EFFECTS OF CAREER-TECHNICAL AND COLLEGE-PREPARATORY HIGH SCHOOL CURRICULA ON EDUCATIONAL ATTAINMENT

by

SINAN GEMICI

(Under the Direction of Jay W. Rojewski)

ABSTRACT

Roughly one-third of all secondary students in the U.S. leave school without a regular high school diploma, and the percentage of U.S. bachelor's degree holders has fallen far below that of other industrialized nations. Since educational attainment is an important determinant of labor productivity and technological progress, it is critical to ascertain the effects of different high school curricula on educational attainment outcomes. The purpose of this study was to examine the effects of career-technical education (CTE) and college-preparatory high school curricula on secondary and postsecondary educational attainment. Given recurring debates over the resource intensity of secondary CTE, educational attainment outcomes for individuals enrolled in CTE curricula were of particular interest.

Data from the National Longitudinal Survey of Youth 1997 (NLSY97; U.S. Bureau of Labor Statistics, 2009a) were analyzed using a rigorous causal-comparative research design that applied multiple imputation and propensity score matching to control for missing data and selection bias, respectively. CTE curricula had a statistically significant positive effect on regular high school completion when compared to general-track curricula. No CTE curriculum effects were found for GED or any level of postsecondary educational attainment. College-preparatory curricula had a statistically significant positive effect on four-year college degree completion when compared to general-track curricula. No college-preparatory curriculum effects were found for any other level of educational attainment. The positive effects of CTE curricula on high school diploma attainment should prompt policymakers to rethink the role of CTE in U.S. public education and consider it a strategic asset in boosting high school completion rates. The positive effects of college-preparatory curricula on four-year postsecondary attainment corroborate the notion that such programs of study are best-suited for academically-inclined students who manage to persist throughout high school. Given the positive impact of CTE and college-preparatory curricula at different educational attainment levels, future research should closely examine the causal effects of dual CTE/college-preparatory curriculum concentrations on student outcomes.

INDEX WORDS: high school curriculum effects, educational attainment, CTE, multiple imputation, propensity score matching

EFFECTS OF CAREER-TECHNICAL AND COLLEGE-PREPARATORY HIGH SCHOOL CURRICULA ON EDUCATIONAL ATTAINMENT

by

SINAN GEMICI

B.A., Monterey Institute of International Studies, 1998

M.B.A., Thunderbird School of Global Management, 1999

M.S., University of Idaho, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

EFFECTS OF CAREER-TECHNICAL AND COLLEGE-PREPARATORY HIGH SCHOOL CURRICULA ON EDUCATIONAL ATTAINMENT

by

SINAN GEMICI

Major Professor: Jay W. Rojewski

Committee: Roger B. Hill
Diane Brook Napier
John Mativo

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2010

DEDICATION

I dedicate this dissertation to my loving parents, Kemal and Berta.

And to Ateş, my true companion :)

ACKNOWLEDGEMENTS

First and foremost, I want to thank you, Dr. Jay W. Rojewski, for your unrelenting support and encouragement over the past four years. As my major professor, you have taught me everything I know about the research process and, more importantly, about being an independent scholar. The most precious gift you have given me is the gift of your friendship, which I will always remember. I could not have wished for a better mentor to guide me through this doctoral program.

Special thanks go to you, Dr. Roger B. Hill, for having been there when it really counted. You are a very decent man, and for that I appreciate you. I also express my sincere gratitude to you, Dr. Diane Brook Napier, for providing me the opportunity to work as instructor of record in the Social Foundations of Education program. Teaching my undergraduates was the most cherished and fulfilling experience I had as a doctoral student. Last but not least, I thank you, Dr. John Mativo, for your lightheartedness and constant encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Almost thirty years ago, the landmark report *A Nation at Risk* (National Commission on Excellence in Education, 1983) highlighted the relative decline in America's ability to compete with a rising tide of well-educated and highly-motivated workers abroad. The authors' key message was that "others are matching and surpassing our educational attainments" (para. 1). This warning was echoed by a series of subsequent high-profile publications that emerged throughout the 1990s and into the new millennium (see National Center on Education and the Economy, 1990, 2007; National Research Council, 2001; U.S. Department of Labor, 1991). Collectively, these accounts highlighted the critical need to raise levels of educational attainment and productivity in the American workforce.

International comparative statistics on educational attainment corroborate this need. Within the past decade, the U.S. has moved from being the world leader in postsecondary educational attainment to occupying tenth place in the age group of 25-to 34-year-olds (Douglass, 2006; Organization for Economic Cooperation and Development [OECD], 2007b). In 2005, well over half of the population in Australia and several European Union member countries received a bachelor's degree at the typical age of graduation, compared to 34 percent in the U.S. (Snyder, Dillow, & Hoffman, 2009). Of these American undergraduate degrees, only 17 percent were awarded in mathematics and science, while rates were about twice as high in Germany, Finland, and South Korea.

Meanwhile, China is rapidly raising workforce productivity through aggressive investment in education (Li, Whalley, Zhang, & Xiliang, 2008). These trends reflect the transformed nature of a global labor market in which all but the most menial jobs require at least some sort of postsecondary education (Rojewski, 2002).

Trend data on U.S. income levels covering the period from 1975 to 2006 clearly illustrate the returns on educational attainment. Measured in constant 2006 dollars, income for high school graduates rose by six percent during that period, whereas individuals with some college earned 10 percent more. By far the largest raises were exhibited by individuals holding bachelor's (23%) and graduate or professional degrees (31%; Swanson, 2009). Advanced postsecondary attainment is, however, predicated on a strong foundation at the secondary level. Major federal legislation, such as the No Child Left Behind Act (2001) and the Carl D. Perkins Acts (1990, 1998, 2006), has been implemented to improve the efficacy of secondary schooling and facilitate post-high school transition. Despite these efforts, severe shortcomings regarding the outcomes of education persist. The following account from a recent U.S. Department of Education (2008) study illustrates the full extent of the attainment crisis from a longitudinal perspective:

> Let's examine what happened to a typical group of 20 children born that year [in 1983] who started kindergarten together in 1988. Six of them would not have graduated on time in 2001. Of the 14 who would have graduated on time, 10 would start college, and just five of those 20 kindergartners would have a college degree by spring 2007. These college graduates can expect to earn $1 million

more over their lifetimes than their classmates who dropped out of high school. (p. 2)

The considerable degree of social and racial/ethnic stratification that is reflected in high school completion rates exacerbates the dilemma. Low-income students have a 10 times higher likelihood of dropping out than their high-income peers (Cataldi, Laird, & KewalRamani, 2009). Given the overrepresentation of minorities in low-income strata, African American, Hispanic, and Native American students continue to exhibit higher dropout rates than the remainder of the student population (KewalRamani, Gilbertson, Fox, & Provasnik, 2007; Swanson, 2009). Overall, the current status of educational attainment in the U.S. reflects an alarming waste of talent that undermines the global competitiveness of the American workforce.

One important area that is strongly associated with educational outcomes, including attainment, is the nature and type of secondary curriculum. Traditionally, the school curriculum has been divided into a three-tiered system of distinct tracks (Conant, 1959; Jarolimek, 1981). The term tracking itself denotes the ability-based sorting and grouping of students for instruction (Rubin, 2006). The rationale for curricular differentiation emanates from the notion that learning and instructional effectiveness can be enhanced by forming groups of students who are homogenous in terms of ability, performance, and motivation (Argys, Rees, & Brewer, 1996; Oakes, 1985, 1992). *Career-technical education* (CTE, also referred to as *vocational*) programs are viewed as pathways to both work and various types of further education (Lynch, 2000), whereas *college-preparatory* (also referred to as *academic*) tracks have the purpose of preparing students for postsecondary education at traditional four-year institutions (Hallinan, 2004).

Students without distinct CTE or college-preparatory concentrations are classified as following a *general* curriculum, which typically consists of an unspecified sequence of courses "under the umbrella or influence of a pseudoacademic concentration" (Stone & Aliaga, 2005, p. 127). Although in recent years curriculum placement schemes have been organized along more well-defined lines of student ability and motivation (Gartin, Murdick, Imbeau, & Perner, 2002; Slavin, 1990; Tomlinson, 1999), most students are placed in instructional groups that reflect a de facto reproduction of the traditional three-tiered curriculum structure (Dornbusch, Glasgow, & Lin, 1996; Hallinan, 2004).

The most common determinants of curriculum placement include a mixture of academic achievement and standardized test scores (Hallinan, 2003; Heubert & Hauser, 1999), teacher recommendations (Oakes & Guiton, 1995), personal intentional choice (Delany, 1991), parental and peer influences (Kilgore, 1991; Useem, 1991), as well as organizational and logistical exigencies at the local school level (Garet & DeLany, 1988; Useem, 1992). Tracking has been attacked for exacerbating social stratification given their particularly negative effects on students from underprivileged backgrounds (Burris & Welner, 2005; Burris, Wiley, Welner, & Murphy, 2008; Lucas, 1999; Oakes & Wells, 1998; Rubin, 2003, 2006; Yonezawa, Wells, & Serna, 2002). Specifically, critics excoriate the disproportionately large number of minority and low-income students that are placed into low-ability tracks (Hoffman, 2003). Since lower tracks are typically associated with stunted curricula and poor teaching quality (Heubert & Hauser, 1999), unequal resource allocation leads to the establishment of a status hierarchy within school systems that has particularly negative repercussions on the academic achievement of underserved students (Gamoran & Weinstein, 1998). However, longstanding attempts to de-track schools have had limited success because many teachers (Lee, Dedrick, & Smith, 1991; Riehl & Sipple, 1996), as well as parents of high achieving students (Oakes & Guiton, 1995), have a preference for classes that offer more homogenous ability profiles.

Numerous research studies have addressed the effects of differential curriculum placement on educational outcomes (Loveless, 1999). Ample evidence supports the positive impact of college-preparatory curricula on secondary academic achievement, dropout rates, and college attendance (Broussard & Joseph, 1998; Gamoran & Mare, 1989; Lee, Burkam, Chow-Hoy, Smerdon, & Geverdt, 1998; Lee, Croninger, & Smith, 1997; Natriello, Pallas, & Alexander, 1989). These findings are unsurprising given that students in college-preparatory tracks receive more engaging curriculum (Oakes, 1992, 2008), have higher-quality teachers (Coleman, 1995; Crosby & Owens, 1993; Marsh & Raywid, 1994), are surrounded by more academic role models (Hallinan, 2004), and benefit from an overall more stimulating academic climate that fosters positive educational outcomes (Hallinan, 2003).

A more ambiguous picture has emerged regarding outcomes of CTE curricula. Secondary CTE has been of particular interest to policymakers due to the resource intensity of career-focused high school programs (Cavanagh, 2005b; Gray, 2004). Several studies have shown that CTE curricula have a positive impact on high school completion rates (Kulik, 1998), academic achievement (Stone & Aliaga, 2005), and aspirations toward pursuing two-year postsecondary programs (Rojewski, 1997). However, other investigations have either failed to ascertain beneficial effects of CTE concentrations on high school completion (Agodini & Deke, 2004; Pittman, 1991), or have identified detrimental effects in terms of a reduced likelihood of college attendance (Arum & Shavit, 1995). A third stream of research has found that, when compared to general-track students, an integrated curriculum of CTE and college-preparatory courses can foster positive outcomes, including higher rates of secondary and postsecondary educational attainment (Castellano, Stringfield, & Stone, 2003; Fletcher, 2009; Plank, 2001; Plank, DeLuca, & Estacion, 2008). Overall, decades of research on secondary CTE outcomes have produced consistently inconsistent results, thus posing a major challenge to education policymakers.

Selection bias has been identified as a principal reason for the multitude of contradictory evidence on high school curriculum effects (Lee & Ready, 2009). Selection bias occurs when individuals either self-select into a program or intervention, or are subject to exogenous assignment based on some underlying rationale (Bryson, Dorsett, & Purdon, 2002; Dehejia & Wahba, 2002). In the presence of selection bias it is unfeasible to link participation in a certain curriculum to variations in educational outcomes because comparison groups may be systematically different from each other. While selection bias can be neutralized through experimental designs using random assignment, implementing such designs is often unfeasible in practice due to a host of ethical and logistical barriers (Gemici & Rojewski, 2007; Moore, Graham, & Diamond, 2003; Titus, 2007). Against this backdrop, propensity score matching (PSM) offers a sound methodological alternative to remedy the issue of selection bias with post-intervention observational data. In essence, PSM mitigates the negative effects of selection bias by statistically balancing comparison groups, post hoc, on a set of causally-relevant, observable background variables. Once comparison groups have been balanced, treatment or intervention effects can be estimated free of overt selection bias.

PSM is a highly effective method to apply in the analysis of large-scale datasets. The National Longitudinal Survey of Youth 1997 (NLSY97, U.S. Bureau of Labor Statistics, 2009a) is a large-scale dataset that contains relevant information for examining the effects of differential high school curriculum placement on educational attainment. The dataset comprises an original sample of 8,984 respondents that is representative of U.S. residents in 1997 who were born between 1980 and 1984. Key features of the NLSY97 include the availability of a wide variety of student, family, and education-related background variables, as well as transcript-based high school curriculum information. Thus, the NLSY97 was particularly well-suited for the purpose of this study.

Statement of Purpose

The purpose of this study was to examine the effects of CTE and college-preparatory high school curricula on secondary and postsecondary educational attainment. Outcomes for participants in each of these specialized curriculum types were compared separately to those for individuals who completed a general high school curriculum. Educational attainment was defined as the highest level of formal education completed by an individual in the NLSY97 dataset as of 2007, the most recent year for which NLSY97 data had been released at the time of writing. Specific secondary and postsecondary educational attainment categories included 0 (*no high school diploma or GED*), 1 (*GED*), 2 (*regular high school diploma*), 3 (*two-year college degree*), and 4 (*four-year college degree*). Given recurring debates over the resource intensity of secondary CTE (Cavanagh, 2005b; Gray, 2004), educational attainment outcomes for individuals enrolled in CTE curricula were of particular interest.

Research Questions

1. Is there a statistically significant difference between CTE and general-track students on secondary and postsecondary educational attainment?
2. Is there a statistically significant difference between college-preparatory and general-track students on secondary and postsecondary educational attainment?

