Above	1.97~	1.76	1.44	6.84**	3.13*	2.41*	2.34***
Late Algebra I (after 9th Grade)	0.06***	0.04***	0.04***	0.05***	0.04***	0.04***	0.04***
Early Algebra I (before 9th Grade)	17.58***	25.07***	25.97***	21.20***	26.53***	23.73***	23.15***
Advanced Credits in 9th Grade ^a	15.19***	13.12***	14.95***	24.51***	26.42***	16.89***	17.78***
Advanced Credits in 10th Grade ^a	26.63***	31.13***	39.37***	88.66***	121.7***	35.99***	43.29***

Note ~p<.10, *p<.05, **p<.01, ***p<.001

Odds ratios in this table are based on bivariate multilevel models (students within schools) with no control variables.

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 4 suggests that most school-level variables are only weakly or moderately predictive of IB participation. While students attending magnet schools are 4.7 times more likely to participate in IB, students attending rural schools rather than suburban schools are up to 63% less likely to participate. Students attending schools with higher pupil-teacher ratios are also less likely to participate in IB. A one standard-deviation increase in pupil-teacher ratio is associated with a 19% reduction in the odds of participating in IB. The same pattern holds for schools serving greater numbers of poor students. A one standard-deviation increase in the percentage of students enrolled who are eligible for free or reduced lunch is associated with a 24% reduction in the odds of participating in IB. Another school-level predictor that is consistently related to IB participation is the percentage of the student population that is Asian. A one standard deviation increase in the percentage of Asian students is associated with a 430% increase in the odds of participation in IB. Admittedly, this variable has a very restricted range from 0% to 14%, with a standard deviation of 2.4 percentage points. Other school-level race/ethnicity demographics have weaker relationships with IB participation. A one standard deviation increase in the percentage of Latino students is associated with a 28% decrease in the odds of participation in IB, while a one standard deviation increase in the percentage of African American students is associated with a 13% increase in the odds of participation in IB.

By far, the strongest school-level predictors of IB participation are school-mean FCAT scores in math and reading. A one standard deviation increase in the school's average FCAT math score is associated with a 1,154% increase in the odds of participation in IB, while a one standard deviation increase in the school's average FCAT reading score is associated with a 841% increase in the odds of participation in IB.

TABLE 4.
Bivariate Odds Ratios for School-Level Predictors of Participation in the International Baccalaureate (IB) Diploma Programme

	YEAR OF HIGH SCHOOL GRADUATION						
	2002	2003	2004	2005	2006	2007	2002-07
Number of IB Students	2,927	3,000	3,223	3,507	3,754	3,962	20,373
Number of Non-IB Students	13,108	13,937	14,215	14,247	14,888	15,613	86,008
PREDICTOR VARIABLE							
Regular School (vs. Alternative or Special Ed)	0.50	0.58	0.81	0.32	0.36	2.14	0.58 [~]
Magnet School					3.60 ^{***}	5.98 ^{***}	4.67 ^{***}
Charter School	3.17	0.01	0.00	0.24	0.75	0.36	0.48
New School	1.29	0.01	1.59	0.01	6.21	2.09	1.32
Urban	1.05	0.87	1.02	1.07	1.05	1.00	1.01
Rural	0.53	0.28 [*]	0.28 [*]	0.39 [~]	0.31 [*]	0.45 ^o	0.37 ^{***}
Title I School	0.74	1.01	0.56	0.64	0.34	0.82	0.73
School-Wide Title I	0.69	1.11	0.52	0.74	0.39	0.71	0.70
Pupil/Teacher Ratio ^a	0.93	0.76	0.76 [~]	0.79	0.84	0.82	0.81 ^{**}
Percent Free/Reduced Lunch ^a	0.86	0.75	0.70 [*]	0.76	0.77	0.73 [~]	0.76 ^{***}
Percent Asian ^a	3.90 ^{***}	4.24 ^{***}	4.97 ^{***}	4.71 ^{***}	4.06 ^{***}	4.10 ^{***}	4.30 ^{***}
Percent Hispanic/Latino/Latina ^a	0.66 [~]	0.74	0.65 [*]	0.66 [~]	0.74	0.83	0.72 ^{***}
Percent African American ^a	1.18	1.13	1.14	1.08	1.22	1.02	1.13 [~]
Percent White ^a	0.98	0.98	1.02	1.07	0.93	1.01	1.00
School Size ^a	0.93	1.03	0.99	1.05	1.03	1.24	1.05
School Mean FCAT Math Scores in Grades 9-10		9.47 ^{***}	8.41 ^{***}	10.02 ^{***}	7.77 ^{***}	10.59 ^{***}	11.54 ^{***}
School Mean FCAT Reading Scores in Grades 9-10		8.75 ^{***}	6.01 ^{***}	9.81 ^{***}	7.35 ^{***}	11.36 ^{***}	8.41 ^{***}

Note [~]p<.10, ^{*}p<.05, ^{**}p<.01, ^{***}p<.001

Odds ratios in this table are based on bivariate multilevel models (students within schools) with no control variables.

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Multivariate Prediction of IB Participation

Table 5 shows the results from the multiple logistic regression analyses of individual and school-level predictors of IB participation. As described in the methods section, interpretation of the model slope parameters for individual predictors is difficult given the high degree of confounding and multicollinearity between predictors. Many of the predictors that were significant in the bivariate models are now insignificant in the multivariate model. In addition, the slope estimates for a number of variables (e.g., attendance, unweighted GPA) have actually changed sign in some years, making interpretation potentially confusing.

Nonetheless, the main purpose of this model is not to interpret coefficients for specific variables, but to maximize the predictive power for explaining who does and does not participate in IB (Rosenbaum & Rubin, 1983; Rosenbaum, 2002; Rubin, 2004). As such, collinearity and unstable coefficients are not a concern since including as many predictors as available serves only to improve the accuracy of the predictions (Rosenbaum, 2002). In fact, Table 5 shows that the concordance index for these six models is incredibly high, ranging from 99.2% to 99.5%. The high concordance index suggests that the multivariate models are able to correctly distinguish IB and non-IB students more than 99% of the time. Although resulting in high predictive power, the inclusion of many predictors limits our ability to interpret individual slope parameters. The high predictive power confirms great dissimilarity between IB participants and non-participants.

TABLE 5.
Parameter Estimates from a Multiple Logistic Regression Predicting Participation
in the International Baccalaureate (IB) Diploma Programme

	YEAR OF HIGH SCHOOL GRADUATION					
	2002	2003	2004	2005	2006	2007
Number of IB Students	2,927	3,000	3,223	3,507	3,754	3,962
Number of Non-IB Students	13,108	13,937	14,215	14,247	14,888	15,613
PREDICTOR VARIABLE						
Intercept	-1.84 [~] (0.98)	-6.06 ^{**} (1.85)	-5.73 ^{**} (1.80)	-3.04 [*] (1.43)	-2.53 [*] (1.24)	-7.27 ^{***} (2.08)
Male	-0.08 (0.14)	0.05 (0.17)	0.25 (0.18)	-0.11 (0.17)	0.09 (0.17)	0.12 (0.14)
Asian	-0.06 (0.30)	0.67 [~] (0.37)	0.58 (0.36)	0.92 ^{**} (0.33)	0.59 [~] (0.33)	0.35 (0.29)
African American	-0.13 (0.22)	0.10 (0.27)	0.08 (0.29)	-0.15 (0.31)	0.12 (0.31)	-0.10 (0.25)
Latino/Latina	0.31 (0.28)	0.28 (0.29)	0.38 (0.31)	0.25 (0.29)	0.10 (0.27)	0.18 (0.24)
Native American	-0.58 (1.06)	0.27 (1.01)	-2.43 (2.07)	-1.34 (1.07)	0.09 (1.48)	-0.09 (1.20)
Multiracial	0.02 (1.34)	-1.03 (1.89)	-1.10 (3.01)	3.10 ^{***} (0.90)	0.45 (1.33)	0.24 (1.60)
Non-Resident Alien	-1.84 (2.56)	-0.15 (1.42)	0.01 (1.61)	-0.45 (1.34)	-1.46 (1.45)	1.49 (1.15)
US Citizen	-1.01 [*] (0.49)	0.48 (0.47)	-0.35 (0.49)	-0.04 (0.43)	0.09 (0.42)	-0.28 (0.38)
Born outside of the US	-0.64 (0.48)	0.63 (0.44)	-0.01 (0.42)	0.37 (0.39)	0.35 (0.38)	0.27 (0.34)
English is home language	-0.35 (0.39)	-0.45 (0.42)	-0.10 (0.49)	0.22 (0.42)	0.07 (0.41)	-0.05 (0.33)
Parents speak English	0.17 (0.40)	0.57 (0.42)	-0.42 (0.49)	-0.35 (0.39)	-0.55 (0.39)	-0.12 (0.31)
Limited English Proficiency	-0.28 (0.32)	-0.41 (0.38)	-0.26 (0.40)	-0.29 (0.42)	0.15 (0.39)	-0.40 (0.34)
Special Education Student	0.92 ^{***} (0.23)	0.17 (0.30)	0.14 (0.32)	-0.02 (0.29)	-0.27 (0.29)	-0.25 (0.23)
Gifted Student	0.13 (0.17)	0.21 (0.20)	0.24 (0.22)	-0.01 (0.21)	-0.14 (0.21)	0.19 (0.18)
Free/Reduced Lunch Eligible	0.11 (0.17)	-0.16 (0.19)	-0.04 (0.21)	-0.04 (0.20)	0.21 (0.20)	-0.07 (0.16)

Note [~]p<.10, *p<.05, **p<.01, ***p<.001

TABLE 5. (continued)
**Parameter Estimates from a Multiple Logistic Regression Predicting Participation
 in the International Baccalaureate (IB) Diploma Programme**

PREDICTOR VARIABLE	YEAR OF HIGH SCHOOL GRADUATION					
	2002	2003	2004	2005	2006	2007
Average Attendance Rate	-0.09 (0.08)	-0.09 (0.08)	-0.24* (0.10)	0.09 (0.10)	0.07 (0.11)	0.18~ (0.09)
Retained in Grade at Least Once	0.17 (0.28)	0.52~ (0.30)	0.21 (0.36)	0.07 (0.37)	-0.62 (0.45)	0.11 (0.30)
Unweighted GPA in 9th Grade	-0.75 (0.67)	0.39 (0.94)	-1.25 (0.95)	-0.27 (0.93)	-1.73~ (0.95)	-0.84 (0.67)
Weighted GPA in 9th Grade	0.98 (0.74)	-0.17 (1.08)	1.62 (1.06)	0.51 (1.02)	2.24~ (1.09)	1.11 (0.76)
Unweighted GPA in 10th Grade	-1.19~ (0.64)	-1.18 (0.78)	-1.18 (0.92)	-2.27** (0.81)	-1.45~ (0.84)	-0.74 (0.69)
Weighted GPA in 10th Grade	1.78* (0.72)	1.75* (0.85)	1.77~ (1.04)	2.82** (0.89)	1.76~ (0.94)	1.23 (0.77)
Mean FCAT Math Score Grade 3-8				0.02 (0.17)	0.11 (0.23)	-0.38* (0.19)
Mean FCAT Reading Score Grade 3-8				0.10 (0.13)	-0.03 (0.17)	-0.54** (0.17)
Mean FCAT Math Score Grade 9-10		0.14 (0.12)	0.03 (0.15)	0.04 (0.17)	-0.19 (0.19)	0.08 (0.15)
Mean FCAT Reading Score Grade 9-10		0.22~ (0.12)	0.36** (0.12)	0.25 ^o (0.14)	0.36** (0.14)	-0.03 (0.13)
Basic Math	0.88~ (0.48)	-1.56 (1.18)	1.12 (0.84)	-0.80 (1.27)	-1.02 (1.75)	0.90 (0.90)
Algebra I	0.10 (0.33)	0.07 (0.38)	0.62 (0.51)	-0.17 (0.44)	0.41 (0.47)	-0.49 (0.41)
Geometry	0.31 (0.27)	-0.04 (0.28)	0.13 (0.32)	-0.71* (0.31)	-0.49 (0.33)	-0.16 (0.25)
Trigonometry/ Pre-calculus	0.43 (0.28)	0.27 (0.30)	-0.27 (0.30)	0.05 (0.30)	0.21 (0.31)	0.67** (0.25)
Calculus or Above	-1.61~ (0.85)	-1.42 (1.14)	-1.72~ (1.01)	0.15 (0.98)	-1.35 (0.89)	-1.15 (0.79)
Early Algebra I	0.59* (0.25)	0.73** (0.26)	0.76** (0.27)	0.14 (0.26)	0.19 (0.30)	0.26 (0.23)
Advanced Credits in 9th Grade	1.12*** (0.12)	0.75*** (0.12)	0.95*** (0.15)	1.23*** (0.16)	1.21*** (0.14)	1.04*** (0.12)
Advanced Credits in 10th Grade	1.91*** (0.13)	2.42*** (0.17)	2.65*** (0.18)	2.86*** (0.20)	2.83*** (0.17)	2.24*** (0.14)

Note ~p<.10, *p<.05, **p<.01, ***p<.001

TABLE 5. (continued)
Parameter Estimates from a Multiple Logistic Regression Predicting Participation
in the International Baccalaureate (IB) Diploma Programme

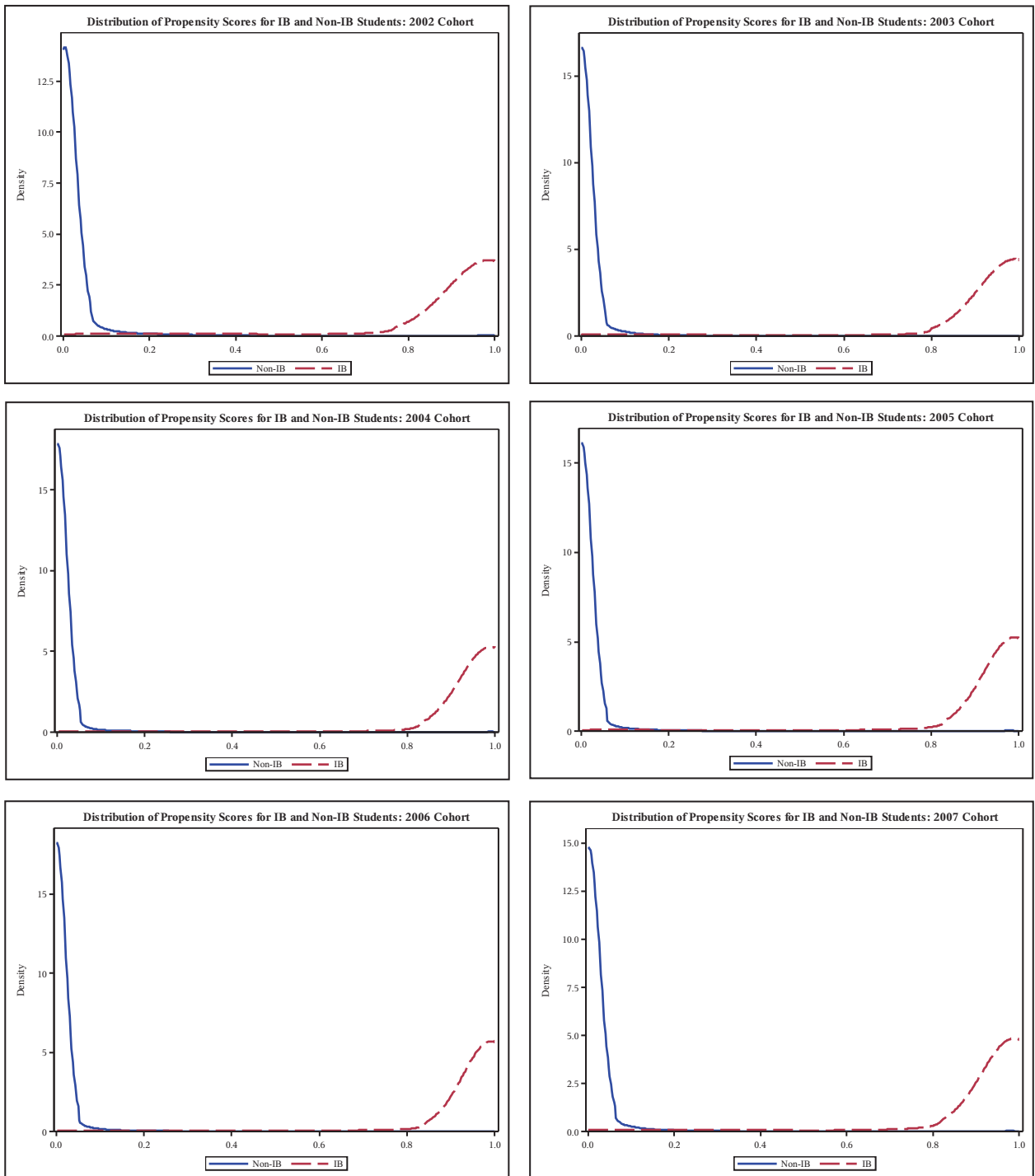
PREDICTOR VARIABLE	YEAR OF HIGH SCHOOL GRADUATION					
	2002	2003	2004	2005	2006	2007
Regular School	-1.85* (0.77)	-1.35 (1.70)	-0.27 (1.52)	-3.07* (1.31)	-2.92** (1.07)	1.16 (1.99)
Magnet School					-0.06 (0.56)	0.41 (0.53)
Charter School	2.30 (1.88)	-20.03 (66.46)	-7.93 (12.38)	1.92 (2.35)	3.09* (1.22)	1.63 (1.28)
New School	1.85 (1.55)	-3.85 (786.73)	-1.92 (2.86)	-0.23 (6.33)	5.24** (1.89)	6.06* (2.62)
Urban	-0.11 (0.39)	-1.01 (0.64)	-0.29 (0.56)	-0.17 (0.60)	0.21 (0.50)	0.15 (0.52)
Rural	-0.03 (0.44)	-0.88 (0.84)	-0.69 (0.73)	-0.29 (0.70)	-0.48 (0.61)	0.12 (0.57)
Title I Eligible	1.47 (1.59)	-16.91 (110.26)	1.82 (1.82)	-2.71 (6.61)	-9.81 (427.90)	1.53* (0.73)
School-wide Title I	-2.08 (1.88)	16.85 (110.25)	-3.34 (2.04)	1.68 (6.73)	8.17 (427.90)	-0.85 (0.76)
Pupil/Teacher Ratio	-0.15 (0.23)	-0.48 (0.43)	-0.18 (0.33)	-0.48 (0.35)	-0.56* (0.23)	-0.22 (0.20)
Percent Free/Reduced Lunch	0.48* (0.22)	0.91* (0.38)	0.65~ (0.37)	1.22** (0.38)	1.19*** (0.35)	0.50 (0.47)
Percent Asian	1.78 (1.36)	-0.69 (2.36)	0.97 (1.66)	1.72 (2.25)	0.01 (2.16)	1.48 (2.41)
Percent Hispanic/Latino/Latina	8.40 (12.73)	-10.73 (21.78)	-0.62 (14.83)	9.12 (20.01)	-1.49 (18.51)	11.22 (20.48)
Percent African American	8.42 (12.11)	-9.66 (20.60)	0.31 (14.03)	9.14 (18.39)	-1.36 (17.05)	10.42 (18.48)
Percent White	10.97 (15.91)	-13.23 (27.37)	-0.01 (18.41)	12.31 (24.53)	-1.72 (22.69)	14.00 (24.82)
School Size	-0.02 (0.24)	0.09 (0.41)	-0.00 (0.38)	0.44 (0.34)	0.41 (0.27)	0.66* (0.29)
School Mean FCAT Math Scores in Grades 9-10		1.20 (0.79)	2.69** (0.81)	-0.54 (0.77)	1.12~ (0.61)	0.16 (0.73)
School Mean FCAT Reading Scores in Grades 9-10		0.75 (0.81)	-1.40~ (0.70)	2.34** (0.82)	0.79 (0.62)	1.87* (0.75)
Model Concordance Index	99.5	99.3	99.5	99.4	99.4	99.2