Conceptual Framework

In quantitative research, theoretical perspectives establish a foundation from which hypotheses about relationships between predictor and outcome variables or constructs can be explained (Creswell, 2003). In a related fashion, conceptual frameworks use principles from relevant areas of scholarship to organize research and provide a meaningful context for the interpretation of findings (Reichel & Ramey, 1987; Smyth, 2004). Most importantly, they should "allow a schema for identifying critical issues and allowing for solutions" (Rojewski, 2009, p. 20). The conceptual framework underlying this study considers the essential relationships between student and school-level covariates, high school curriculum type, and educational attainment.

Covariates

Certain student and school-level variables, or covariates, have been established as predictors of placement into different curriculum tracks. The most important of these student and school-level predictors include academic achievement, ability, socioeconomic status, race/ethnicity, postsecondary plans, disability status, behavioral profile, participation in vocational programs, and school affluence (Agodini, Uhl, &

Novak, 2004; Jones, Vanfossen, & Ensminger, 1995; Oakes, Selvin, Karoly, & Guiton, 1992; Stone & Aliaga, 2005). The same vector of covariates influences high school completion (Jimerson, Egeland, Sroufe, & Carlson, 2000; Lee & Staff, 2007; Plank et al., 2008; Rylance, 1997; Swanson, 2004), as well as postsecondary enrollment and attainment (Hossler & Stage, 1992; Kao & Tienda, 1998; Rojewski, 1997). The fact that these covariates are important predictors of both curriculum placement and educational attainment allows for their use in a propensity score matching model. Propensity score matching uses a defined vector of covariates known to impact treatment selection and outcomes of interest to create equivalent comparison groups. The creation of equivalent groups based on a common vector of influential covariates is a prerequisite for examining causal effects between high school curriculum type and educational attainment.

Curriculum

The nature of instruction is considered “the proximal and most powerful factor in student engagement and learning” (National Research Council & Institute of Medicine, 2004, p. 60). Instruction is often used as a synonym for curriculum, whereby the latter specifically refers to “the attainment of learning objectives” (Laska, 1984, p. 212). Engaging curricula are expected to provide interactive learning opportunities, elicit positive emotional responses, and foster psychological investments in learning (Fredricks, Blumenfeld, & Paris, 2004). Specialized curricula with either career-technical or college-preparatory concentrations can create a meaningful, purpose-driven, and engaging learning environment, leading to better secondary and postsecondary educational attainment. Moreover, the fact that many students consider the theory-driven nature of the general-academic curriculum to be irrelevant for their lives often results in a gradual disengagement process from school (Kelly & Price, 2009). For those students, career-oriented high school programs, which are inherently context-based and application-focused (Advisory Committee for the National Assessment of Vocational Education, 2003), may represent an effective alternative curriculum.

Educational Attainment

Educational attainment is defined as “the highest grade completed within the most advanced level attended in the educational system of the country where the education was received” (United Nations, 2007, p. 176). Educational attainment in the U.S. is directly linked to economic growth and technological progress (National Research Council, 2001; OECD, 2003), since higher levels of educational attainment lead to increases in labor quality and productivity gains (Jorgenson, Ho, & Stiroh, 2003). Ancillary outcomes of educational attainment include improved health status (Hammond, 2003; Kenkel, 1991), reduced welfare dependency (An, Haveman, & Wolfe, 1993), and reduced crime rates (Yamada, Yamada, & Kang, 1991). Against this backdrop, the conceptual framework underlying the present study allows for a causal-comparative examination of the relationship between high school curriculum type and educational attainment. In this study, educational attainment is conceptualized as the highest level of formal education completed by an individual in the NLSY97 dataset as of 2007, the most recent year for which NLSY97 data were released at the time of writing. Specific educational attainment levels include no high school diploma or GED, GED, regular high school diploma, two-year college degree, and four-year college degree.

Model

The conceptual framework for this study connects student and school-level covariates, high school curriculum type, and educational attainment. Select student and school-level covariates influence placement into one of three broad high school curriculum types. CTE and college-preparatory curricula represent mutually exclusive treatment conditions that are compared separately against a general curriculum, which represents the control condition. Since propensity score matching allows for the creation of equivalent curriculum groups based on an observable set of covariates known to impact both selection into treatment and outcomes of interest, the causal effects of high school curriculum type on educational attainment can be determined free of overt selection bias. For students enrolled in CTE curricula, potential attainment effects may emanate from the inherently context-based and application-focused nature of CTE curricula when compared to general curricula. For students enrolled in college-preparatory curricula, attainment differences may emanate from a more purpose-driven academic orientation and access to higher-quality educational resources when compared to general curricula. Figure 1.1 provides a visual model for the present study's conceptual framework.

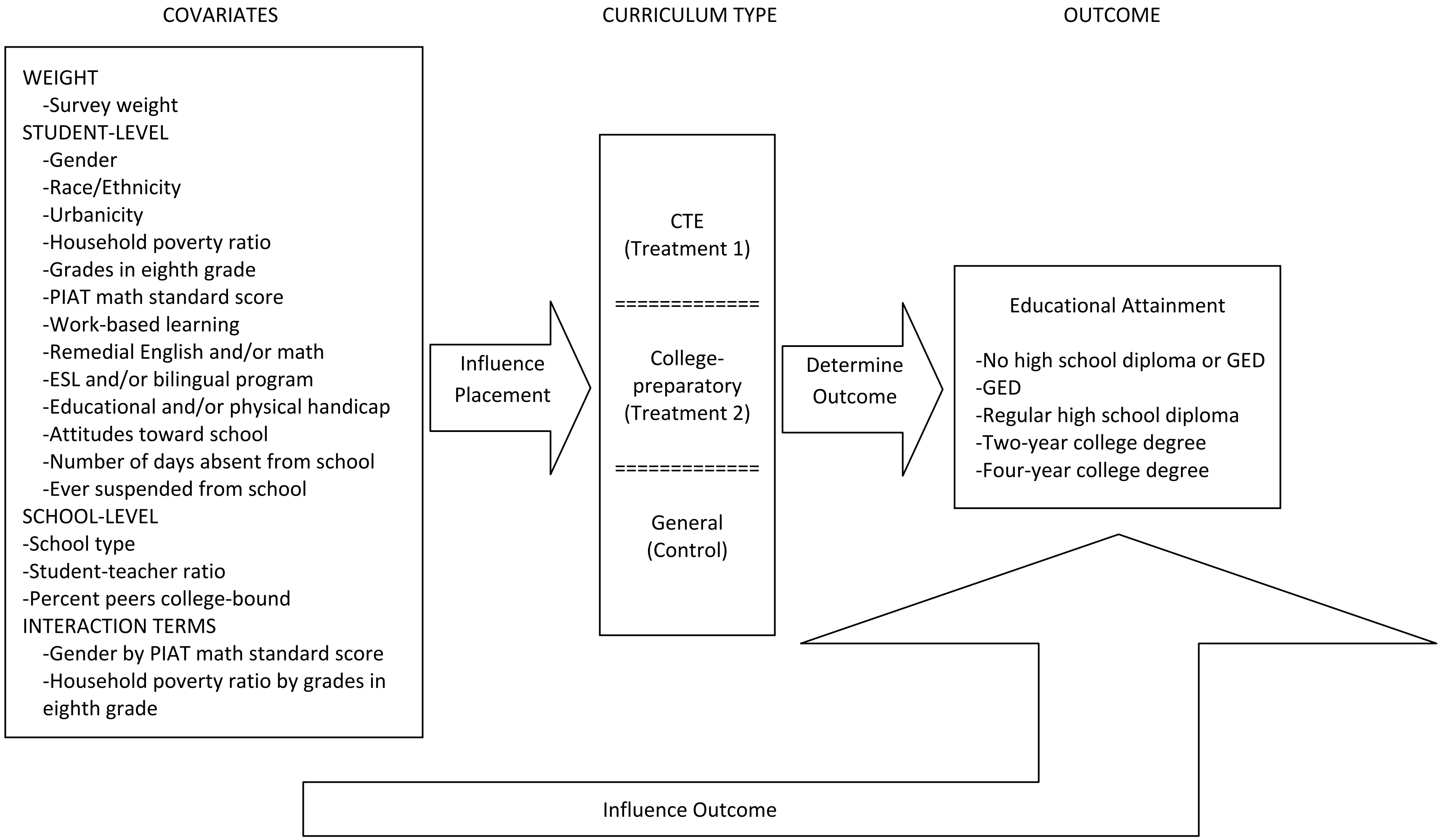

Figure 1.1. Conceptual framework for the study. The survey weight covariate addresses the complex sampling methods used in NLSY97. Interaction terms are included to improve the fit of the logistic regression model for propensity score estimation.

Importance of the Study

Given that educational attainment is an important determinant of labor productivity and technological progress (OECD, 2003), clarity about the usefulness of different high school curricula in bringing about positive educational attainment outcomes is critical. Decades of research on curriculum effects have failed to provide such clarity, mainly due to the inadequate treatment of selection bias in the design of many studies (Lee & Ready, 2009). By using propensity score matching as an efficient means for eliminating overt selection bias, this study seeks to generate more robust evidence on the effects of different high school curricula on secondary and postsecondary educational attainment. The importance of such new evidence is threefold. First, government agencies have a vital interest in data on curriculum effects based on methodologically sound research designs in order to make informed funding and policy decisions. This strong interest in data on the effects of curriculum tracking is reinforced by the increasing importance placed on international comparisons of student achievement. Second, local school personnel require information about attainment outcomes of high school tracking to improve curriculum design and delivery. Finally, stronger research evidence may assist students and parents in making the right high school curriculum choices that will allow them to reach a desired level of educational attainment.

CHAPTER 2

REVIEW OF LITERATURE

Curriculum

Curriculum is a "set of learning objectives … and the curricular content that is utilized for the attainment of these objectives" (Laska, 1984, p. 211). Schiro (2008) provided a concise overview of the four principal curriculum ideologies that have evolved historically and are in existence today, including the scholar academic, social efficiency, learner-centered, and social reconstruction curriculum ideologies. The objective of the scholar academic curriculum ideology is to foster the acquisition of cultural values through the academic disciplines, such as literature, history, and mathematics. In the context of this ideology, Hirsch (1987) promoted the idea of *cultural literacy* that could be attained by grounding the curriculum in the academic disciplines, whereby each discipline represents a particular component of literate culture. Likewise, Adler (1940) proposed the *great books* approach, whereby a list of culturally essential readings would constitute the core curriculum for a liberal education. Synonymous terminology for the scholar academic curriculum ideology includes *academic rationalist* (Eisner, 1974) and *knowledge-centered* (Ellis, 2004) ideologies.

The social efficiency curriculum ideology aims at serving society as a whole by training youth in occupational and workplace skills. As such, the value of learning lies in the purpose-driven acquisition of skills and capabilities that are viewed as vehicles of social and economic productivity. Bobbitt (1918) formally introduced the social efficiency approach as the foundation of the school curriculum, whereby specific curriculum objectives revolved around abilities, attitudes, and forms of knowledge that were measurable and demonstrable. Learning was to be organized efficiently in terms of financial and time investments. Overall, the social efficiency approach seeks to equip youths with the skills needed to function efficiently in society. The social efficiency curriculum ideology is also known as *technological* (McNeil, 1977) and *social behaviorist* (Schubert, 1996).

Focusing on the needs of students, the goal of a learner-centered curriculum ideology is to bring out the talents and strengths inherent in every individual. Fostering personal growth is the principal concern and occurs as a result of the interaction between the person and the environment. Thus, learner-centered curricula provide the structural environment to maximize opportunities for personal growth. Dewey (1916, 1948) is among the best-known proponents of the learner-centered curriculum ideology and highlights the importance of personally experiencing reality by involving students in the learning process. The curriculum guides teachers in creating opportunities for dynamic communication through small group interaction (Gandini, 2004). Synonyms for the learner-centered curriculum ideology include *therapist* (Fenstermacher & Soltis, 1992) and *experiential-cognitive* (Posner, 1992).

Finally, recognizing the social injustices emanating from racial/ethnic, gender, and economic inequalities, the social reconstruction ideology suggests that the "purpose of education is to facilitate the construction of a new and more just society that offers maximum satisfaction to all of its members" (Schiro, 2008, p. 6). The curriculum functions as a central medium to inform individuals about societal shortcomings and enable them to construct an improved social order. Accordingly, the role of educators is that of activists who lead the battle for social and economic justice (Giroux, 2006). In the literature, the social reconstruction curriculum ideology is also referred to as *society-centered* (Ellis, 2004) and *social-meliorist* (Kliebard, 2004).

Historical Developments in Curriculum Tracking

Historically, the general curriculum in the U.S. reflected the scholar academic model, for the original purpose of schooling in the latter part of the 19th century was aimed at preparing a small elite for higher education (Gordon, 2008). Early on, deliberations over college-preparatory and vocational functions of the school curriculum sparked controversies over the fundamental purpose of education. Within the context of race, Washington (1903) and Du Bois (1903) debated whether vocational education would be detrimental or beneficial for African Americans. Washington strongly supported industrial education for African Americans because he considered economic stability through the acquisition of industrial skills the prerequisite for a life marked by autonomy and freedom. Du Bois contested Washington's perspective, which he thought would perpetuate the dependence of African Americans on the academically educated and economically powerful White upper class. Instead, Du Bois suggested sending the most talented tenth of African Americans to colleges and universities for a liberal arts education. He saw this group as providing the intellectual leadership that would eventually solidify the freedom of all African Americans. Although this controversy was firmly grounded in the issue of race, it planted the seed for the debate about the true purposes of vocational and academic education that continues to shape education policy even today (Gordon, 2008).