Note ~p<.10, *p<.05, **p<.01, ***p<.001

Comparison of Propensity Scores for IB and Non-IB Students

The propensity scores from the multilevel multiple logistic regression model were used as estimates of the probability that each student participated in IB. Figure 2 shows density plots of propensity scores by year. In each of these plots, one thing is very clear—there is little overlap in the distribution of estimated propensity scores between IB students and non-IB students. The propensity scores for IB students are heavily left-skewed, while the propensity scores for the non-IB students are even more heavily right-skewed. The vast majority of IB students' propensity scores are lumped mostly at the high end (i.e., between .80 and 1.0), while the propensity scores for the non-IB students are lumped mostly at the low end (i.e., between 0.0 and .10). Still, the long tails of the distributions suggest that at least some non-IB students have high propensity scores, and some IB students have low propensity scores.

FIGURE 2.
Density Plots of Estimated Propensity Scores
for Six Cohorts of IB and Non-IB Students



Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 6 groups students into five strata of propensity scores and shows the numbers of IB and non-IB students in each strata by year and in total. Across all the years, over 90% of IB students have propensity scores greater than .80, while over 97% of non-IB students have propensity scores less than .20. This pattern suggests that IB students are, in general, very different from the larger population of students in Florida. Given the large sample of students in the dataset (about 100,000 cases), however, we are able to identify nearly 300 non-IB students with propensity scores between .80 and 1.0, and over 800 additional non-IB students with propensity scores between .20 and .80. Then again, comparing a sample of over 20,000 IB students to a sample of only 1,100 non-IB students suggests that IB students are quite unlike the vast majority of students in general.

TABLE 6.
Counts of IB and Non-IB students in Five Propensity Score Strata

GRADUATION YEAR	PROPENSITY SCORE STRATA (PREDICTED PROBABILITY OF IB PARTICIPATION)				
	0.00 – 0.20	0.20 – 0.40	0.40 – 0.60	0.60 – 0.80	0.80 – 1.00
2002					
Non-IB Students	12,839 97.9%	170 1.3%	42 0.3%	12 0.1%	45 0.3%
IB Students	89 3.0%	72 2.5%	53 1.8%	69 2.4%	2,644 90.3%
2003					
Non-IB Students	13,775 98.8%	70 0.5%	34 0.2%	21 0.2%	37 0.3%
IB Students	75 2.5%	44 1.5%	39 1.3%	47 1.6%	2,795 93.2%
2004					
Non-IB Students	14,066 99.0%	74 0.5%	24 0.2%	12 0.1%	39 0.3%
IB Students	59 1.8%	29 0.9%	29 0.9%	48 1.5%	3,058 94.9%
2005					
Non-IB Students	14,068 98.7%	80 0.6%	30 0.2%	20 0.1%	49 0.3%
IB Students	70 2.0%	28 0.8%	28 0.8%	67 1.9%	3,314 94.5%
2006					
Non-IB Students	14,725 98.9%	82 0.6%	32 0.2%	10 0.1%	39 0.3%
IB Students	60 1.6%	37 1.0%	27 0.7%	65 1.7%	3,565 95.0%
2007					
Non-IB Students	15,376 98.5%	106 0.7%	43 0.3%	23 0.1%	65 0.4%
IB Students	78 2.0%	58 1.5%	55 1.4%	85 2.1%	3,686 93.0%
TOTAL					
Non-IB Students	84,849 98.7%	582 0.7%	205 0.2%	98 0.1%	274 0.3%
IB Students	431 2.1%	268 1.3%	231 1.1%	381 1.9%	19,062 93.6%

Reducing Selection Bias Through Propensity Score Stratification and Full Matching

Propensity score stratification or matching is often used in regression models as a mechanism for reducing selection bias (Rosenbaum, 2010). The notion is that by blocking on the strata or matching on the propensity score, we are holding constant the likelihood of participating in IB given that students in the same strata or matched group have similar propensity scores. This approach should reduce or eliminate the selection bias inherent in the unadjusted relationships between IB participation and student and school characteristics. Tables 7 through 10 show the bivariate odds ratios for each predictor before and after propensity score stratification and full matching. The tables also show the percent reduction in selection bias, calculated as the relative change in the logistic regression slope coefficient (i.e., $[(\beta - \beta_{adj})/\beta] \times 100\%$). Tables 7 through 10 do not include pair matching results because the effective bias reduction from pair matching must be evaluated simultaneously with comparisons of matched and unmatched students (see the next section for those results).

Table 7 shows that using propensity score stratification and propensity score full matching dramatically reduces the selection bias associated with student demographic predictors. What had been highly significant odds ratios showing major differences for IB participation based on gender, race, nationality, language, poverty, and disability/ability are now non-significant under both stratification and full matching. The relative reduction in selection bias is at least 86% and well over 90% for most variables. Although some variables show bias reductions greater than 100%, these should not be taken to suggest a reversal of the bias, as the adjusted relationships are not significantly different from even odds of 1.0.

TABLE 7.
Bivariate Odds Ratios for Student Demographic Predictors of Participation
in the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	ODDS RATIOS FOR IB vs. NON-IB STUDENTS			PERCENT REDUCTION IN SELECTION BIAS
	UNADJUSTED	PROPENSITY STRATIFICATION ADJUSTED	PROPENSITY MATCHING ADJUSTED	
Male	0.81***	0.99	0.98	94%, 89%
Race/Ethnicity (Caucasian Reference)				
Asian	3.09***	1.17	1.13	86%, 89%
African American	0.27***	0.99	1.00	99%, 100%
Hispanic/Latino/Latina	0.59***	1.05	1.05	108%, 110%
Native American	0.68~	0.97	1.15	n/s, n/s
Multiracial	0.51**	1.16	1.53	121%, 162%
US Residency Status				
Nonresident Alien	2.13**	0.92	1.23	111%, n/s
US Citizen	1.41***	0.88	0.89	138%, 134%
Born outside the US	1.00	1.12	1.12	n/s, n/s
Family Language				
English	1.25***	0.87	0.90	163%, 150%
Parent speaks English	1.19***	0.92	0.91	146%, 155%
School Program Participation				
Limited English Proficiency	0.14***	0.91	0.97	95%, 98%
Special Education Student	0.42***	0.99	0.96	98%, 95%
Free/Reduced Lunch Eligible	0.30***	0.98	0.99	98%, 99%
Gifted Student	6.97***	1.05	1.06	98%, 97%

Note ~p<.10, *p<.05, **p<.01, ***p<.001;
n/s denotes non-significant change in odds ratios (i.e. p>.10)

Odds ratios in this table are based on bivariate multilevel models (students within schools).