More than a decade after the Washington-Du Bois controversy, Dewey (1916) and Prosser (see Prosser & Allen, 1925) disputed the nature of education and, specifically, the purpose of vocational education. Although both appreciated the value of vocational education, they did so from different perspectives. Dewey promoted a learner-centered curriculum ideology, in which vocational education would be an essential way to ensure the education of students for a democratic society (Lakes, 1985). Students were to be engaged in the learning process to become active participants in their world (Hyland, 1993). Dewey criticized conservative pedagogical theory that, in his view, isolated a student's various capacities for learning. Instead, he advocated for an organic, unifying, and reflective education model in which vocational education functioned as an integral part of the comprehensive curriculum (Scheffler, 1995). Dewey further posited that knowledge of shop work, cooking, sewing, and other trades and crafts would best prepare students to understand the science of tools and processes used in work. As such, he concluded that vocational education would guarantee *all* citizens the right to full participation in industrial policy and decision making.

In contrast to Dewey's (1916) concept of education for democracy, Prosser's (see Prosser & Allen, 1925) perspective on vocational education was driven by a social efficiency curriculum ideology. Schools should prepare individuals for those occupations for which they would be most suited in terms of ability and intelligence. With the human resource needs of rapidly expanding industrialization in mind, vocational education was deemed an efficient way of creating an educational environment that would emulate the tasks and conditions of the work place (Drost, 1967). Opportunities for personal growth were secondary to the needs of industry that demanded training in narrowly defined manual skills. Against this backdrop, policymakers at the time preferred Prosser's social efficiency doctrine over Dewey's learner-centered approach. Prosser's ideas were codified in the Smith-Hughes Act of 1917 that set the foundation for occupation-specific secondary vocational education programs (Hyslop-Margison, 2001). As such, the Smith-Hughes Act mandated a new curriculum aimed at the working class instead of exclusively catering to the needs of college-bound upper-class students (Gray, 1991).

The expansion of industrialization, combined with the rapid influx of immigrants between 1890 and 1940, led to a shift in the U.S. public school system, whereby the sorting of students into different curriculum tracks was considered an efficient way of creating homogeneous groups of academic ability and future occupational orientation (Mayer, 2008). The curriculum was changed to educate students from diverse backgrounds for citizenship in the broadest sense, and the different curriculum tracks meant to prepare students for a wide range of future occupations covering all societal strata (Powell, Farrar, & Cohen, 1985). Critics, however, have highlighted that tracking was introduced as a mechanism of social control with the objective to foster social and cultural reproduction (Anyon, 1988; Apple, 1982; Spring, 2009). This criticism is supported by demographic and student outcome data showing that vocational tracks contain a disproportional number of minority and lower class students who achieve less favorable education outcomes (Boesel, Hudson, Deich, & Masten, 1994; Silverberg, Warner, Fong, & Goodwin, 2004).

Besides the impact of industrialization and large-scale immigration, the tracking of students according to ability was strongly supported by the development of intelligence tests as predictors of academic performance (Kliebard, 1992). Such tests were used to facilitate sorting procedures into different tracks and occupational destinations. Eventually, a three-tiered school system consisting of college-preparatory, general, and vocational tracks was created during the 1940s (Conant, 1959; Jarolimek, 1981; Mayer, 2008). This new track-based system had three well-differentiated purposes (Scott & Sarkees-Wircenski, 2008):

> The function of the secondary school today is at least threefold - the preparation of one group of pupils for college, the provision of terminal vocational education for another segment, and the furnishing of a program of education for general living which will meet the needs of both these groups and also of other pupils who are not going to college and for whom the vocational curriculums are not suitable. (Harriger, 1948, p. 163)

Modern Developments in Curriculum Tracking

Over the past two decades, the rigid three-tiered tracking scheme of CTE, college-preparatory, and general curricula has been complemented by additional measures aimed at further distinguishing students by individual ability and motivation. Examples of such measures include differentiated instruction and ability grouping. Differentiated instruction refers to the creation of flexible learning environments and curricula to address individual differences in students' skills, strengths, and needs (Gartin, Murdick, Imbeau, & Perner, 2002; Tomlinson, 1999), whereas ability-grouping is defined as "any school or classroom organization plan that is intended to reduce the heterogeneity of instructional groups" (Slavin, 1990, p. 471). Following these models, individual courses are typically divided into an ability continuum ranging from basic and regular to honors and advanced placement courses (Hallinan, 1994). Despite these more detailed differentiation schemes, most students are placed in instructional groups that reflect a de facto reproduction of the traditional three-tiered curriculum structure (Dornbusch, Glasgow, & Lin, 1996; Hallinan, 2004).

Today, college-preparatory curricula continue to have the purpose of preparing students for postsecondary education at traditional four-year institutions (Hallinan, 2004), whereas the goals of general curricula continue to be ill-defined. This lack of direction often has negative repercussions for students enrolled in general curriculum tracks (Hallinan, 1994; Oakes, 1994; Oakes, Selvin, Karoly, & Guiton, 1992). In making reference to the general track, Stone and Aliaga (2005) observed that “the rest of the students are left to wander haphazardly through their high school years, mostly under the umbrella or influence of a pseudoacademic concentration” (p. 127).

CTE curricula have experienced the most pronounced changes over the past two decades. Formerly considered a terminal high school track for work-bound youth, reform efforts that started in the late 1980s have transformed CTE into a track with a clear mandate to prepare students for both the transition into work *and* post-secondary education (Gray, 2004). Reform efforts were codified in the Carl D. Perkins Vocational and Applied Technology Education Act (1990), which had the goal to integrate secondary CTE and core academic courses and position secondary CTE more clearly as a direct pathway to postsecondary education and training. The Carl D. Perkins Vocational and Technical Education Act (1998) and the Carl D. Perkins Career and Technical Education Act (2006) further intensified the notion of a seamless transition from secondary CTE to postsecondary education and training. Collectively, these legislative initiatives were driven by an increased need to raise employment options through higher educational attainment (Foster, 2007). The CTE curriculum itself has been defined as earning 3.0 or more credits in one of 10 broad occupational program areas, including agriculture, business, marketing, health care, protective services, technology, trade and industry, food service and hospitality, child care and education, and personal and other services (Levesque, 2003).

Curriculum Placement Mechanisms

It is generally assumed to be more efficient to teach groups of students who are homogeneous in terms of ability (see Argys, Rees, & Brewer, 1996; Hallinan, 1994; Kilgore, 1991; Oakes 1985, 1992). The mechanisms that lead to such homogeneous ability groups, or curriculum tracks, are complex and subject to a variety of influences. Test scores build the foundation for teacher recommendations and are, thus, one of the principal factors that play into the track placement decision. Heubert and Hauser (1999) recognized the potentially negative long-term impact of test-based track placement on student ability. They posited that once students were placed in lower tracks due to low test scores, poor curricula, deficient teaching quality, and low teacher expectations generally associated with lower-track courses would forcibly lead to low scores on subsequent tests. Therefore, Heubert and Hauser concluded that the practice of placing students in different curriculum tracks based on test scores was a potential dilemma that would perpetuate negative student outcomes.

Several ancillary factors beside test scores complement the track assignment process. Such factors include intentional student choice, influence exerted by parents and peers, as well as organizational exigencies at the local school level (Mickelson & Velasco, 2006). To the same degree that students have gained more influence over tracking decisions, the impact of peer cultures has increased in importance (Kilgore, 1991). Depending on the particular school system, the influence parents have over course and track placements is also considerable. Useem (1991) studied the magnitude of parental influence on school policies in the middle grades, a period during which students' career plans are still relatively vague. Her study revealed that parental influence over course placements was strongly related to social class background, whereby the extent to which external input was welcomed, restricted, or thwarted was dependent on parents' socioeconomic status and education level.

Organizational exigencies play an important role in track placement decisions. Through a complex analysis of student transcript data, Garet and DeLany (1988) discovered that track placements were largely dependent on the composition of the student body and the organization of the curriculum. Specifically, schools implemented track placements through an ex ante sorting process "in which positions in the curriculum (courses or sections) are created, and students are matched with positions" (p. 75). This sorting process, however, was often frustrated by resource pressures arising from day-to-day operations, whereby efficient scheduling took precedent over matching students with adequate courses based on ability levels. Overall, Hallinan (2003) emphasized that factors unrelated to achievement were subjective and could lead to bad matches between student ability and a given curriculum, with negative repercussions on learning outcomes.

One of the most comprehensive studies on track assignment factors was conducted by Oakes and Guiton (1995). Based on data from a two-year longitudinal investigation, the study closely examined the dynamics of tracking decisions at three metropolitan high schools on the West Coast. The authors identified an intricate web of elements influencing tracking decisions at different levels and based on different rationales. Specifically, they discovered that schools and teachers considered students' abilities and motivations to be predetermined, with little room for improvement in their educational development through curricular interventions. Consequently, the curriculum structure was organized around accommodating rather than altering or improving students' performance characteristics. Since race, social class, and ethnicity were considered by school personnel as predictors of ability and motivation, these demographic factors played an additional role in curriculum decisions. The authors summarized their insights:

> We conclude, then, that high school tracking decisions result from the synergy of three powerful factors: differentiated, hierarchical curriculum structures; school cultures alternatively committed to common schooling and accommodating differences; and political actions by individuals within those structures and cultures aimed at influencing the distribution of advantage. (p. 30)

This conclusion provides a comprehensive perspective on the complex interplay of different factors that influence track placement mechanisms.

Disadvantages Associated with Curriculum Tracking

Even though curriculum tracking has dominated the U.S. educational landscape for decades, the practice has been met with strong and continuous opposition since the mid-1980s (Oakes, 1985). The term detracking has been coined to denote a change in traditional curriculum placement practices and to enroll all students in a college-preparatory curriculum, regardless of their race/ethnicity, social class background, or ability level (Alvarez & Mehan, 2007; Lewis, 2007; Wells & Serna, 1996). The principal allegation made by supporters of the detracking movement revolves around the widening of the racial achievement gap by creating track-based segregation along racial/ethnic and socioeconomic status lines (Oakes, 2008). In a study questioning traditional views of student achievement, Burris and Welner (2005) argued that the racial achievement gap emanated from a tracking-based imbalance in educational opportunities. Since African American and Hispanic minority students are overrepresented in lower tracks that afford students less educational opportunities and resources (Hoffman, 2003), their achievement forcibly suffers when compared to that of their higher-tracked peers. Moreover, since minority status is highly correlated with low socioeconomic status (Lucas, 1999), lower class students have a drastically reduced chance of being placed into a challenging curriculum track, even if they are academic high-achievers (Vanfossen, Jones, & Spade, 1987). A series of studies about the negative effects of tracking on students from underprivileged backgrounds has consistently confirmed the same patterns of disadvantage in a wide range of school settings (see Boaler & Staples, 2008; Burris, Wiley, Welner, & Murphy, 2008; Oakes & Wells, 1998; Rubin, 2003, 2006; Yonezawa, Wells, & Serna, 2002). Collectively, these studies provide evidence for the racial and class-based stratification effects that occur through tracking in contemporary American public education.

While teachers may appreciate the arguments put forth by supporters of detracking, many prefer classes with a relatively homogeneous ability profile that they perceive as easier to teach than mixed ability classes (Oakes & Lipton, 1992). Teachers often internalize issues around low student ability and discipline by lowering their expectations regarding their students and their own instructional self-efficacy, eventually resulting in lower job satisfaction (Ashton & Webb, 1986; Riehl & Sipple, 1996). Thus, school organizational factors, including track level assignments, have a direct influence on teacher self-efficacy (Cooper, Burger, & Seymour, 1979; Raudenbush, Rowan, & Cheong, 1992). Lee, Dedrick, and Smith (1991) provided further evidence showing that teacher satisfaction and self-efficacy were directly tied to their control over the curriculum and the perceived ability level of their students.

Similar to teachers' reservations regarding detracked curriculum structures, many parents of high achieving students who are enrolled in homogeneous college-preparatory curricula push for tracking to remain in place (Rubin, 2008). Wells and Serna (1996) examined how upper-class parents used their economic, political, and cultural capital in resisting detracking reform at 10 racially and socioeconomically mixed schools. The *elite* status of parents opposed to detracking was characterized:

> The elites discussed here had children enrolled in the detracking schools and thus constitute the subgroup of local elites active in shaping school policies. Their practices were aimed at maintaining a track structure, with separate and unequal educational opportunities for "deserving" elite students and "undeserving" or non-elite students. (p. 95)

The authors identified four general approaches used by elite parents to undermine detracking efforts in order to preserve at least some of the privileges of high track placement for their children. The most frequently employed strategy consisted of threats to leave the school if tracking privileges were revoked. This phenomenon of elite flight was deemed highly effective given that the economic, political, and cultural capital and prestige associated with local upper-class parents is a vital factor in schools' ability to remain politically viable in the community. The second strategy consisted of systematically co-opting schools' institutional elites made up of educators and administrators who have power and authority within the school system and who feel a sense of commitment to cater to the needs of local elite parents. These co-opting mechanisms are mainly based around demands to enable their children to score highly on standardized tests, which they felt was best accomplished through a rigid tracking structure. The third strategy consisted of concerted efforts to gain buy-in of middle-class parents who themselves aspire to become part of local decision-making and power structures by putting themselves forward for election to governing bodies within the school system. Finally, the authors discovered influential parents' attempts to bribe into taking measures to preserve academic privileges for their children in return for their willingness to support at least some minor detracking reforms. These bribes often take the form of special high-profile course or program offerings into which a disproportionate number of advanced local elite students are recruited. Overall, Wells and Serna's study uncovered the unique tactics employed by upper class parents in ensuring that continued preferential access to a high-quality curriculum is given to their children. These efforts are motivated by the widely held notion that curriculum track placement is a predictor of student outcomes, including educational attainment.

Status Deprivation Effects of Tracking

` Most national school systems worldwide follow some sort of ability-based tracking scheme (LeTendre, Hofer, & Shimizu, 2003; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006). However, while in Anglo-Saxon countries tracking schemes are mostly organized within schools, many European countries implement tracking in specialized schools that focus on offering one distinct type of either vocational-technical or academic track (Brunello & Checchi, 2007). Schools that offer within-school tracking are also referred to as *multilateral* schools, whereas schools that specialize in a particular curriculum are referred to as *categorical* schools (Van Houtte & Stevens, 2009). Yet, no matter whether countries follow a multilateral model as in the U.S. and the U.K. or a categorical model as in Germany and Scandinavian countries, there is increasing evidence from international comparative tests that the timing of tracking seems to be much more impactful on student outcomes than the particular tracking model employed. Specifically, results from the Program for International Student Assessment (OECD, 2007a) demonstrate that countries with late onsets of tracking (usually after the age of 15) consistently outperform others such as the U.S. in which ability grouping occurs early on (Cavanagh, 2005a). The best performing countries, including Finland, Canada, Hong-Kong China, and Japan, have largely untracked, comprehensive school systems (Oakes, 2008), whereas the U.S. and other tracking-intensive school systems perform significantly worse.