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 8 shows that the selection bias reduction for student academic indicators is not as complete. Although the bias associated with attendance and grade retention are reduced to non-significant levels after stratification and matching, all of the GPA and prior test score predictors remain statistically significant after stratification, and most remain significant after matching, despite dramatic reductions in selection bias. The bias associated with GPAs in 9th and 10th grades was reduced by 82% to 93%, while the bias associated with prior test scores was reduced by 91% to 97%. That said, the unadjusted bias for GPA and prior test scores was enormous, reflecting a 285% to 749% increase in the odds of participating in IB for each standard deviation increase in GPA or FCAT scores. After adjustment, these increases in odds of participation in IB are shrunken to between 15% and 22% after stratification, and to no greater than 16% after matching.

TABLE 8.
Bivariate Odds Ratios for Student Performance Indicators as Predictors of Participation in the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	ODDS RATIOS FOR IB vs. NON-IB STUDENTS			PERCENT REDUCTION IN SELECTION BIAS
	UNADJUSTED	PROPENSITY STRATIFICATION ADJUSTED	PROPENSITY MATCHING ADJUSTED	
Average Attendance Rate ^a	1.95***	1.01	1.02	98%, 97%
Retained in Grade at Least Once	0.13***	0.87	0.94	93%, 97%
Prior Grade Point Average^a				
Unweighted 9th Grade GPA	3.52***	1.17***	1.11**	88%, 92%
Unweighted 10th Grade GPA	2.85***	1.21***	1.15***	82%, 87%
Weighted 9th Grade GPA	5.18***	1.18***	1.11**	90%, 93%
Weighted 10th Grade GPA	4.09***	1.22***	1.16***	86%, 90%
Prior FCAT State Test Scores^a				
Mean FCAT Math Score in Grades 3-8	7.42***	1.15*	1.06	93%, 97%
Mean FCAT Reading Score in Grades 3-8	5.37***	1.17**	1.07	91%, 96%
Mean FCAT Math Score in Grades 9-10	7.49***	1.18***	1.10*	92%, 95%
Mean FCAT Reading Score in Grades 9-10	6.17***	1.18***	1.10*	91%, 95%

Note ~p<.10, *p<.05, **p<.01, ***p<.001;
n/s denotes non-significant change in odds ratios (i.e. p>.10)

Odds ratios in this table are based on bivariate multilevel models (students within schools).

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 9 shows that selection bias associated with course-taking patterns is also mostly reduced to non-significant levels, with relative bias reductions of at least 85%. The indicator of taking Trigonometry/Pre-calculus by 10th grade remained slightly more prevalent among IB students (i.e., by 25% to 38%) after stratification or matching. The number of advanced credits in 10th grade also maintained a small positive bias after matching (i.e., a 1 SD increase was associated with a 16% increase in the odds of IB participation). Under the stratification adjustment, both advanced credits variables showed bias reductions greater than 100%, with statistical significance for both the adjusted and unadjusted odds ratios. This finding implies a change in the direction of the relationship and a possible over-adjustment of this particular bias after controlling for the propensity scores and strata, IB students are less likely to have as many advanced credits as non-IB students. Then again, the adjusted odds ratios for these two variables, although statistically significant, are just barely significant and very close to even odds (i.e., 1.0). Given that the matching adjustment did not produce the same reversal of sign, this finding is likely reflective of an over-adjustment due to misspecification in the simpler stratification model.

TABLE 9.
Bivariate Odds Ratios for Early High School Course-Taking Indicators as Predictors of Participation in the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	ODDS RATIOS FOR IB vs. NON-IB STUDENTS			PERCENT REDUCTION IN SELECTION BIAS
	UNADJUSTED	PROPENSITY STRATIFICATION ADJUSTED	PROPENSITY MATCHING ADJUSTED	
Highest Math Through 10th Grade (reference: Algebra II)				
Basic Math	0.03***	0.88	0.91	96%, 97%
Algebra I	0.01***	1.05	1.07	101%, 101%
Geometry	0.03***	1.04	1.09	101%, 102%
Trigonometry/ Pre-calculus	8.20***	1.25~	1.38**	89%, 85%
Calculus or Above	2.34***	1.53	1.40	n/s, 61%
Late Algebra I (after 9th Grade)				
	0.04***	0.91	0.92	97%, 97%
Early Algebra I (before 9th Grade)				
	23.15***	1.10	1.08	97%, 97%
Advanced Credits in 9th Grade)^a				
	17.78***	0.89*	1.05	104%, 98%
Advanced Credits in 10th Grade)^a				
	43.29***	0.91~	1.16**	102%, 96%

Note ~p<.10, *p<.05, **p<.01, ***p<.001;
n/s denotes non-significant change in odds ratios (i.e. p>.10)

Odds ratios in this table are based on bivariate multilevel models (students within schools).

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 10 shows that most of the school-level predictors that exhibited selection bias before adjustment experience dramatic reductions in bias after adjustment. Factors such as rural location, pupil/teacher ratio, percent free/reduced lunch, and percents of Asian and Hispanic students show near complete reductions in their selection bias after stratification. Magnet school status and mean FCAT scores in 9th and 10th grades remain significant predictors after stratification adjustment despite bias reductions between 83% and 91%. Under stratification, there is no significant reduction in bias associated with school type (regular vs. alternative or special-ed) or percent African American. Under the matching adjustment, the only variables whose bias was not completely removed were percent African American and percent White; the analyses also show a slight over-adjustment for percent Free/Reduced Lunch.

TABLE 10.
Bivariate Odds Ratios for School-Level Predictors of Participation
in the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	ODDS RATIOS FOR IB vs. NON-IB STUDENTS			PERCENT REDUCTION IN SELECTION BIAS
	UNADJUSTED	PROPENSITY STRATIFICATION ADJUSTED	PROPENSITY MATCHING ADJUSTED	
Regular School (vs. Alternative or Special Ed)	0.58 [~]	0.62 [~]	0.95	n/s, 90%
Magnet School	4.67 ^{***}	1.29 [*]	1.04	83%, 97%
Charter School	0.48	0.87	0.99	n/s, 99%
New School	1.32	1.81	1.14	n/s, n/s
Urban	1.01	0.90	0.93	n/s, n/s
Rural	0.37 ^{***}	0.85	0.88	83%, 87%
Title I School	0.73	1.02	1.09	107%, 128%
School-Wide Title I	0.70	0.99	1.07	98%, 118%
Pupil/Teacher Ratio ^a	0.81 ^{**}	0.96	0.98	80%, 89%
Percent Free/Reduced Lunch ^a	0.76 ^{***}	1.04	1.07 [*]	115%, 126%
Percent Asian ^a	4.30 ^{***}	1.03	0.99	98%, 101%
Percent Hispanic/Latino/Latina ^a	0.72 ^{***}	0.98	1.02	94%, 107%
Percent African American ^a	1.13 [~]	1.06 [~]	1.06 [~]	n/s, n/s
Percent White ^a	1.00	0.96	0.94 [~]	n/s, n/s
School Size ^a	1.05	1.01	0.98	n/s, n/s
School Mean FCAT Math Scores in Grades 9-10	13.88 ^{***}	1.27 ^{***}	1.02	91%, 99%
School Mean FCAT Reading Scores in Grades 9-10	8.41 ^{***}	1.23 ^{***}	1.01	90%, 100%

Note [~]p<.10, ^{*}p<.05, ^{**}p<.01, ^{***}p<.001;
n/s denotes non-significant change in odds ratios (i.e. p>.10)

Odds ratios in this table are based on bivariate multilevel models (students within schools).

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Reducing (or Increasing) Selection Bias through Propensity Pair Matching

Paired propensity score matching is another means of reducing selection bias using propensity scores (Rosenbaum, 2010). Unlike full matching, which stratifies the full sample, pair matching creates strata with only two subjects—one from each group. Here, pair matching links IB students to only one control student with a similar propensity score, and unmatched IB students are dropped from subsequent analyses. Thus, it is important to gauge not only how similar are the matched IB and non-IB students, but also to compare the characteristics of the matched IB students to unmatched IB students. If those IB students who were successfully matched are not representative of the larger population of IB students, then we may have reduced selection bias for only a subset of the original sample. In that case, any findings would not generalize to the larger population. In other words, we would have estimates of the impacts of IB for a sample of students who don't really look like IB students.