Status deprivation effects based on placing students in lower, less prestigious tracks is considered a principal cause of negative student outcomes found in highly-tracked school systems (Hargreaves, 1967; Rosenbaum, 1976; Schwartz, 1981). Two more recent studies corroborate the detrimental effects of early tracking on student disengagement and poor attitudes toward school. Berends (1995) examined the effects of educational stratification on students' orientations and attitudes toward school. Using data from High School and Beyond, the author found students in general and vocational tracks to be less engaged academically and exhibiting stronger anti-school orientations when compared to their college-preparatory track peers. He explained these outcomes through social bonding theory (Hirschi, 1969), whereby the activities and people of an institution are instrumental in creating individuals' attachment, commitment, involvement, and belief in institutional norms. As such, general and vocational track students' attitudes and behaviors towards school change during high school, as they become increasingly aware of the disadvantaged prestige status inherent in lower tracks when compared to college-preparatory tracks. This awareness leads to reduced social bonding with school as an institution, causing negative repercussions for students in terms of educational outcomes.

An interesting recent study by Van Houtte and Stevens (2009) compared multilateral and categorical schools in Belgium to explain negative effects of within-school tracking through reduced social bonding. Their study compared variations in educational involvement of students who were in mixed-track schools with those who were in specialized schools that offered an exclusive vocational-technical curriculum. Findings were intriguing in that vocational students in multilateral schools showed less engagement than vocational students who were in categorical institutions. These findings suggest that "in multilateral schools, vocational students compared themselves with academic-track students, consistent with the hypothesis of increased status deprivation, resulting in even stronger anti-school attitudes" (p. 943). When extrapolated to the U.S., it can be concluded that the commonly-practiced within-school tracking model may indeed be more detrimental to educational outcomes of lower-tracked students when compared to following a between-school tracking model.

Factors Impacting Curriculum Placement and Educational Attainment

This study was based on the hypothesis that participation in certain high school curricula has a causal effect on secondary and postsecondary educational attainment. Particular background factors, or variables, at the student and school levels have been identified in the literature to impact both curriculum choice and educational attainment. Influential background variables at the student level include gender, race, urbanicity, socioeconomic status, academic achievement, postsecondary educational aspirations, English proficiency, disability status, and student risk behavior (Levesque, 2003; Rojewski, 1997; Silverberg, Warner, Fong, & Goodwin, 2004; Stone & Aliaga, 2005). Influential background variables at the school level include the type of school control (i.e., public or private), the school’s socioeconomic status, and peer effects (Hanushek, Kain, Markman, & Rivkin, 2003; Jones, Vanfossen, & Ensminger, 1995). Students and school level background variables that influence curriculum placement and educational attainment have been identified in several studies, as outlined below.

Predictors of high school curriculum placement. Jones et al. (1995) conducted one of the most comprehensive studies of individual and school-level predictors of high school curriculum placement. Using multinomial logistic regression, which allows the determination of student and school-level variable effects on high school curriculum placement while controlling for influential confounders, the authors found that the odds of assignment to college-preparatory programs significantly increased if students were female, non-Hispanic, from a high socioeconomic status background, academic high achievers who exhibited high educational expectations, or enrolled in a high socioeconomic status school. General-track predictors included being male, Hispanic, as well as having a low socioeconomic status, educational expectations, and academic achievement. Finally, the odds of vocational track placement increased for students who were Black, exhibited lower educational expectations and grades, and attended a lower socioeconomic status high school. Overall, while the study clearly demonstrated a dependence of curriculum placement decisions on student and school-level background variables, the authors concluded that school and organizational factors may exert a greater influence on track placement than purely student-specific variables. Specifically, "students with similar individual characteristics may find themselves in different tracks, depending on the characteristics of the schools they attend" (p. 296).

A second influential study (Agodini, Uhl, & Novak, 2004) considered particularly those factors that influence participation in secondary career-technical education. The authors emphasized the importance of identifying the predictors of vocational track placement in order to develop more effective vocational reform policies aimed at facilitating the transition from school to postsecondary education for career-focused students. Data from the National Education Longitudinal Study of 1988 (NELS:88) were analyzed and produced several interesting findings regarding CTE track placement mechanisms. First, results confirmed existing notions about vocational students in that low academic achievement and educational aspirations were principal predictors of CTE curriculum placement. Students in the lowest third of eighth-grade reading achievement exhibited a seven percent higher likelihood of enrolling in a CTE curriculum than their otherwise identical peers who performed in the highest third. Male students were more likely to become vocational concentrators, as were students from lower SES backgrounds.

Even though Agodini et al.'s (2004) study corroborated most of the other commonly held predictors of CTE curriculum participation, including urbanicity, school type, student risk behavior, and postsecondary plans, their study yielded two interesting insights regarding the role of race/ethnicity and disability status. Whereas Black and White students exhibited the same odds for participation in CTE, Hispanic students were *less* likely to do so. This finding contradicts results from prior research that found higher CTE participation rates for Black and Hispanics (see Ekstrom, Goertz, & Rock, 1988; Oakes & Guiton, 1995; Oakes, Selvin, Karoly, & Guiton, 1992) and higher college-preparatory participation rates for White and Asian students (Braddock, 1990; Oakes, 1990). Agodini et al. hypothesized that lower Hispanic participation may be due to a desire to remain with their Hispanic peers who are enrolled in general-track curricula, as well as the fact that participation in limited English proficient programs leaves less time for taking vocational courses.

Students with special needs have been found to exhibit overall higher rates of CTE participation (Levesque, 2003). However, Agodini et al.'s (2004) investigation revealed that the relationship between special needs status and CTE participation stems from confounding factors rather than being rooted in a students' disability status itself. In fact, a statistically significant difference in participation rates of students with special needs disappeared once the effects of low academic achievement, low educational expectations, and low socioeconomic status were controlled. It may thus be possible that issues of multicollinearity may have influenced previous research with regard to explanations for the curriculum placement of students with special needs.

Two more recent studies (Fletcher, 2009; Stone & Aliaga, 2005) have examined predictors of high school curriculum placement. Both studies confirmed the important role of gender, race/ethnicity, and socioeconomic status in track placement. Beyond the influence of basic demographic variables, Stone and Aliaga identified academic achievement as a key predictor. Moreover, school size has been found to impact student achievement (Lee & Smith, 1997; Mayer, Mullens, & Moore, 2000; Mosteller, Light, & Sachs, 1996) and can, thus, be considered an indirect predictor of curriculum placement. Likewise, attitudes toward school is a direct influencer of student achievement (Germann, 1988), rendering attitudes an additional indirect predictor of curriculum placement.

Predictors of educational attainment. Several studies exist that focus on the predictors of secondary educational achievement. An increased focus has recently been placed on the effects of gender on secondary educational attainment. For instance, a rigorous assessment of high school transcripts between 1990 and 2000 found that girls consistently earned a higher number of core academic credits including mathematics, science, English, and social studies than boys, whereas no gender difference was ascertained in non-core credits (Perkins, Kleiner, Roey, & Brown, 2004). Recent research (Shettle et al., 2007) confirmed that females complete higher level courses, resulting in overall more academically challenging curricula that are more reflective of college-preparatory than vocational programs.

Other demographic and student/family-level variables have been identified as predictors of secondary educational attainment, the most prominent of which are socioeconomic status and race/ethnicity. Students from high socioeconomic backgrounds have a considerable advantage in terms of high school completion, as do students who are

White (Haveman, Wolfe, & Spaulding, 1991; Kaufman, & Bradbury, 1992; Natriello, McDill, & Pallas, 1990). In contrast, low academic achievement, frequent absenteeism, and disciplinary problems are risk factors that diminish the likelihood of high school completion (Gleason & Dynarski, 1998; Kaufman & Bradbury, 1992). School-level factors further contribute to high school completion. Particularly, large school size, low school socioeconomic status, and increases in state-mandated minimum course graduation requirements have been found to negatively influence student persistence and secondary educational attainment (Fitzpatrick & Yoels, 1992; Lillard & DeCicca, 2001; Werblow & Duesbery, 2009). Finally, attitudes toward school are an indirect predictor of educational attainment. Attitudes toward school are associated with student motivation and engagement (Skinner & Belmont, 1993), which are directly tied to student achievement (Turner, Thorpe, & Meyer, 1998) as a precursor to educational attainment. Overall, the same set of variables that has been found to predict high school curriculum placement also predicts secondary educational attainment.

While numerous studies have investigated predictors of postsecondary *enrollment*, there is a remarkable gap in the literature regarding predictors of actual postsecondary educational *attainment*. Most factors that predict postsecondary enrollment are identical with those that impact secondary educational attainment. Such basic student and school-level factors include socioeconomic status (Kao & Tienda, 1998), postsecondary educational aspirations (Rojewski, 1997), the influence and postsecondary aspirations of peers (Alfeld, Hansen, Aragon, & Stone, 2006; Harnish & Lynch, 2005), and characteristics of the school itself, such as school location and socioeconomic status

(Konstantopoulos, 2006). Overall, more favorable student and school level indicators lead to more desirable postsecondary educational attainment outcomes.

Beyond these common factors, two postsecondary-specific predictors have been ascertained. Although somewhat dated, Fuller, Manski, and Wise (1982) carried out an econometric analysis of the National Longitudinal Study of the High School Class of 1972 dataset to determine financial parameters in the pursuit of postsecondary education. Access to financial aid was a major determinant in individuals' decision to enroll in postsecondary education. The second postsecondary-specific factor revolved around family influences. Stage and Hossler (1989) examined family influences on the college attendance plans of ninth graders. At this predisposition stage of postsecondary decision-making the authors found that besides common demographic factors (i.e., gender, race/ethnicity, socioeconomic status), parental educational expectations and encouragement were instrumental in students' postsecondary planning. Finally, a recent study by Legutko (2008) examined various aspects of family influence on the postsecondary decisions of rural high school students. Particularly, the study focused on the effects of parent and sibling education, as well as the perceived family financial/class standing on students' postsecondary education decisions. With the exception of sibling effects, parental influences significantly impacted students' decision to pursue postsecondary education.

Clearly, the decision to enroll in postsecondary education offers very limited insights into actual postsecondary attainment, for 35 to 44 percent of two-year and 17 to 19 percent of four-year college degree students fail to actually complete their postsecondary degrees (Bradburn, 2002). Against this backdrop, it seems surprising that very few studies have examined factors impacting actual postsecondary educational attainment in terms of degree completion. One relevant study was conducted in the area of economically disadvantaged and urban youths. Rojewski (1997) investigated the impact of secondary CTE participation on students' work experience and postsecondary aspirations. Conceptually, his investigation set out to test assertions of the ability of secondary CTE to instill higher educational and occupational aspirations in students (see Kablaoui & Paulter, 1991). Using data from a large-scale national education dataset, Rojewski's findings contradicted that assertion, for increases in postsecondary aspirations were actually associated with decreasing involvement in secondary CTE courses. Since educational aspirations are a strong predictor of actual educational attainment (Qian & Blair, 1999), the degree to which students participate in secondary CTE can be considered an indirect predictor of postsecondary attainment.

Only one recent study directly examined predictors of postsecondary educational attainment. Specifically, Fletcher (2009) examined demographic and school curriculum variables in an attempt to isolate predictors of degree attainment. He found that gender, race/ethnicity, parental education, household income, high school track placement were important predictors of two-year college degree attainment. Specifically, being female and placed in a college-preparatory high school track significantly increased the likelihood of attaining a two-year college degree, whereas being African American and having parents with low levels of education significantly decreased the odds of successful completion of an Associate's degree. The same significant predictors were ascertained for the four-year college level.

Summary of relevant predictors. The literature on factors impacting high school curriculum placement and secondary and postsecondary educational attainment consistently demonstrates the influence of certain student, family, and school-level variables. Gender, race/ethnicity, urbanicity, socioeconomic status, academic achievement, participation in secondary CTE programs, English language ability, special needs status, attitudes toward school and educational aspirations, student risk behaviors, peer influences, as well as school type and socioeconomic status all exert either direct or indirect influence on high school curriculum choice and educational attainment. As such, a combination or subset of these variables should be used in any analysis that seeks to examine causal effects between high school curriculum type and educational attainment.

Student Engagement and Motivation

Student engagement in school and the learning process is considered a key factor in determining educational achievement outcomes (Fredricks, Blumenfeld, & Paris, 2004). Student engagement is a multidimensional construct that consists of three interrelated sub-constructs, including behavioral, emotional, and cognitive engagement (Fredricks et al.).

Behavioral engagement. Behavioral engagement entails a willingness to conform to the behavioral expectations of the school environment (Finn, Pannozzo, & Voelkl, 1995; Finn & Rock, 1997), a commitment to actively participate in the learning process (Birch & Ladd, 1997; Buhs & Ladd, 2001), and a feeling of dedication to the school as an institution (Finn, 1993). Behavioral engagement is closely linked to the ability of curriculum and instruction to foster student involvement through authentic pedagogy, interchange, and reciprocity in the classroom (Newmann, Marks, & Gamoran, 1996). By the same token, a structural environment that provides students with a high number of opportunities for interactive learning facilitates growth in academic achievement (Sørensen & Hallinan, 1978, 1986).

Emotional engagement. Emotional engagement emanates from a students' affective reactions toward the learning environment (Connell & Wellborn, 1991; Skinner & Belmont, 1993). Emotional engagement is conceptually linked to the taxonomy of affective domain (Krathwohl, Bloom, & Masia, 1964) and theory of reasoned action (Fishbein & Ajzen, 1975). The taxonomy of affective domain spans various stages during which an individual successively develops affective behaviors toward an object. Laforgia (1988) noted that these affective behaviors are "placed along a hierarchical internalization continuum which ranged from the mere awareness of phenomena (receiving) to the development of a life outlook (characterization)" (p. 408). With rising levels of internalization, the affect exercises increasing control and guidance over the individual's behavior (Seels & Glasgow, 1990). The theory of reasoned action holds that attitude is a principal determinant of intention, which itself is the key predictor of human behavior (Langdridge, Sheeran, & Connolly, 2007; Olson & Zanna, 1993). Since success in the affective domain is an antecedent to success in the cognitive domain, affective dimensions greatly impact students' emotional engagement with the learning process and related levels of educational achievement.

Cognitive engagement. Cognitive engagement refers to a willingness to invest in learning and involves a commitment to putting forth the necessary effort to acquire intellectually demanding material (Fredricks et al., 2004). It goes beyond the compliance with school norms and expectations that are characteristic of behavioral engagement.

Instead, it entails notions of embracing academic challenges, exceeding standards or requirements, and making vested psychological investments in mastering knowledge and skills (Connell & Wellborn, 1991; Newmann, Wehlage, & Lamborn, 1992). Moreover, cognitive engagement includes an element of approaching education-related objectives strategically. An active management of the learning process and meta-cognitive planning and self-regulation strategies are integral characteristics of cognitive engagement (Pintrich & De Groot, 1990).

Motivation. Motivation directly affects academic achievement because it drives a student's learning effort (Hallinan, 2004). A greater involvement in, and control of, the learning process is associated with increases in motivation (Hidi & Harackiewicz, 2000; Wallace, 1996).The construct of motivation is generally divided into extrinsic motivation, which entails external rewards and punishment, and intrinsic motivation, whereby students are motivated to achieve through the learning environment itself (Ryan & Deci, 1996). Cognitive engagement, for instance, draws from intrinsic motivation, which leads students to adopt learning rather than performance goals and results in a sustained commitment to self-regulated learning (Ames, 1992; Ames & Archer, 1988; Vauras, Rauhanummi, Kinnunen, & Lepola, 1999). Particular gains in student motivation can be achieved when schooling can stimulate higher-order thinking, provides depth of knowledge, and is perceived as important beyond classroom walls (Newmann & Wehlage, 1993).

Student Engagement and Curriculum

The nature of instruction is considered "the proximal and most powerful factor in student engagement and learning" (National Research Council and Institute of Medicine, 2004, p. 60). Instruction is often used as a synonym for curriculum, whereby the latter specifically refers to “the attainment of learning objectives” (Laska, 1984, p. 212). Engaging curricula are expected to provide interactive learning opportunities, elicit positive emotional responses, and foster psychological investments in learning. However, most students who drop out of school cite aversion toward school and related disengagement as the principal reason for dropping out (Berktold, Geis, & Kaufman, 1998). The curriculum is directly linked to this disengagement process (Tyler & Lofstrom, 2009). The theory-driven nature of the general-academic curriculum is considered by many students as irrelevant and intensifies the disengagement process (Kelly & Price, 2009). For those students, career-oriented high school programs, which are inherently context-based and application-focused (Advisory Committee for the National Assessment of Vocational Education, 2003), may represent an effective alternative curriculum. Through their application-focused nature, CTE curricula may engage students by offering numerous opportunities for interactive learning (behavioral engagement) through which affective reactions to the learning environment can be elicited (emotional engagement), thus enhancing the willingness to invest in learning and embrace academic challenges (cognitive engagement). Overall, the higher engagement potential CTE can offer to some students may produce more positive educational outcomes, including secondary and postsecondary educational attainment.

Educational Attainment

Educational attainment is broadly defined as “the highest grade completed within the most advanced level attended in the educational system of the country where the education was received” (United Nations, 2007, p. 176). Notions of educational attainment as an essential determinant of economic growth emanate from human capital theory.

Human Capital Theory

Lucas (1988) outlined the fundamental nature of human capital theory by providing the following definition:

> By an individual's 'human capital' I will mean, for the purposes of this section, simply his [sic] general skill level, so that a worker with human capital *h(t)* is the productive equivalent of two workers with ½*h(t)* each, or a half-time worker with *2h(t)*. The theory of human capital focuses on the fact that the way an individual allocates his time over various activities in the current period affects his productivity, or his *h(t)* level, in future periods. (p. 17)

Human capital theory holds that individuals invest in education and training to acquire knowledge, skills, and experience that function as assets in the labor market (Becker, 1964; Schultz, 1963). Educational assets are viewed in a manner similar to capital assets and have the objective to increase an individual's labor market value through enhanced productivity (Olaniyan & Okemakinde, 2008). Higher levels of educational attainment represent a more valuable means of production that enables individuals to yield higher rates of return in terms of wages or salaries (Becker, 1993). At the macroeconomic level, higher educational attainment then translates into stronger national economic growth (Psacharopoulos & Woodhall, 1997; Schultz, 1971). While human capital theory has been met with criticism for disregarding the intrinsic importance of education that encompasses social, religious, moral, and emotional dimensions (Robeyns, 2006), the idea of productivity and growth effects being based on educational human capital investments has been empirically validated (see Barro, 1991; Gemmel, 1996).

Human capital theory is inherently linked to theories of economic growth. Two major theoretical approaches to economic growth include the *neoclassical* and *endogenous* models. Developed by Solow (1957), neoclassical growth theory is based on an economic aggregate production function to explain the combined impact of technological development, labor, and capital on macroeconomic output, or growth. Solow's approach represented a new way of "segregating variations in output per head due to technical change from those due to changes in the availability of capital per head" (p. 312). While groundbreaking at the time, one limitation of the Solow model was that it did not consider the effects of education on economic growth. This shortcoming was remedied by Mankiw, Romer, and Weil's (1992) augmentation of the Solow model, whereby education was considered an important factor of production directly linked to gross domestic product growth per worker (OECD, 2010).

Although the augmented neoclassical growth model accounts for static effects of education as a factor of production, it fails to explicitly link investment into human capital to technological progress. This shortcoming prompted the development of the endogenous growth model (Lucas, 1988; Romer, 1990) as the second major theory of economic growth. Romer focused on the growth impact of technological change induced by investment in human capital. According to him, growth due to technological innovations is considered endogenous (i.e., stemming from innovation through human capital investment) because individuals intentionally seek education and training in order to respond to labor market incentives such as employment opportunities and higher wages. Lucas further highlighted the critical effects of human capital accumulation through schooling as well as on-the-job training on economic growth. His work corroborated prior notions about the strength of educational attainment effects being proportional to the extent to which a nation is technological advanced (see Nelson & Phelps, 1966). Since education is instrumental in fostering the creation and diffusion of technological change (Benhabib & Spiegel, 1994, 2005), investment in higher levels of educational attainment contributes to higher levels of economic growth (Jorgenson & Fraumeni, 1992). Different types of high school curriculum can be considered different types of human capital investment that may lead to variations in educational attainment and resulting economic effects. These economic effects are illustrated in the next section.

Economic Effects of Educational Attainment

Educational attainment in the U.S. is directly linked to economic growth and technological progress (National Research Council, 2001; OECD, 2003), since higher educational attainment leads to increases in labor quality and productivity gains (Jorgenson, Ho, & Stiroh, 2003). During 1958 and 1999, such attainment-driven increases in labor productivity contributed .3 percent to the average annual economic growth of 3.4 percent during that same period (Jorgenson, Ho, & Stiroh, 2002). Besides strict labor-based productivity increases, educational attainment produces ancillary growth effects by increasing the use and demand for technological innovations (Acemoglu, 1998). Attainment-driven demand for information technologies contributed .8 percent to annual economic growth during the 1990s (Oliner & Sichel, 2000).

Assessing the current state of educational attainment in the U.S. is a complex endeavor that entails considerations of both quantity and quality of education. Nominal increases in the quantity of educational attainment have been impressive. Since 1940, the U.S. population has experienced a three-fold increase in high school attainment, accompanied by a five-fold increase in college attainment rates (Crissey, 2009). Despite this substantial boost in nominal secondary and postsecondary educational attainment over the past several decades, actual attainment rates have not seen sizeable improvements since the early 1980s (Ho & Jorgensen, 1995). The mismatch between nominal and actual attainment growth emanates from the fact that "much of the increase in schooling since the 1970s is due to the dying out of older generations with comparatively little education, rather than steadily growing educational attainment among younger generations" (Kodrzycki, 2002, p. 39). While the U.S. remains in a leading position regarding the average quantity of secondary schooling (Barro & Lee, 2001), growth in college attainment rates has been projected to be minimal for the period until 2020 (Ellwood, 2001).

The quality of education, generally measured in terms of student achievement on standardized tests (Kodrzycki, 2002), is another important aspect of educational attainment. In fact, the quality of educational input is more strongly associated with economic growth than the average years of education received by a given population (Barro, 2001; Hanushek & Woessmann, 2009). This fact is reflected in the performance of U.S. secondary students on large-scale international comparative assessments of student achievement, which have become important indicators of international prestige and competitiveness (Baker & Wiseman, 2005; Kamens & McNeely, 2009). The significance attributed to tests such as the Program for International Student Achievement (PISA; OECD, 2007a) and similar assessments is deeply rooted in human capital theory, whereby a well-educated workforce is the gateway to economic success in a globally competitive market environment. Student achievement on such tests, particularly in science and mathematics, is viewed as a predictor of individual productivity, remedy against unemployment, and general indicator of an education system's ability to produce workers whose skills adhere to globally-accepted standards (Atkin & Black, 1997; National Center on Education and the Economy, 2007).

When public education expenditures are considered vital human capital investments into a country's future economic well-being, accountability becomes a critical element of the education policy process (Hopmann, 2008). The provision of accountability is a major function of large-scale international tests (Abu-Alhija, 2007), which is why policymakers often use them as proxy measures of returns on educational investments. These returns hinge upon the quality much rather than the quantity of education. An important recent study (OECD, 2010) examined the economic growth potential of PISA performance improvements and corroborated the direct relationship between the quality of educational attainment and economic growth from an internationally comparative perspective. Raising the PISA scores of students in medium and low performing countries to levels achieved by students in high performing countries would yield substantial gross domestic product increases based on gains in labor productivity. In projections covering the period between the present and the year 2090, the study quantified the economic impact of such education reforms. The authors determined that "bringing all countries up to the average performance of Finland, OECD's best performing education system in PISA, would result in gains in the order of

USD 260 trillion" (p. 6). By far the largest education reform-based gross domestic product gains would be realized in the U.S. (see Figure 2.1).

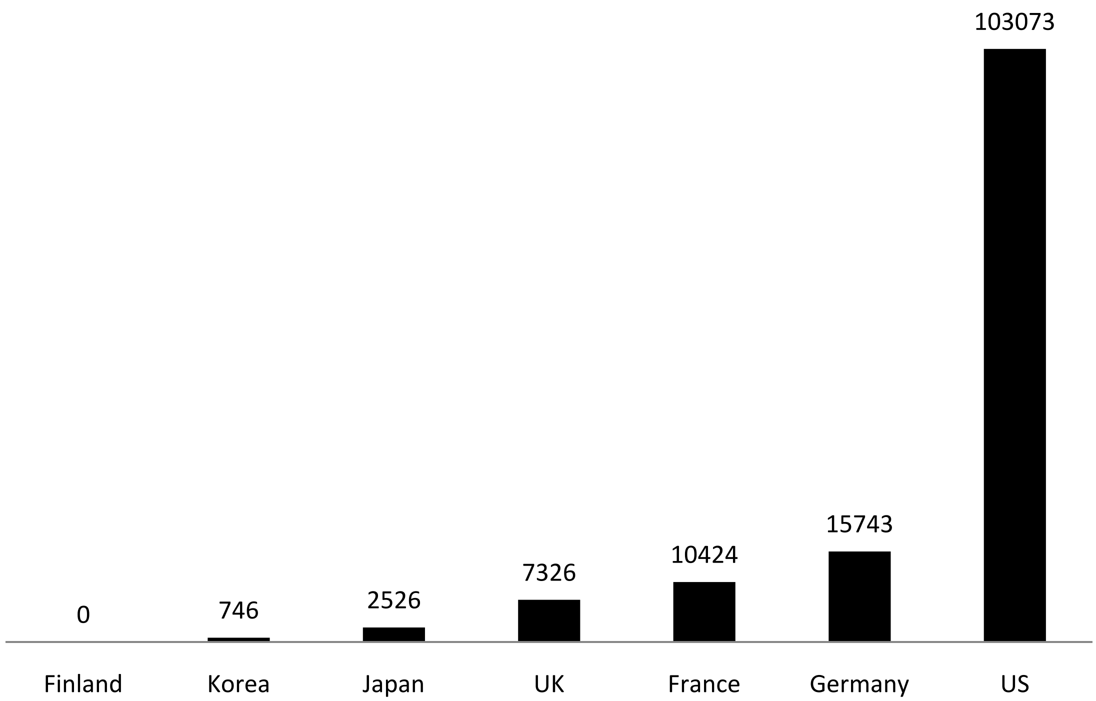

Figure 2.1. Economic growth impact of improvements in qualitative educational attainment (in billion USD; adapted from OECD, 2010). The figure shows the discounted value of future increases in gross domestic product for the period between the present and the year 2090 as a result of education reforms that would improve student performance in each PISA participating country to the level achieved by students in Finland.

Gender, Race/Ethnicity, and Societal Effects of Educational Attainment

The relationship between educational attainment and economic indicators such as wages at the individual level has been empirically established (see Angrist & Krueger, 1991; Card, 1999), However, when examining the educational wage premium for individuals with higher educational attainment the effects of gender and race/ethnicity need to be considered. Longstanding gender discrepancies on educational attainment have been reversed, whereby females now consistently outperform males at both secondary and postsecondary levels (Freeman, 2004). Yet, despite higher average educational attainment levels, college-educated women experience a five percent earnings gap when compared to their male counterparts in the first year after graduation; this gap widens to 12 percent 10 years after graduation (Goldberg Dey & Hill, 2007).