Table 11 presents results for differences in student demographics after pair matching. Whereas there are no significant differences between matched IB and non-IB students, the matched IB students are very different from the non-matched IB students. Asian IB students were 45% less likely to be matched. African American and Hispanic IB students were, respectively, 91% and 46% more likely to be matched. IB students who were US citizens were 21% less likely to be matched. IB students whose primary family language was English were 15% less likely to be matched. IB students who had been identified at some point from 3rd through 10th grade as English language learners were 2.5 times (254%) more likely to be matched. IB students who had been selected for Special Education services at some point from 3rd through 10th grade were 1.4 times (139%) more likely to be matched. IB students who had received free or reduced-price lunch at some point from 3rd through 10th grade were 1.8 times (175%) more likely to be matched. Lastly, IB students who had been selected for a gifted/talented program at some point from 3rd through 10th grade were 55% less likely to be matched.

TABLE 11.
Student Demographics for Matched and Unmatched Students from the
International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	IB STUDENTS		MATCHED NON-IB STUDENTS (C)	ODDS RATIOS	
	UNMATCHED (A)	MATCHED (B)		A vs. B	B vs. C
Male	42.3%	43.5%	44.2%	1.08	0.97
Race/Ethnicity (Caucasian Reference)					
Asian	13.8%	6.6%	6.3%	0.55***	1.06
African American	9.9%	16.0%	16.3%	1.91***	1.00
Hispanic/Latino/Latina	13.3%	18.6%	17.0%	1.46***	1.12
Native American	0.4%	0.4%	0.5%	1.49	0.87
Multiracial	0.2%	0.1%	0.4%	0.44	0.41
US Residency Status					
Nonresident Alien	0.3%	0.3%	0.0%	n/a	n/a
US Citizen	91.5%	88.25%	89.3%	0.79*	0.89
Born outside the US	13.5%	15.4%	14.9%	1.06	1.04
Family Language					
English	85.6%	83.0%	83.7%	0.85~	0.95
Parent speaks English	83.3%	81.3%	82.0%	0.89	0.95
School Program Participation					
Limited English Proficiency	1.3%	4.2%	4.2%	2.54***	0.98
Special Education Student	5.5%	8.7%	9.3%	1.39**	0.93
Free/Reduced Lunch Eligible	22.1%	34.2%	34.2%	1.75***	1.00
Gifted Student	47.8%	26.7%	25.7%	0.45***	1.06

Note ~p<.10, *p<.05, **p<.01, ***p<.001;
n/s denotes non-significant change in odds ratios (i.e. p>.10)

Odds ratios in this table are based on bivariate multilevel models (students within schools).

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 12 presents results for differences in students' prior performance indicators after pair matching. The matched IB and non-IB students differed only in terms of 9th grade GPA. There was difference of .33 grade points favoring IB students on both weighted and unweighted GPA, with odds ratios showing that a one standard deviation difference in GPA increased the odds of enrolling in IB by about 20 percent.

In contrast, all ten indicators in this table show large differences between matched IB and non-matched IB students in prior performance. A one standard deviation increase in attendance rate was associated with a 27 percent reduction in the odds of being matched. Being retained in a grade at least once before the 10th grade was associated with a 3.6 times (361%) greater odds of being matched. Higher GPA in 9th and 10th grades was associated with substantial reductions in the odds of being matched, with the largest effect for weighted 9th grade GPA—a one standard deviation increase in GPA was associated with a 55 percent reduction in the odds of being matched. Lastly, higher mean FCAT scores were associated with substantial reductions in the odds of being matched, with the largest effect for mean FCAT math score across Grades 3 through 8—a one standard deviation increase math FCAT scores in was associated with a 67 percent reduction in the odds of being matched.

TABLE 12.
Student Performance Indicators for Matched and Unmatched Students
from the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	IB STUDENTS		MATCHED NON-IB STUDENTS (C)	ODDS RATIOS	
	UNMATCHED (A)	MATCHED (B)		A vs. B	B vs. C
Average Attendance Rate ^a	97%	96%	96%	0.73***	1.00
Retained in Grade at Least Once	2.2%	6.5%	7.1%	3.61***	0.91
Prior Grade Point Average^a					
Unweighted 9th Grade GPA	3.46	3.24	3.13	0.58***	1.22***
Unweighted 10th Grade GPA	3.38	3.18	3.15	0.66***	1.06
Weighted 9th Grade GPA	3.74	3.41	3.32	0.45***	1.20***
Weighted 10th Grade GPA	3.66	3.37	3.34	0.53***	1.06
Prior FCAT State Test Scores^a					
Mean FCAT Math Score in Grades 3-8	379	351	350	0.33***	1.04
Mean FCAT Reading Score in Grades 3-8	373	345	343	0.42***	1.05
Mean FCAT Math Score in Grades 9-10	376	355	354	0.35***	1.05
Mean FCAT Reading Score in Grades 9-10	373	348	348	0.41***	1.01

Note ~p<.10, *p<.05, **p<.01, ***p<.001;

Odds ratios in this table are based on bivariate multilevel models (students within schools).

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 13 presents results for differences in students' prior courses after pair matching. The matched IB and non-IB students differed only in terms of advanced credits taken in 9th and 10th grades. Matched non-IB students had .5 more advanced credits in 9th grade than matched IB students, with an odds ratio showing that a one standard deviation difference in advanced credits was associated with a 20 percent decrease the odds of enrolling in IB. On the other hand, matched IB students had .4 more advanced credits in 10th grade than matched non-IB students, with an odds ratio showing that a one standard deviation difference in advanced credits was associated with a 19 percent increase the odds of enrolling in IB.

TABLE 13.
Early High School Course-Taking Indicators for Matched and Unmatched Students
from the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	IB STUDENTS		MATCHED NON-IB STUDENTS (C)	ODDS RATIOS	
	UNMATCHED (A)	MATCHED (B)		A vs. B	B vs. C
Highest Math Through 10th Grade (reference: Algebra II)					
Basic Math	0.1%	1.2%	1.0%	7.83 ***	1.34
Algebra I	0.1%	9.0%	8.2%	67.56 ***	1.19
Geometry	2.7%	26.3%	25.0%	15.79 ***	1.14
Trigonometry/Pre-calculus	36.0%	12.1%	10.8%	0.34 ***	1.22
Calculus or Above	2.2%	1.1%	0.7%	0.54 *	1.63
Late Algebra I (after 9th Grade)	13.7%	42.6%	41.9%	4.84 ***	1.03
Early Algebra I (before 9th Grade)	86.3%	57.4%	58.0%	0.21 ***	0.97
Advanced Credits in 9th Grade^a	5.0	1.0	1.5	0.12 ***	0.80 ***
Advanced Credits in 10th Grade^a	4.6	1.4	1.1	0.06 ***	1.19 ***

Note ~p<.10, *p<.05, **p<.01, ***p<.001;

Odds ratios in this table are based on bivariate multilevel models (students within schools).

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Table 14 presents results for differences in school characteristics after pair matching. The matched IB and non-IB students differed only in terms of the prevalence of Asian students in their schools, and school mean FCAT reading and math scores in 9th and 10th grades. Matched non-IB students had a slightly higher proportion of Asian students in their schools, with an odds ratio showing that a one standard deviation increase in percent Asian was associated with a seven percent decrease the odds of enrolling in IB. Matched non-IB students also had slightly higher school-mean FCAT reading and math scores, with an odds ratio showing that a one standard deviation increase in school-mean FCAT scores in 9th and 10th grades was associated with a 12 percent lower odds of enrolling in IB.

Much larger differences were observed in comparisons of matched and unmatched IB students. Compared to other IB students, matched IB students had an 83 percent lower odds of attending a regular school (as opposed to alternative or special education schools). Matched IB students had an 81 percent lower odds of attending a magnet school and an 11.9 times (1186%) higher odds of attending a charter school. Matched IB students were 4 times more likely to attend a rural school and were 53 percent less likely to attend a school that was eligible for school-wide Title I assistance. Matched IB students had lower proportions of Asian and African American students in their schools, with odds ratios showing that a one standard deviation increase in percent Asian was associated with a 50 percent decrease the odds of being matched, and that a one standard deviation increase in percent African American was associated with a 29 percent decrease the odds of being matched. Correspondingly, matched IB students had higher proportions of White students in their schools, with odds ratios showing that a one standard deviation increase in percent White was associated with a 47 percent increase the odds of being matched. School size for matched and non-matched IB students was similar when averaged across students; however, the odds ratios from the multilevel model with school random effects showed that a one standard deviation increase in school size was associated with a 26 percent reduction in the odds of being matched. Lastly, school mean FCAT math and reading scores in grades 9 and 10 were considerably lower for matched IB students. A one standard deviation increase in school mean FCAT math and reading scores was associated, respectively, with a 72 and 76 percent reduction in the odds of being matched.