In terms of race/ethnicity, educational attainment quality outcomes continue to be stratified along familiar patterns, with White students continuing to outperform African American and Hispanic peers on the National Assessment of Educational Progress (Rampey, Dion, & Donahue, 2009; Vanneman, Hamilton, Baldwin Anderson, & Rahman, 2009). The skewed distribution of educational attainment along racial/ethnic lines had direct effects on the educational wage premium during the 1980s and 1990s, whereby African Americans and Hispanics earned significantly less than Whites (Bradbury, 2002). However, while absolute income differences persist, labor market returns on educational attainment for African Americans and Hispanics have recently exceeded those for other racial groups in the U.S. (Long, 2009).

Besides producing direct economic benefits, educational attainment is associated with ancillary outcomes that affect a nation’s social outcomes. One of the major positive social effects of educational attainment is improved health status (Kenkel, 1991). Access to health services, more favorable health-related practices, and a better ability to cope with stress have been determined as the principal intermediate factors besides economic effects that link educational attainment to better overall health status (Hammond, 2003). These health benefits of education have further been found to extend to children and other family members directly (Currie & Moretti, 2003; Wolfe & Zuvekas, 1997), as well as in terms of reduced teenage pregnancies (An, Haveman, &Wolfe, 1993). Beyond private social effects, educational attainment produces beneficial public social effects such as reduced welfare dependency (An et al., 1993) and reduced crime rates (Yamada, Yamada, & Kang, 1991). Educational attainment has been found to decrease recidivism (Nuttall, Hollmen, & Staley, 2003), and the magnitude of its effects on reducing crime exceed that of the effects of higher income (Witte & Tauchen, 1994).

Overall, it is clear that higher levels of educational attainment can be causally linked to productivity, labor market, and broader social benefits, although gender and race/ethnicity have differential effects on the extent to which individuals can capitalize on such attainment gains. The following section considers educational attainment based on high school curriculum type.

Curriculum Effects on Educational Attainment

Secondary Educational Attainment

Secondary educational attainment entails (a) graduation with a diploma from a regular high school program, and (b) alternative completion of a GED.

Regular high school diploma. Secondary educational attainment is directly tied to issues of student persistence and dropout. The National Education Summit in 1989 (see Vinovskis, 1999) proclaimed the reduction of dropout rates as a major education policy focus. Four distinct types of dropout rates, including the event dropout rate, status dropout rate, status completion rate, and averaged freshman graduation rate are commonly reported. The event dropout rate comprises those students who leave school each year without completing a high school program, whereas the status dropout rate indicates the percentage of young adults between the ages of 16 to 24 who are out of school and who have not earned a high school credential (Schargel & Smink, 2001). The status completion rate comprises those individuals who are not in high school and who have earned a high school diploma or equivalency credential, irrespective of when the credential was earned (Laird, Cataldi, KewalRamani, & Chapman, 2008). Finally, the averaged freshman graduation rate consists of public high school freshmen who graduate with a regular diploma four years after starting ninth grade (Laird et al.). A summary of the different dropout measures is provided in Table 2.1.

Table 2.1
Dropout and School Completion Measures (taken from Tyler & Lofstrom, 2009)

Rate	Age group	Description	Data source	GED status
Event dropout rate	15-24	Percentage of high school students who dropped out of grades 10-12	Current Population Survey	GEDs do not count as dropouts
Status dropout rate	16-24	Percentage of people who are not enrolled in high school and who do not have a high school credential	Current population survey	GEDs do not count as dropouts
Status completion rate	18-24	Percentage of young adults who have left high school and who hold a high school credential	Current population survey	GEDs are counted as having a high school credential
Averaged freshman (ninth-grade) graduation rate	NA	Percentage of public high school students who graduate with a regular diploma four years after starting ninth grade	Common core of data	GEDs are not counted as graduates

Increasing the rate of student persistence has remained an essential concern given the negative personal and societal repercussions associated with high school dropout. Personal repercussions of dropping out entail higher unemployment and lower earnings (Wirt et al., 1998), whereas societal repercussions comprise reduced tax revenue, higher welfare costs, and increased crime rates (Tyler & Lofstrom, 2009). It is, therefore, not surprising that numerous investigations have been carried out that explore the impact of curriculum type on high school completion.

Ample evidence supports the positive impact of college-preparatory curricula on secondary academic achievement and completion rates. While somewhat dated, Gamoran and Mare (1989) conducted a sophisticated analysis of High School and Beyond data to determine curriculum impact on mathematics achievement and high school completion. The underlying purpose of the study was to examine subliminal tracking effects on educational stratification, learning outcomes, and student persistence. In an attempt to scrutinize productivity arguments that are commonly used as justifications for curriculum tracking, the authors cautioned that "stratification in schools may produce higher achievement overall, but increased productivity may be mainly due to higher outcomes among high-track students, so that inequality rises as well" (p. 1149). Findings clearly demonstrated that students in college-preparatory curricula exhibited substantially higher mathematics achievement and high school completion rates when compared to their lower track peers. The authors concluded that enrolling all students in a college-preparatory high school curriculum would yield higher average rates of achievement and graduation for all students. Other studies (Broussard & Joseph, 1998; Lee, Burkam, Chow-Hoy, Smerdon, & Geverdt, 1998; Lee, Croninger, & Smith, 1997) have corroborated this conclusion. It needs to be emphasized, however, that this conclusion was based on data that reflected an era before the implementation of groundbreaking CTE reforms.

Results from studies evaluating the impact of CTE curricula on high school completion have yielded more heterogeneous results. Several studies have shown that CTE curricula have a positive impact on high school completion rates (Kulik, 1998), academic achievement (Stone & Aliaga, 2005), and aspirations toward pursuing two-year postsecondary programs (Rojewski, 1997). However, other investigations have either failed to ascertain beneficial effects of CTE concentrations on high school completion (Agodini & Deke, 2004; Pittman, 1991), or have discovered detrimental effects in terms of a reduced likelihood of college attendance (Arum & Shavit, 1995). A third stream of research has found that, when compared to general-track students, an integrated curriculum of CTE and college-preparatory courses can foster positive outcomes, including higher rates of secondary and postsecondary attainment (Castellano, Stringfield, & Stone, 2003; Fletcher, 2009; Plank, 2001; Plank, DeLuca, & Estacion, 2008).

Rasinski and Pedlow (1994) conducted a frequently cited study on the effects of CTE course taking on students' propensity to drop out of high school. Logistic regression techniques were applied to a sample from NELS:88 to determine the effects of CTE course-taking during grades nine and 10 on dropout in grades 11 or 12. Findings suggested that CTE curricula had an indirect positive impact on student persistence by providing an opportunity to improve grades and, thus, class rank relative to other students. The authors hypothesized that improvement in class rank raised students' self-esteem, helping them in their decision to stay in school. Besides these indirect effects, direct positive effects of CTE course-taking on reducing dropout were reported for vocational concentrations in agriculture and technical/communications courses.

Rasinski and Pedlow's (1994) study was limited by two factors. First, the inclusion of achievement and class rank as covariates in the logistic regression model resulted in a loss of one-third of the initial analysis sample due to high rates of missing data on those two variables. Given the magnitude of the missing data problem, it is likely that listwise deletion may have unduly biased analysis results. Second, the fact that the authors only considered vocational course-taking during the first two years of high school seems problematic given that a lot of vocational course-taking occurs in grades 11 and 12. Overall, however, the authors demonstrated the ability of CTE curricula to diminish dropout rates either indirectly or, as in the case of agriculture and technical/communications courses, directly.

The positive impact of CTE on dropout proposed by Rasinski and Pedlow (1994) was not ascertained in other studies. Notably, Pittman (1991) used a path-analytic approach to examine data from the High School and Beyond study to contrast the influence of CTE curriculum effects versus those of social factors in students' dropout decision. Pittman uncovered an intricate web of relationships that affect the dropout decision. Within this web, curriculum type had the weakest relationship with dropout status, whereas the effect of school environmental factors such as peer effects and the general school climate had a more pronounced influence. Pittman concluded that while CTE may have some influence in raising students' perceived utility of school, the emphasis in enhancing high school completion rates should be on improving factors related to strengthening student-teacher relationships and making sure that at-risk students become an integral part of the school environment.

Agodini and Deke (2004) conducted a focused investigation to determine the relationship between high school CTE enrollment and dropping out. The underlying purpose of their study was to verify claims about CTE being perceived as more engaging and relevant by non-college-bound and low achieving students who are usually deemed at risk of dropping. Using data from the National Education Longitudinal Study of 1988

(NELS:88), the authors conducted a two-step analysis in which regression analysis was used to model the decision to enroll in CTE as opposed to a basic academic curriculum while controlling for potential confounders. In a subsequent step, average probabilities of dropping out were calculated based on track placement. The authors determined no difference in dropout rates between CTE and basic academic students, even when subgroups were examined in terms of postsecondary aspirations, academic achievement, socioeconomic status, and school difficulty. The only discernable curriculum effect for CTE students was when they were concentrators with a heavy vocational course load as opposed to explorers. Overall, results from Agodini and Deke corroborated those from two important randomized-control studies that found no effects of career academies on dropout rates (Kemple & Scott-Clayton, 2004), and even negative effects of career magnet programs on-time graduation (Crain et al., 1999).

The strength of Agodini and Deke's (2004) study lies in its strong research design that uses regression approaches effectively to control for the influence of confounding variables on curriculum effects. Moreover, their design takes into consideration the timing of when and how many CTE credits were taken and directly relates this information to the point in their high school career when students actually dropped out. One limiting factor of the study was its use of an older data source. In fact, while results reflect the state of CTE during the late 1980s and early 1990s, the effects of more recent Carl D. Perkins legislation with its focus on the integration of CTE and academic curricula could not be considered due to the use of NELS88.

A notable study by Plank et al. (2008) recently provided interesting new evidence about the effects of secondary CTE on student dropout. The authors conducted a survival analysis of NLSY97 data with a specific focus on CTE-to-academic course-taking ratios. Hazards models were used to simulate the risk and timing of dropping out based on course choice. As opposed to Agodini and Deke's (2004) findings that associated stronger CTE concentration with reduced risks of high school dropout, Plank et al. found an intriguing curvilinear effect, whereby a 2:1 ratio of core academic-to-CTE courses was associated with minimizing the risk of dropping out. In contrast, both lower and higher ratios of CTE coursework produced an increased drop out risk. The study corroborates earlier research (Plank, 2001) that had hinted at a similar curvilinear relationship. Plank et al.'s study has important implications for policy in that it provides evidence supporting the positive impact of balanced CTE curricula on secondary educational attainment. This substantiates the importance of Perkins 1998 and 2006 legislation, which fostered dual concentrations and similar measures that integrated CTE and core academic coursework.

GED. Although a rich body of knowledge exists regarding labor market and postsecondary education outcomes of GED holders, research that explores secondary curriculum effects on GED attainment are virtually nonexistent. Therefore, this section provides a general overview of literature on the GED.

The GED is a high school equivalency credential that is obtained via a test consisting of 240 multiple choice items and one essay that covers knowledge in five content areas, including writing, social studies, science, reading, and mathematics (American Council on Education, 2010). The GED was first offered in 1942 as an option for World War II veterans to complete their secondary education (Meeker, 2008). In the late 1960s, the GED was made accessible to the general population, resulting in a rapid growth of the credential (Ou, 2008). Boesel, Alsalam, and Smith (1998) identified four important social functions of the GED. First, the GED is considered a stimulus to human capital investment that offers dropouts the certification of work-relevant skills in return for the time and effort expended for test preparation. Second, by the simple fact that dropouts have to actively pursue and study for a GED, the diploma functions as a sorting mechanism that offers dropouts who are more motivated potential labor market benefits. Third, as a recognized certification of cognitive skills, the GED provides pathways to employment and postsecondary education that may not otherwise be available to dropouts. Finally, the GED can function as a confidence-builder that increases self-esteem and eventually improve the life circumstances of those who obtain it.

Although a GED formally represents an alternative pathway to completing secondary education, it should not be treated as an equivalent educational attainment category for practical purposes, since GED holders achieve lower average rates of employment and income (Heckman, & LaFontaine, 2006; Sum, 1996; Tyler, 2003), exhibit lower rates of postsecondary enrollment (Tyler & Lofstrom, 2008), and are less likely to complete a degree if they do enroll in postsecondary education or training (Cameron & Heckman, 1993). Moreover, the availability of GEDs as an alternative route to completing secondary education has been found to increase dropout rates, especially in light of more stringent high school graduation requirements (Chaplin, 1999; Lillard & DeCicca, 2001). Heckman, LaFontaine, and Rodriguez (2008) found that decreasing GED pass rates by six percent caused a 1.3 percent decline in dropout rates, with large effect sizes for both older (i.e., old for grade) and minority students. However, when compared to all dropouts, those who obtain a GED generally experience more positive labor market and postsecondary outcomes (Garet, Jing, & Kutner, 1996; Kroll & Qui, 1995; Murnane, Willett, & Boudett, 1995; Tyler, Murnane, & Willett, 2000). With regard to earnings, effects are particularly beneficial for female dropouts with weak basic mathematics skills (Tyler, Murnane, & Willett, 2003) and foreign-born men who completed their secondary education outside of the U.S. (Clark & Jaeger, 2002).

Dropouts, GED recipients, and high school graduates exhibit differences in life outcomes even when controlling for sociodemographic factors and early cognitive skills. Ou (2008) analyzed data from the Chicago Longitudinal Study (a panel of low-income minority youths from inner city Chicago) to examine GED effects on adult well-being, as expressed by differences in income, crime, teenage pregnancy, life satisfaction/optimism, and substance abuse. While no differences were found between dropouts and GED recipients in terms of crime and teenage pregnancy, outcomes for wages, life satisfaction/optimism, and substance abuse exhibited three stratified levels with outcomes mirroring the different levels of education.

The literature demonstrates that GED holders have more positive labor market, postsecondary education, and adult well-being outcomes when compared to other dropouts. The GED appears to fulfill its principal social functions (Boesel et al., 1998), as outlined above. A GED is, however, clearly less desirable than a regular high school diploma with regard to almost all personal and societal dimensions. Overall, the absence of literature that directly examines secondary curriculum effects on GED attainment points represents a gap that needs to be addressed.