TABLE 14.
School-Level Predictors for Matched and Unmatched Students
from the International Baccalaureate (IB) Diploma Programme

PREDICTOR VARIABLE	IB STUDENTS		MATCHED NON-IB STUDENTS (C)	ODDS RATIOS	
	UNMATCHED (A)	MATCHED (B)		A vs. B	B vs. C
Regular School (vs. Alternative or Special Ed)	99.9%	98.4%	98.9%	0.17 **	0.68
Magnet School	60.4%	46.2%	47.7%	0.19 ***	0.94
Charter School	0.0%	0.7%	0.6%	11.86 **	1.13
New School	0.0%	0.3%	0.2%	5.58	1.34
Urban	34.3%	30.0%	32.2%	1.44	0.89
Rural	5.4%	11.7%	11.8%	4.00 ***	0.95
Title I School	12.7%	13.3%	13.6%	0.74	0.97
School-Wide Title I	10.4%	9.3%	9.2%	0.47 *	1.02
Pupil/Teacher Ratio ^a	19.4	19.6	19.6	1.11	1.02
Percent Free/Reduced Lunch ^a	26.9%	26.1%	25.5%	1.21	1.04
Percent Asian ^a	5.1%	4.3%	4.6%	0.50 ***	0.93 *
Percent Hispanic/Latino/Latina ^a	15.7%	17.1%	16.6%	1.04	1.03
Percent African American ^a	27.1%	22.9%	23.2%	0.71 ***	0.99
Percent White ^a	51.8%	55.4%	55.4%	1.47 ***	1.00
School Size ^a	2247	2259	2254	0.74 **	1.01
School Mean FCAT Math Scores in Grades 9-10	362	353	355	0.24 ***	0.88 *
School Mean FCAT Reading Scores in Grades 9-10	358	347	349	0.28 ***	0.88 **

Note ~p<.10, *p<.05, **p<.01, ***p<.001;

Odds ratios in this table are based on bivariate multilevel models (students within schools).

^a Odds ratios for continuous variables represent difference in odds associated with a one standard deviation increase in the predictor.

Comparing Postsecondary Indicators for IB and non-IB Students by Matched Status

Our final analyses compare postsecondary indicators related to access to and performance in college for IB and non-IB students broken out by whether they were matched or unmatched. First, we compare SAT and ACT scores among these four groups of students (i.e., unmatched non-IB, matched non-IB, unmatched IB, and matched IB). Next, we present college enrollment rates for each of the four groups. Finally, we use multilevel linear and logistic regression to compare these outcomes for IB and non-IB students with and without propensity score adjustments.

Table 15 shows mean SAT and ACT scores across these four groups along with missing data rates. The missing data rates in Table 15 are important because each of these scores is observed only if the student chooses to take the SAT or ACT test. For example, data are missing for the majority of unmatched non-IB students in the study sample likely because they did not take the SAT or ACT tests. Therefore, comparisons of the missing data rates provide information about differences in the percentages of students who chose to take these college entrance tests. The general trend in the average test scores shows that unmatched non-IB students have the lowest scores, matched non-IB students have substantially higher scores, matched IB students have still higher scores, and unmatched IB students have the highest scores by far. The missing data rates for these test scores show a similar but opposite trend—unmatched non-IB students have the highest missing data rates, matched non-IB students have substantially lower rates of missing data, matched IB students have even lower rates of missing data, and unmatched IB students have the lowest rates of missing data by far.

TABLE 15.
SAT and ACT Test Score Averages and Missing Data Rates for International Baccalaureate (IB) Diploma Programme Participants and Non-Participants

POSTSECONDARY INDICATOR	NON-IB STUDENTS		IB STUDENTS	
	UNMATCHED	MATCHED	MATCHED	UNMATCHED
SAT Math Score	505.8 (49%)	561.9 (22%)	575.3 (19%)	628.5 (4%)
SAT Verbal Score	504.1 (49%)	557.7 (22%)	579.9 (19%)	626.5 (4%)
SAT Writing Score	480.0 (90%)	524.3 (85%)	545.3 (83%)	608.6 (81%)
ACT Math Score	21.2 (68%)	23.5 (50%)	23.9 (50%)	26.4 (46%)
ACT Reading Score	22.1 (68%)	24.2 (50%)	25.3 (50%)	27.6 (46%)
ACT English Score	20.7 (68%)	23.1 (50%)	24.0 (50%)	26.4 (46%)

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 16 shows postsecondary enrollment rates across the four groups of students. While 86% of unmatched IB students enrolled in postsecondary studies in the summer or fall immediately following their high school graduation, a slightly lower percentage of matched IB students (84%) and matched non-IB students (83%) did so. A substantially lower percentage of unmatched non-IB students (76%) enrolled in postsecondary studies immediately following high school graduation. These students also had a low rate of enrollment in 4-year institutions (55%) and a very low rate of enrollment in selective institutions³ (19%). Matched non-IB and matched IB students were quite similar in their 4-year institution enrollment rates, with 73 versus 70 percent enrollment, respectively. Unmatched IB students had a 4-year institution enrollment rate of 78 percent. Matched non-IB, matched IB, and unmatched IB students were quite similar in their rates of enrollment in selective institutions: 36, 34, and 34 percent, respectively.

³ Identified by Barrons in 2008 as “very competitive,” “highly competitive,” or “most competitive.”

TABLE 16.
College Enrollment Rates for International Baccalaureate (IB)
Diploma Programme Participants and Non-Participants

POSTSECONDARY INDICATOR	NON-IB STUDENTS		IB STUDENTS	
	UNMATCHED	MATCHED	MATCHED	UNMATCHED
Immediate College Enrollment	75.7%	83.4%	84.1%	86.0%
Enrollment in a 4-Year Institution	55.0%	72.6%	69.5%	78.0%
Enrollment in a Selective Institution	18.8%	36.4%	33.9%	34.0%

Note. Missing data rates for enrollment indicators are unknown given that non-enrollment is observed as missing data.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Table 17 shows results from multilevel linear and logistic regression models comparing outcomes for IB and non-IB students with and without propensity score adjustments. Very large differences in SAT scores were observed for all three sections of the test; math, verbal, and writing scores were between 119 and 126 points higher for IB students. After propensity score adjustments, the advantage in SAT scores for IB students shrank substantially with the largest adjustments occurring under propensity stratification (with the continuous propensity score estimate as an additional covariate) and the smallest adjustments occurring under full matching. A similar pattern was found for ACT scores, with differences in Math ACT scores between IB and non-IB students becoming insignificant under propensity stratification and pair matching.

Large differences were also observed with regards to postsecondary enrollment. IB students were almost 2 times more likely to enter college immediately after high school, they were 2.6 times more likely to enroll in a 4-year institution, and they were over 2 times more likely to enroll in a selective college (see footnote 3). After propensity stratification or full matching, these differences were completely absent, with insignificant odds ratios near unity. After propensity pair matching, IB students were only 9% more likely to enroll in college immediately after high school, while the difference for 4-year institution enrollment rates actually reversed, with IB students predicted to be 17% less likely to enroll in a 4-year institution.

Taken as a whole, these results suggest that propensity score techniques reduce selection bias when comparing IB and non-IB students; however, the adjusted differences observed between IB and comparison students should not be interpreted as causal impacts of IB given the problems associated with extrapolation and the inability to match all IB students to similar non-IB students.

TABLE 17.
Differences in Postsecondary Indicators for International Baccalaureate (IB)
Diploma Programme Participants and Non-Participants
with and without Propensity Score Adjustments

POSTSECONDARY INDICATOR	UNADJUSTED	PROPENSITY STRATIFICATION	PROPENSITY FULL-MATCHING	PROPENSITY PAIR-MATCHING
Continuous Outcomes (Mean Differences)				
SAT Math Score	120.90 ***	14.03 ***	29.00 ***	15.22 ***
SAT Verbal Score	119.10 ***	21.39 ***	35.30 ***	25.12 ***
SAT Writing Score	126.30 ***	20.06 **	30.25 ***	26.68 **
ACT Math Score	5.28 ***	0.35 ~	1.00 ***	0.38
ACT Reading Score	5.41 ***	0.87 ***	1.52 ***	1.15 ***
ACT English Score	5.62 ***	0.71 ***	1.36 ***	1.02 ***
Categorical Outcomes (Odds Ratios)				
Immediate College Enrollment	1.94 ***	1.02	1.04	1.09 ***
Enrollment in a 4-Year Institution	2.57 ***	0.95	1.06	0.83 ***
Enrollment in a Selective Institution	2.15 ***	0.95	1.00	0.89

Note ~p<.10, *p<.05, **p<.01, ***p<.001;

Adjusted mean differences and odds ratios in this table are based on bivariate multilevel models (students within schools).

CONCLUSIONS AND IMPLICATIONS



There is tremendous interest in the potential impacts of credit-based transition programs like International Baccalaureate, but any attempts to examine those impacts must deal with selection bias that results from the voluntary participation of schools and students. Failure to do so makes it impossible to determine whether the performance of participating students was actually influenced by the program, or whether the outcomes for these students would have been just as good without the program. This study revealed that, when looking at the statewide population in Florida, the selection bias associated with voluntary participation in IB is very large, and that mechanisms for dealing with selection bias using propensity scores may not be sufficient. In other words, comparing IB and non-IB students in this statewide context is like comparing apples and oranges, and using propensity score methods to adjust for these differences require strong assumptions and extrapolation into regions with very thin data.

Our results show that IB students in Florida differ from other students in terms of individual demographics, academic performance, course-taking, and the characteristics of the schools they attend. Although predictive of IB participation, individual demographic variables were not the strongest predictors. IB students were only slightly more likely to be female, 3 times more likely to be Asian versus White, and 2 to 3 times more likely to be White versus Latino or African American. IB students were also less likely to be English language learners, have a disability, or be eligible for free/reduced lunch. The strongest predictor of IB participation among the individual demographic variables was gifted/talented status, with IB students more than 6 times as likely to be gifted compared to non-IB students.

Individual student academic performance and course-taking indicators were by far the strongest predictors of IB participation. A one-standard deviation increase in GPA or prior test scores in math and reading translated to between a three-fold and eight-fold increase in the odds of participating in IB. Even stronger was the prediction of course-taking patterns in 9th and 10th grades. Students who took Algebra I early (i.e., before 9th grade) were 23 times more likely to participate in IB, while students who took Algebra I late (i.e., after 9th grade) were 25 times less likely to participate. Students who took more advanced courses (i.e., honors, AP) in 9th and 10th grades were 18 to 43 times more likely to participate in IB. Clearly, IB students are much more likely to have exceptional academic records, and their individual academic performance is much more predictive of participation in IB than their gender, race, or family background.

A number of school-level variables were predictive of IB participation, but these relationships were generally much weaker than student-level factors. The strongest school-level predictors showed that students attending schools with high test scores were 8 to 12 times more likely to participate in IB, students attending magnet schools were over four times more likely to participate in IB, and students in rural schools were nearly three times less likely to participate in IB. Racial composition of schools was also related to IB participation, with IB more prevalent in schools with larger Asian and African American populations, and smaller Hispanic/Latino populations. The slightly increased prevalence of IB in schools serving African American students may be confounded with the popularity of IB as a magnet program, especially given that magnet programs in Florida were intended to improve racial balance in schools (Chen, 2007), and that African American students are less likely to participate in IB despite the greater likelihood of IB program availability in the schools many of them attend. The converse was true for Asian students in that IB programs are more prevalent in schools that serve larger populations of Asian students, and Asian students are also much more likely to enroll in IB. The reasons behind Asian students' preference for IB and the increased prevalence of IB in schools that serve more Asian students is a potential topic for future research.

The first major conclusion from these results is that while school and student demographics are related to IB participation, the best predictors are individual academic performance indicators. This conclusion aligns quite well with the design of IB as a highly rigorous college preparatory curriculum, one that tends to attract the best and the brightest high school students. IB has a reputation as an elite academic program, and that certainly rings true in these results. But, the most commonly available indicators of high school students' academic performance such as GPA and test scores tell only part of the story. Far better prediction of IB participation can be made using information on students' course-taking patterns in early high school—IB students tend to take challenging courses well before they enroll in IB, suggesting that the selection process starts much earlier than enrollment in IB at the start of 11th grade.

The second major conclusion from this work is that a comprehensive logic model of the selection mechanism is essential for any observational study. The myriad factors found to predict IB participation highlight the importance of a logic model based on a comprehensive literature review and conceptual framework whenever statistical analyses are used to model a selection process. From our previous research about

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

schools' adoption of IB, their recruitment of students into the program, and the characteristics of students who enroll in IB (see Perna et al., in press), we identified dozens of indicators from the Florida K-20 Education Data Warehouse that revealed dramatic differentiation between IB and non-IB students. In addition, we found that the strongest predictors of participation were not the indicators most commonly used to address selection bias in prior research on IB (i.e., student demographics and test scores). Instead, our analyses show that the strongest predictors of IB participation were indicators of academic challenge and success in prior grades; specifically, enrolling in advanced courses during 8th, 9th and 10th grades. This conclusion suggests that future research on IB and other credit-based transition programs should dig deeper into administrative data and include indicators derived from middle school and high school transcripts.

Our logic model also identifies predictors of participation that are not available in our dataset and thus not included in our analyses, such as measures of student motivation and family influences. Obviously, obtaining relevant data on these factors is complicated. However, the predictive power of these factors above and beyond that captured by the typical academic indicators may be substantial. Therefore, any study that uses statistical methods to adjust for overt selection bias (e.g., propensity scores), but does not include measures of student motivation or family influence in its models, may leave a substantial bias uncontrolled. Even sensitivity analyses may not assuage concerns if the strength of the relationship between student motivation or family influence and IB participation is as high as that for indicators of course-taking patterns.

The third major conclusion from these results is that most IB students in Florida are very, very different from non-IB students. The differences are evident in the very large odds ratios for many of the academic indicators predicting IB participation. Our analyses suggest that IB students, prior to their enrollment in IB, are unlikely to participate in programs for at-risk students (i.e., English language learners, Special Education students, or economically-disadvantaged students), tend to have excellent grades, and take accelerated courses even before they reach high school. The accumulation of these many differences between IB and non-IB students becomes very evident in the distributions of propensity scores for the two groups. There is very little overlap in the propensity scores for IB and non-IB students. Even though the propensity score adjustments seem to reduce the selection bias, the lack of overlap in the propensity scores suggests that our models are reliant on blocked comparisons in which thousands of IB students are compared to only a couple

hundred non-IB students (i.e., at the higher end of the propensity score distribution) and thousands of non-IB students are compared to only a few hundred IB students (i.e., at the lower end of the propensity score distribution). Consequently, the stability of any models of impacts on student outcomes using these propensity scores will be quite poor, with the most critical regions of the model (e.g., outcomes for non-IB students who are similar to IB students, and vice versa) based on only a tiny fraction of the available sample. More importantly, as Donald Rubin (2004) articulates:

“If there is little or no overlap in the distributions of the estimated propensity scores in the treatment groups, there is no hope of drawing valid causal inferences from these data without making strong external assumptions involving model-based extrapolation, because the estimated propensities will all be essentially either 0 or 1. . . . sometimes a data set cannot support a decent causal inference” (p. 354).

It certainly seems that our study of the population of IB students across the entire state of Florida is one of those cases where decent causal inference is simply not possible.⁴

Our findings do not mean that all studies of IB or other credit-based transition programs that use propensity score methods are suspect. In fact, one potential explanation for why our propensity scores had so little overlap is that we used a statewide sample in our analyses. Since much of the difference in propensity scores in our study can be attributed to the exceptional academic records of IB students prior to the 11th grade, it may be easier to establish comparability of IB and non-IB students in contexts where access to and participation in IB is not limited to the

⁴ In addition to propensity score methods, we also considered several alternative methods for dealing with selection bias including differences-in-differences models, regression discontinuity, and instrumental variables. Unfortunately, the nature of IB participation during the study period ruled out each of these alternatives for the vast majority of IB schools and students. Differences-in-differences could not be used given that IB programs have a two-year initiation phase, requiring comparisons of small cohorts of students from only a handful of schools who graduated at least four years apart. Regression discontinuity could not be used given that every IB program had flexible admission criteria and no clear cutoff could be identified for most indicators in most schools. Instrumental variables could not be used given the absence of an instrument that was correlated with students' outcomes. As such, we conclude that any attempts to deal with selection bias through propensity score or other methods will, at best, work only for a relatively small and unrepresentative subset of our sample, thus undermining our ability to draw causal inferences about IB for this statewide population.

academic elite. Such is the case in recent studies of IB in Chicago (Coca et al., 2012; Roderick et al., 2009; Saavedra, 2011), where program participants are much more demographically and academically diverse than the majority of IB students in our Florida sample. That said, the comparability of propensity scores and the threat of model extrapolation should be assessed visually and computationally in any such study (i.e., after assessment of the inclusion of relevant predictors of participation), with special attention paid to the implications for both internal and external validity.

The notion of better comparability in more narrow contexts could also be taken to justify our comparisons of IB student's outcomes in Florida for those students with less extreme propensity scores. In essence, the problem of model extrapolation and sparse data might be avoided by using pair-matching or multiple-matching to restrict analyses in which outcomes are compared between IB and non-IB students to a subsample of students with similar propensity scores. Of course, in these pair matching analyses, only a fraction of the total IB sample were matched, thus limiting the generalizability of findings to the population of students who could actually be matched—the IB students whose prior academic indicators are not quite as exceptional. Not surprisingly, the IB students we were able to match looked quite different from the broader population of IB students. Nonetheless, the adjusted differences in student outcomes after pair matching were quite similar to those after propensity score stratification. This is actually not surprising given that the greatest precision in the propensity stratification model occurs within those strata where there are a large number of both IB and non-IB students. In other words, propensity stratification results may also largely ignore large portions of the IB and non-IB samples because there simply aren't enough students from both groups represented in those strata. As such, results from any of our analyses involving propensity score matching or stratification are unlikely to meet Rubin's appeal for "decent causal inference."