Postsecondary Educational Attainment

While considerable research has been conducted to determine regular high school curriculum effects on student persistence and high school completion, a remarkable dearth exists in the literature regarding effects on postsecondary education. Most studies consider either actual or *desired* enrollment in postsecondary education (e.g., DeLuca, Plank, & Estacion, 2006; Plank, 2001), yet do not take a longitudinal approach in which *actual* postsecondary attainment is examined. Recently, three studies have focused on the latter aspect. One important such study (Silverberg et al., 2004) consolidated results from the literature to assemble a comprehensive account of the state of CTE prior to the 2006 reauthorization of Perkins IV, whereas two investigations (Fletcher, 2009; Novel, 2009) analyzed data from NLSY97 to determine the likelihood of postsecondary enrollment and completion based on curriculum type.

Perhaps the most comprehensive report on CTE curriculum effects with a focus on postsecondary education was the congressionally-mandated National Assessment of Vocational Education (Silverberg et al., 2004). The three-year study had the objective to assemble a comprehensive assessment of the state of CTE after the passage of Perkins III in 1998 in order to guide Congressional decision-making about the upcoming reauthorization of Perkins IV. Perkins III had a strong focus on fostering CTE students' transition from school to postsecondary education and training. In fact, enrollment in, retention in, and completion of postsecondary education and training were considered key measures of CTE program success. The report found that the Perkins III-based integration of occupational and academic coursework effectively increased CTE students' academic achievement and preparation for college. Postsecondary enrollment rates of CTE students increased at a considerably higher rate than those of students in general/unspecified curricula. A further interesting finding was reported regarding the timing of enrollment in postsecondary education. Short-term postsecondary enrollment effects are often negative because many CTE students take advantage of their higher earnings potential right after high school. However, short-term negative postsecondary enrollment effects disappear as more CTE students eventually do enroll in postsecondary education or training programs. Nonetheless, the authors pointed at statistics provided by Agodini, Uhl, and Novak (2002), whereby only 53 percent of CTE concentrators had earned a postsecondary degree or certificate eight years after high school graduation, compared to 66 percent of non-CTE students. The authors concluded that while the integration-focused legislative approach of Perkins III produced overall positive effects on CTE student outcomes, “secondary vocational education itself is not likely to be a widely effective strategy for improving academic achievement or college attendance without substantial modifications to policy, curriculum, and teacher training” (p. xviii).

Novel (2009) studied the likelihood of postsecondary enrollment and attainment based on participation in an enhanced CTE+ high school curriculum. As such, her investigation sought to evaluate whether Perkins 1990 and 1998 legislation was implemented effectively. This legislation had the underlying premise to prepare career-technical students for both the workplace *and* postsecondary education by enabling students to combine a career-technical program of study with a concentration in college-preparatory courses. Using logistic regression with NLSY97 data, Novel determined that students in CTE+ programs had indeed a significantly higher probability of enrolling in and completing a four-year college degree, alongside their college-preparatory peers. Of the three high school curricula under investigation (i.e., CTE, CTE+, academic/general), no high school curriculum type emerged as a positive predictor of two-year college completion. When combining two-year and four-year college attainment, the study found pure CTE concentrators to be outperformed by both CTE+ and academic/general in terms of the likelihood to complete any level of postsecondary educational attainment. Novel concluded that findings corroborated the importance of enhancing traditional career-technical curricula with rigorous sequence of academic courses in order to drastically increase postsecondary transition success. This outcome was deemed particularly positive with regard to the 2006 re-authorization of Perkins legislation, which further intensified efforts to increase postsecondary attainment levels among CTE students by promoting integrated CTE+ curricula.

Fletcher (2009) conducted a study similar to Novel (2009) by examining the relationship between high school curriculum type, degree attainment, and occupational earnings. His study was also highly similar to this present investigation in that he analyzed the same dataset (i.e., NLSY97) and used the same educational attainment variable. Using a multinomial logistic regression approach to analyze secondary and postsecondary attainment, Fletcher found that students enrolled in dual, or CTE+, curriculum concentrations exhibited a significantly higher likelihood to earn a two-year college degree when compared to general-track students. General-track students had the worst overall attainment outcomes, in that they were less likely than any group (i.e., CTE, dual-track, college-preparatory) to earn degrees at either the secondary or postsecondary levels. College-preparatory students, by contrast, exhibited the best degree attainment outcomes, particularly at the four-year college degree level.

Fletcher's (2009) study represents a commendable effort to examine high school curriculum effects on educational attainment, yet his investigation contains several methodological shortcomings, some of which severely qualify the value of his findings.

First, the author used curriculum classifications based on self-report data. While other studies (e.g., Stone & Aliaga, 2005) rely on self-report data to form curriculum/treatment groups, self-report information is considered less accurate than transcript information (National Center for Education Statistics, 2009), which is available in the NLSY97 dataset. Second, Fletcher's study is based on a weak selection model for curriculum group assignment, for it solely considers gender, race/ethnicity, mother's highest grade completed, father's highest grade completed, and household income as predictors for curriculum placement. Even though the importance of these predictors is beyond dispute, Fletcher rightfully pointed at the exclusion of other important predictors as an important limitation. Third, the issue of missing data was not treated in a principled manner. Instead, the author listwise deleted observations with missing values, which greatly increases the possibility of biased results. Finally, the study demonstrated an erroneous understanding of selection bias. Fletcher controlled for selection bias by randomly sampling participants from within the NLSY97 dataset. Such simple random selection is inadequate, for the resulting sample is anything but random due to the complex sampling and design effects inherent in large-scale observational datasets such as NLSY97. Overall, this accumulation of methodological limitations renders the robustness of Fletcher's findings questionable. The present investigation addressed these shortcomings in a more principled manner.

Treatment of Missing Data

Missing data are a common occurrence in a variety of empirical research contexts (Downey & King, 1998). Missing data due to nonresponse can occur because of noncontact, refusal to cooperate, or specific barriers that impede an eligible respondent from participating (Groves & Couper, 1998). Survey researchers generally distinguish between unit nonresponse and item nonresponse. The former refers to the absence of any sort of data from an eligible respondent due to noncontact or outright refusal to participate, while the latter denotes a situation in which a respondent answers some items but fails to answer others (Elliott, Edwards, Angeles, Hambarsoomians, & Hays, 2005). Wave nonresponse occurs in longitudinal surveys where participants' responses may be missing for one or more survey waves (Kalton, 1986). In experimental studies missing data may occur due to attrition, meaning that a participant decides to drop out before data collection has been completed (Given, Keilman, Collins, & Given, 1990). Finally, erroneous data entry, disclosure restrictions, and similar procedural factors can lead to incomplete data (Tsikriktsis, 2005).

The missing values that emanate from these and other scenarios routinely obstruct data analysis because most statistical procedures require a complete data matrix (Schafer, 1997). Incomplete data can result in (a) reduced statistical efficiency or power, (b) difficulties in data analytic procedures using standard software packages, and (c) biased analysis results due to the potential existence of systematic differences between missing and observed data (Barnard & Meng, 1999; Roth, Switzer, & Switzer, 1999). From a research design perspective, missing data can have consequences for construct validity because incomplete information reduces the accuracy with which a construct can be measured (McKnight, McKnight, Sidani, & Figueredo, 2007). The detrimental effects caused by missing data are particularly challenging within the context of survey research due to the sizeable number of responses and respondents involved (Raaijmakers, 1999).

Consequently, incomplete data represent a dilemma in the analysis of survey-based large-scale datasets.

Historically, cases with missing values were either ignored or the missing observations were substituted with imprecise approximations of the missing data points based on simplistic replacement procedures. The statistical costs incurred by such unsophisticated approaches were frequently prohibitive in terms of case loss and/or analysis bias. To remedy the detrimental effects of such crude methods, Dempster, Laird, and Rubin (1977) developed the expectation maximization (EM) algorithm, whereby a likelihood function is used to draw parameter estimates from a particular distribution that is assumed to underlie the missing data. Based on Rubin's (1976) coherent framework of inference from incomplete data, EM was the first modern-day *stochastic* (i.e., considering randomness) missing data technique (Schafer & Graham, 2002). A decade after introducing EM, Rubin (1987) developed the multiple imputation (MI) method that is based on the creation of several complete datasets in which the missing values are replaced by different random draws from a distribution of plausible values. By analyzing each complete dataset separately before pooling parameter estimates, MI is able to better incorporate the uncertainty inherent in the missing data, thus producing more robust parameter estimates (Enders, 2001; Schafer, 1999a).

This brief historic overview of handling missing data illustrates a progression from applying naïve approaches to more principled ones that incorporate the randomness reflected in the missing data. This progression has been supported by the general proliferation of computing power and the widespread incorporation of advanced missing data methods in standard statistical software. Before revisiting various missing data methods in more detail, it is important to understand the nature of missing data patterns and mechanisms.

Missing Data Patterns and Mechanisms

Missing data can occur in random or nonrandom patterns within a data matrix. Methodologists differentiate between three types of patterns, including univariate, monotone, and arbitrary patterns. *Univariate* patterns occur when a specific variable contains missing values, while all other variables in the dataset are fully observed. Likewise, univariate patterns include situations in which a group of variables exists whose values for a given case or individual are either entirely complete or entirely missing (Schafer & Graham, 2002). *Monotone* patterns occur when individuals decide to drop out from a study/survey before formal completion (Fielding, Fayers, & Ramsay, 2009; Minini & Chavance, 2004). Such respondent behavior results in a pattern in which all observations are complete up to the point of dropout, upon which all remaining data are missing. For instance, if within a y_i variable data matrix a particular case or individual has a nonmissing value on variable Y_3, the same case or individual has nonmissing values on all preceding variables Y_1 and Y_2 (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001). If Y_3 is the last variable for which data were collected before dropout, all further variables Y_4 to Y_i will have missing values. Finally, *arbitrary* patterns arise when missing data display no systematic, discernable structure within a given data matrix. This occurs when each case or individual exhibits a different pattern of missing values (McKnight et al., 2007).

Besides different structural patterns, the theoretical literature distinguishes between several response mechanisms that may underlie missing data (Little & Rubin, 1987; Rubin, 1976). Missingness can be classified under three distinct mechanisms that "offer different explanations for how missingness is probabilistically related to the values of variables in the dataset" (Enders, 2006, p. 315). When data are *missing completely at random* (MCAR) the missingness of a value in variable *Z* is unrelated to any other data point within variable *Z* or any other variable in the dataset. MCAR is the most restrictive missing data mechanism and, heuristically, means that missing data occur independent of any other observed or unobserved factors pertinent to a given study (Horton & Lipsitz, 2001). Thus, MCAR is considered a strong assumption that can be difficult to uphold in practice (Little & Rubin, 1987; Muthén, Kaplan, & Hollis, 1987). Many standard statistical software packages use Little's (1988a) MCAR test to determine whether missing data in a given matrix are, in fact, missing completely at random. If Little's chi-square test statistic rejects the null hypothesis of data being MCAR at conventional significance levels, then the assumption of random missingness cannot be upheld. Consequently, the application of missing data techniques that require data to be MCAR would yield biased analysis results.

A less stringent mechanism is known as *missing at random* (MAR), whereby the missingness of a value in variable *Z* is unrelated to any other data point within variable *Z*, but *is* related to one or more of the other variables in the dataset. MAR is also referred to as *ignorable nonresponse* because the probabilities of missingness do not depend on the missing data themselves (Allison, 2002). Under MAR, results based on data from respondents with complete data are generalizable to those with missing data, since cases or individuals with missing data on a given variable differ only by chance from those with complete data (Tsikriktsis, 2005).

The third mechanism, *missing not at random* (MNAR), denotes a situation in which the missingness of a value in variable *Z* is a function of other values in variable *Z*. Data that are MNAR represent *nonignorable nonresponse*, which greatly complicates the missing data problem because a model for the distribution of the missingness needs to be specified in addition to a model for the complete dataset (Schafer & Graham, 2002). No straightforward statistical remedy is available to deal with situations of nonignorable nonresponse, and results based on samples with MNAR observations are usually not generalizable to a wider population (Byrne, 2001). Given the complexities inherent in dealing with data that are MNAR, MAR is a working assumption in the application of modern imputation methods (Enders, 2006). Figure 2.2 illustrates the different missing data mechanisms for data that are MCAR, MAR, and MNAR.

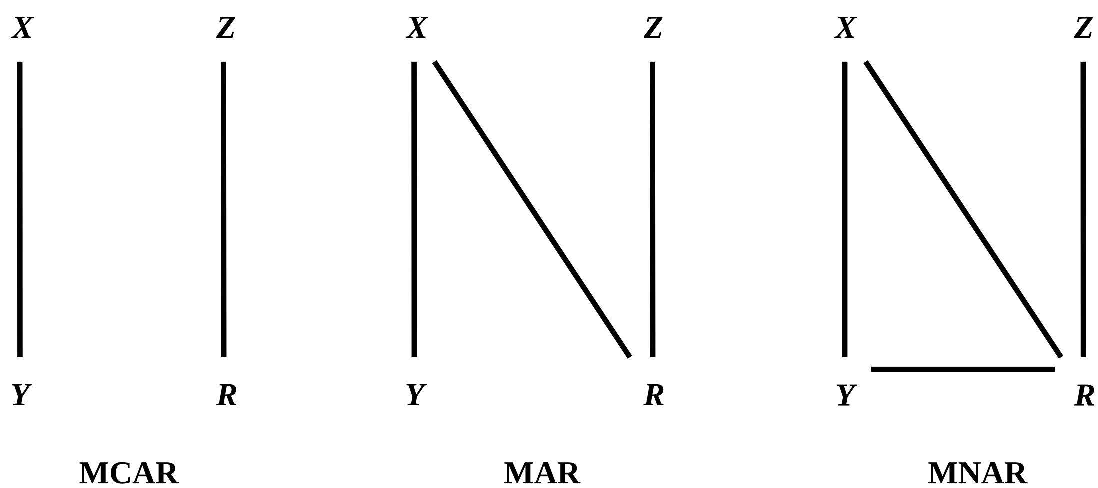

Figure 2.2. Missing data mechanisms (taken from Schafer & Graham, 2002). *X* represents variables that are completely observed, *Y* represents a variable that is partly missing, *Z* represents the component of the causes of missingness unrelated to *X* and *Y*, and *R* represents the missingness.