The fourth and final conclusion from these results has implications for improving access to IB and other credit-based transition programs. Simply put, the amazing accuracy with which participation in IB can be predicted suggests that students are set along a well-defined IB-like track well before they reach the 11th grade. IB has a reputation for high standards, exceptional rigor, and recruiting the most capable and motivated students. To some degree, our results simply confirm that IB has had great success in Florida recruiting the best and brightest students. Efforts to expand IB to a broader population of students may provide new opportunities to study program impacts where differences between IB and non-IB students are less extreme.

On the other hand, as we describe in earlier work (see Perna et al., in press), IB has embarked on a mission to increase access to its programs. One way to increase access is simply to relax enrollment criteria and lower requirements for those students who might participate. The problem is that doing so may change the very nature of the IB program. An alternative approach is to better prepare a broader population of students for enrollment in IB once the opportunity arises. The recent development of the IB Primary Years Program (PYP) and Middle Years Program (MYP) is intended to improve early preparation. Yet the same issues of selection bias exist, so it may again be difficult to isolate the causal impacts of the PYP and MYP when studied on a broad scale.

Future research may be able to identify specific contexts in which causal inference can be made. The most promising opportunities for this approach are situations where IB programs are over-enrolled and students must apply for admission through a lottery. Although rare, these situations do exist for the PYP, MYP and IB Diploma Programme. Even in these instances, however, the students who apply for admission to these programs may not look like the broader population of IB participants. So once again, we are forced to choose a balance between internal and external validity. We might be able to get the right answer—What is the impact of IB for *this* group of students?—but it might not be the answer to the right question—What is the impact of IB in general?

REFERENCES



- Adolescence, 174, 289-312 Allison, P.D. (2001). *Missing Data*. Thousand Oaks, CA: Sage Publications.
- Barbour, M., & Mulcahy, D. (2006). An inquiry into retention and achievement differences in campus based and web based AP Courses. *Rural Educator*, 27, 8-12.
- Chen, G. (2007, December 4). What is a magnet school? *Public School Review*. Retrieved October 31, 2011 from <http://www.publicschoolreview.com/articles/2>
- College Board. (2012). *The 8th annual AP report to the nation*. No. 12b-5036). New York: Author.
- College Board (YEAR). *The 7th annual AP report to the nation*. No. 100084202. New York: Retrieved from <http://apreport.collegeboard.org/>
- College Board (YEAR). *Achieving equity*. Retrieved January 9, 2012 from <http://professionals.collegeboard.com/k-12/assessment/ap/equity>
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B.*, 39, 1–38.
- Education Commission of the States. (2012). *Advanced placement: State provides financial incentives for AP courses*. Retrieved from <http://mb2.ecs.org/reports/Report.aspx?id=1125>
- Enders, Craig K. (2010). *Applied Missing Data Analysis* (1st ed.). New York: Guildford Press.
- Estacion, A., Cotner, B. A., D'Souza, S., Smith, C. A. S., & Borman, K. M. (2011). *Who enrolls in dual enrollment and other acceleration programs in florida high schools?* No. Issues & Answers Report, REL 2012–No. 119). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Foust, R. C., Hertberg-Davis, H., & Callahan, C. M. (2009). *Adolescence*, 174, 289-312. Students' perceptions of the non-academic advantages and disadvantages of participation in Advanced Placement courses and International Baccalaureate Programs.

-
- Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6): 721–741.
- Godfrey, K. (2009). *Advanced placement summer institute: A survey of Florida AP teachers*. No. RN-39). New York: College Board.
- Hughes, K. L., Karp, M. M., Fermin, B. J., & Bailey, T. R. (2005). *Pathways to college: Access and success*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.
- International Baccalaureate North America. (2008). *Examination review & data summary 2008: Profile of diploma programme test takers*. New York: International Baccalaureate Organization.
- International Baccalaureate Organization. (2010). *IB answers*. Retrieved January 9, 2012 from https://ibanswers.ibo.org/app/answers/detail/a_id/33
- Karp, M. M., Bailey, T. R., Hughes, K. L., & Fermin, B. J. (2004). *State dual enrollment policies: Addressing access and quality*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.
- Kisker, C. B. (2006). Integrating high school and the community college: Previous efforts and current possibilities. *Community College Review*, 34(1), 68-86.
- Klopfenstein, K. (2004). Advanced placement: Do minorities have an equal opportunity? *Economics of Education Review*, 23(2004), 115-131.
- Lerner, J. B., & Brand, B. (2006). *The college ladder: Linking secondary and postsecondary education for success for all students*. Washington, DC: American Youth Policy Forum.
- Lerner, J. B., & Brand, B. (2008). *Review of state policies supporting advanced placement, international baccalaureate, and dual credit programs*. Washington, DC: American Youth Policy Forum.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data, Second Edition*. New York: John Wiley & Sons.
- Office of Program Policy Analysis & Government Accountability (OPPAGA) (2009). *Modifying advanced placement program incentive funding could produce significant cost savings*. Tallahassee, FL: Author.

Apples and Oranges:

Comparing the
Backgrounds and
Academic Trajectories
of International
Baccalaureate (IB)
Students to a Matched
Comparison Group

Perna, L.W. (2005). The key to college access: A college preparatory curriculum. In W.G. Tierney, Z. B. Corwin, & J.E. Colyar (Eds.), *Preparing for College: Nine Elements of Effective Outreach* (pp. 113-134). Albany, NY: State University of New York Press.

Perna, L.W., May, H., Yee, A., Ransom, T., Rodriguez, A., & Fester, R. (2013). Unequal Access to Rigorous High School Curricula: An Exploration of the Opportunity to Benefit From the International Baccalaureate Diploma Programme (IBDP). *Educational Policy*.

Planty, M., Bozick, R., and Ingels, S.J. (2006). Academic Pathways, Preparation, and Performance - A Descriptive Overview of the Transcripts from the High School Graduating Class of 2003-04 (NCES 2007-316). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Roderick, M., Nagoaka, J., Coca, V. & Moeller, E. (2009). *From High School to the future: Making hard work pay off, the road to college for students in CPS's academically advanced programs*. Chicago, IL: Consortium on Chicago School Research, University of Chicago.

Rosenbaum, P., & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55.

Rosenbaum, P.R. (2010). *Design of observational studies*. New York, NY: Springer-Verlag.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Rubin, D.B. (2004). Teaching statistical inference for causal effects in experiments and observational studies. *Journal of Educational and Behavioral Statistics*, 29, 103-116.

Saavedra, A.R. (2011). *The academic impact of enrollment in international baccalaureate diploma programs*. WR-867-EDU. Pittsburgh, PA: RAND Education.

Scott, T.P., Tolson, H., & Lee, Y. (2010). Assessment of advanced placement participation and university academic success in the first semester: Controlling for selected high school academic abilities. *Journal of College Admission*, (208), 26-30.

Siskin, L. S., Weinstein, M., & Sperling, R. (2010). *TO BE IB: Creating support structures and services for title I high schools implementing the international baccalaureate programme*. NYU Wagner Research Paper No. 2011-11. New York: Institute for Education and Social Policy, New York University.

-
- Steiner, P.M. (2011). *Matching Strategies for Observational Data with Multilevel Structure*. Paper presented at the spring conference of the Society for Research on Educational Effectiveness (SREE), Washington D.C., March 3-5, 2011.
- Strange, M., Johnson, J., Showalter, D., & Klein, R. (2012). *Why rural matters 2011-12: The condition of rural education in the 50 states*. Washington, DC: The Rural School and Community Trust.
- U.S. Department of Education (2012). *Advanced placement test fee program*. Retrieved from <http://www2.ed.gov/programs/apfee/index.html>
- U.S. Department of Education, National Center for Education Statistics (2011). *Digest of education statistics table 162. public high schools that offered and students enrolled in dual credit, advanced placement, and international baccalaureate courses, by school characteristics: 2003*. Washington, DC: Author. Retrieved from http://nces.ed.gov/programs/digest/d10/tables/dt10_162.asp
- Waits, T., Setzer, J. C., & Lewis, L. (2005). *Dual credit and exam-based course in U.S. public high schools: 2002-03*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Western Interstate Commission for Higher Education (WICHE). (2006). *Accelerated learning options: Moving the needle on access and success*. Boulder, CO: Author.

University of Pennsylvania
Teachers College, Columbia University
Harvard University
Stanford University
University of Michigan
University of Wisconsin-Madison
Northwestern University

A COLLABORATIVE PUBLICATION BETWEEN



Center for Research in
Education & Social Policy