Traditional Missing Data Methods

The taxonomy of missing data mechanisms is directly related to the statistical options available to researchers. Traditional options include complete case analysis, complete variables analysis, mean substitution, regression-based imputation, and hot-deck imputation (Enders, 2006). For purposes of clarity, traditional approaches have been conceptually divided here into *case reduction* and *deterministic* techniques.

Case reduction approaches. Complete case and complete variables analysis are based on eliminating the missing data problem through case reduction. Complete case analysis, also referred to as *listwise deletion*, entails simply discarding all cases in a dataset that exhibit missing data on one or more variables (Hill, 2004). This approach is often used by researchers because it can be implemented without any additional computational effort and may be used in conjunction with all sorts of statistical methods (Allison, 2002). Complete variables analysis, also referred to as *pairwise deletion*, is a variable-by-variable approach whereby "only those cases with missing values on a particular bivariate pair are discarded" (Enders & Bandalos, 2001, p. 432). The advantage of pairwise over listwise deletion is that the former restricts the deletion of cases with missing data to situations in which a given statistical analysis procedure would actually make use of a specific variable containing the missing data (Roth, 1994). Whereas complete variable analysis has been shown to yield slightly more accurate results when compared to complete case analysis (Raymond, 1986; Raymond & Roberts, 1987), both methods are generally considered naïve because they discard cases for which information may be at least partially available.

Data reduction approaches are considered adequate in situations in which the amount of missing observations is small. While no general guidelines exist as to the exact meaning of *small*, five percent (Schafer, 1997) and 10 percent (Kline, 1998), respectively, have been suggested as acceptable upper limits. When applied in scenarios with higher rates of missingness, case reduction eliminates important information contained in the original data matrix, which can result in dramatic case loss causing inefficient and biased parameter estimates (Arbuckle, 1996; Graham, Hofer, & MacKinnon, 1996; King, Honaker, Joseph, & Scheve, 2001). Case loss is also directly linked to the issue of statistical power, for sample size is "the only parameter that can be used to design empirical studies with high statistical power" (Verma & Goodale, 1995, p. 144). Multivariate analyses are particularly prone to excessive case loss, given that they typically contain a large set of different variables on which missing data can occur (Schafer & Olsen, 1998). A final limitation is that case reduction approaches are efficient only under the assumption that data be MCAR, which is a strong assumption.

Deterministic approaches. Mean substitution and regression imputation are considered *deterministic* procedures because they replace missing values with a simple fixed estimate of the hypothesized true value (Schulte Nordholt, 1998). The key advantage of deterministic approaches over case reduction methods lies in the preservation of sample size. As implied by its name, mean substitution simply replaces all missing data points in a given variable with one arithmetic mean value of that variable (Enders, 2006). It is no less naïve than case reduction techniques, for replacing missing values with the variable mean value reduces the overall variance in the data (Perez, Dennis, Gil, Rondon, & Lopez, 2002), therefore leading to considerable bias even for data that are MCAR (Huisman, 2000; Little & Rubin, 1987).

Regression imputation is a slightly more refined approach that replaces missing data points with the predicted values from a linear regression model (Little, 1988b). The regression model itself is specified using a set of auxiliary variables from the dataset (Rao, 1996). For regression imputation to work, at least a moderate degree of covariance must exist between the variables with missing data and all other variables within the same data matrix (Kline, 1998). An additional random error term can be added to the regression imputation model in order to introduce additional variance and more realistically model the missing data (Roth et al., 1999). Similar to other deterministic and case reduction approaches, regression imputation requires data to be MCAR (Carter, 2006). While easy to implement, regression imputation has been found to introduce undue bias by artificially inflating correlations between variables (Schafer & Olsen, 1998).

Cold-deck and hot-deck procedures have long been used to deal with missing data in survey research. In contrast to mean substitution and regression imputation, cold-deck and hot-deck procedures do not rely on the creation of synthetic values (Chen & Shao, 2001). The distinction between cold-deck and hot-deck imputation is sometimes unclear. Cold-deck imputation is used in longitudinal surveys that consist of several data collection waves. If a certain case or individual exhibits an observed value on a given variable in a previous survey wave, but a missing value on the same variable in a current survey wave, then the previous wave's observed value is assigned (Chaudhuri & Stenger, 1992). Whereas cold-deck imputation is based on data from *different* datasets on the *same* case, hot-deck imputation uses the actual value from a *different* case or individual in the *same* dataset (Schulte-Nordholt, 1998). In essence, hot deck imputation identifies a case (also referred to as a *donor*) in the same dataset that is similar across all variables to the case containing the missing value and simply replaces the missing observation with the donor's value (Roth, 1994). The use of distance measures in combination with hot-deck imputation, whereby the closest-fitting donor value is identified and used for replacement, has been shown to improve imputation accuracy (Chen & Shao, 2000; Switzer, Roth, & Switzer, 1998). An advantage of hot-deck procedures using similar donors is that the use of nonsensical replacement values is generally avoided (Roth). A disadvantage of hot-deck imputation lies in its tendency to produce biased estimates of correlations and regression weights (Roth & Switzer, 1995), and to produce biased parameter estimates even when data are MCAR (Brown, 1994). Finally, undue bias can arise when no sufficiently similar donor case is available in the dataset

Overall, while deterministic approaches are easy to compute and implement they underestimate the variance of the variable for which values are being imputed, which is problematic when the data are to be used for advanced statistical analysis procedures (Schulte-Nordholt, 1998). Moreover, deterministic missing data methods routinely underestimate parameter standard errors, thus increasing the likelihood for Type I error in any given hypothesis tests. Modern estimation procedures, such as expectation maximization and multiple imputation, remedy these shortcomings.

Modern Missing Data Methods

Modern missing data techniques use stochastic approaches that require the generation of random numbers (Huisman, 2000; Schulte-Nordholt, 1998). These techniques have gained widespread popularity, since they have demonstrated consistently superior estimation properties in terms of parameter bias and efficiency (Enders & Bandalos, 2001; Graham & Schafer, 1999; Muthén et al., 1987; Schafer, 1997; Schafer & Olsen, 1998). Moreover, they are the only methods that yield unbiased parameter estimates when data are MAR (Enders, 2001).The most frequently-used modern missing data techniques include expectation maximization and multiple imputation.

Expectation maximization. Expectation maximization (EM; Dempster et al., 1977) is a maximum-likelihood approach that arrives at missing value estimates through an iterative approximation process (Ruud, 1991). Maximum-likelihood estimation is an iterative procedure that "searches over different possible population values, finally selecting parameter estimates that are most likely (have the "maximum likelihood") to be true, given the sample observations" (Eliason, 1993, p. v). Based on the iterative approach of maximum-likelihood estimation, EM conceptually solves a complex missing data problem by repeatedly solving simpler complete data problems (Schafer, 1997). EM is a two-step process that consists of an *expectation* and a *maximization* step. During the expectation step, the covariance matrix of the available data and resulting parameter estimates are used to determine the conditional expectations of the missing data (Enders, 2006). The maximization step consists of re-calculating parameters using maximum-likelihood estimates based on actual and re-estimated missing data from the expectation step (Little & Rubin, 1987).

Maximum-likelihood methods in general have been found to yield relatively unbiased estimates and small standard errors (Agresti & Finlay, 1997). Simulation studies using EM have found the method to perform very well under a variety of different missing data scenarios (see Graham & Donaldson, 1993; Ibrahim, 1990; Ibrahim, Chen, & Lipsitz, 1999). EM is often used to model the reasons that cause data to be missing (Dempster et al., 1977). One disadvantage of the EM approach is that it is computationally intensive and sensitive to misspecifications of the imputation model. Another drawback inherent in EM is the fact that, with the exception of the *EM by the method of weights* approach (Ibrahim, 1990; Lipsitz & Ibrahim, 1996), its use is limited to the imputation of continuous data. This is a critical issue given that many applied multivariate analyses use categorical data. Multiple imputation can overcome this limitation, for it allows the imputation of both continuous and categorical data points in multivariate scenarios (Schafer, 1997).

Multiple imputation. First notions of multiple imputation (MI) were introduced by Rubin (1978) as a reaction to the nonresponse problem in the analysis of large-scale surveys. Almost a decade later, Rubin (1987) presented a comprehensive framework for the use of multiple imputation (MI) as a highly versatile, general-purpose missing data method. However, it was not until the late 1990s that MI became more widely used based on advances in computational power (Sindharay, Stern, & Russell, 2001). Today, it has been well established that MI provides highly accurate estimates in conditions under which deterministic approaches may yield biased results (Little & Rubin, 1989; Schafer, 1997; Schulte-Nordholt, 1998).

MI is a so-called *Monte Carlo* approach, a general term for computational techniques that generate statistical results by repeating an artificially created chance process using random numbers (Barreto & Howland, 2006; Mooney, 1997). It is based on the creation of $m > 1$ complete datasets that are analyzed individually before pooling discrete parameter estimates and standard errors into one unified set of results (Schafer, 1999a). The replacement of each missing data point with several simulated values is a key characteristic that distinguishes MI from all other methods (Rubin, 1996; Schulte-Nordholt, 1998).

MI has specific advantages over traditional approaches. For one, traditional missing data methods often result in serious underestimation of the true sample variance, as they fail to account for the additional variance that is inherent in the missing data (Rao, 1996). By replacing each missing observation with several slightly different plausible values, MI incorporates the randomness inherent in the missing data, thus mitigating the problem of variance underestimation (Rubin & Schenker, 1986). The randomness created due to the uncertainty about the missing data is translated directly into the width of the confidence interval that accompanies a given parameter estimate (Van Buuren, 2007). Also, unlike other Monte Carlo approaches, MI is able to yield precise missing value estimates without a large number of computation cycles (Schafer & Graham, 2002). Another advantage of MI is its universal applicability to a wide variety of different data analysis contexts (Schafer, 1999a). Once several imputed datasets have been created, researchers can carry out standard statistical analysis procedures without further considering the nature of the imputed data (Sinharay et al., 2001). This property is particularly important for researchers who require a robust solution to a practical missing data problem. While in certain situations EM may outperform MI, Schafer emphasized that in applied research contexts “where missing data are a nuisance rather than a major focus of scientific enquiry, a readily available, approximate solution [such as MI] with good properties can be preferable to one that is more efficient but problem-specific and complicated to implement" (p. 4). The totality of these advantages makes MI a particularly user-friendly choice for imputing missing data in large-scale public-use datasets (Rubin, 1987).

MI is a highly efficient method that requires very few imputation cycles to obtain accurate estimation results (Schafer & Olsen, 1998). The efficiency of an imputation estimate is calculated by using

$$\left(1+\frac{\gamma}{m}\right)^{-1}$$

where γ is the rate of missing information and m denotes the number of complete imputed datasets. For 30 percent missing data, $m = 3$ imputations yield an estimation efficiency of 91 percent. Likewise, increasing the number of imputations to $m = 5$ increases efficiency to over 94 percent. While increasing the number of imputation cycles beyond five or 10 yields marginal efficiency improvements, such increases are expedient only in situations in which the rate of missing data is very high (Schafer, 1997; Van Buuren, Boshuizen, & Knook, 1999).

Whereas until recently MI could only be carried out using specialized software such as MICE (Van Buuren & Oudshoorn, 2000) or NORM (Schafer, 1999b), standard software packages such as SPSS, SAS, or STATA now feature built-in MI capability. Once MI has been carried out using any given specialized or standardized software application, the resulting $m > 1$ complete datasets are analyzed separately using standard statistical methods (Raghunathan et al., 2001). Straightforward arithmetic procedures are then used to obtain final parameter and standard error estimates, which are also referred to as *repeated imputation inferences* (Barnard & Meng, 1999). These arithmetic procedures can be divided into four steps. During the initial step, the $m > 1$ individual parameter estimates are averaged to obtain one pooled parameter estimate. This can be accomplished by using

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m} \hat{Q}_i$$

where m is the number of imputations and $\hat{Q}_i$ is the parameter estimate from the ith imputed dataset. The next three steps focus on calculating the standard error of the parameter estimate. First, the within-imputation variance is calculated by averaging individual standard errors over the $m > 1$ imputations. This can be accomplished by using

$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m} \hat{U}_i$$

where $\hat{U}_i$ is the variance estimate from the ith imputed dataset, and m is the number of imputations. Second, the between-imputation of the parameter estimates is calculated across all imputation cycles. This can be accomplished using

$$B = \frac{1}{m}\sum_{i-1}^{m} \left(\hat{Q}_i - \bar{Q}\right)^2$$

Finally, taking the square root of the total variance of the parameter estimate, or

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B$$

yields the MI standard error for the pooled parameter estimate. Overall, four simple arithmetic procedures produce stable parameter and standard error estimates from imputed data that take into consideration the randomness inherent in the missing data (Rubin, 1987). The approach of actively incorporating uncertainty is the element responsible for MI's superior performance when compared to a variety of other missing data approaches (Schafer, 1999a).

While MI is frequently applied by researchers in medicine (Abraham & Russell, 2004; Barnard & Meng, 1999; Kmetic, Joseph, Berger, & Tenenhouse, 2002; Molenberghs, Burzykowski, Michiels, & Kenward, 1999; Raghunathan, 2004; Zhou, Eckert, & Tierney, 2001), statistics (Allison, 2003; Demirtas, 2004; Olinsky, Chen, & Harlow, 2003), and economics (Brownstone & Valletta, 1996; Kofman & Sharpe, 2003; Schank, Schnabel, & Wagner, 2007), the extant literature in education research features few studies that employ MI to address incomplete data problems (Peugh & Enders, 2004; Smits, Mellenbergh, & Vorst, 2002 are notable exceptions). Even fewer studies in any field of scholarly inquiry have implemented MI in combination with PSM. Based on a simulation study about the combined use of MI and PSM, Hill (2004) concluded that MI (a) outperformed traditional missing data approaches such as complete case and complete variable analyses, and (b) was suitable for a wider range of missing data models and matching methods. Applied research further corroborated positive theoretical findings about the efficiency of combining MI with PSM, particularly in causal-comparative evaluation studies using large-scale observational datasets (see, e.g., Hill, Reiter, & Zanutto, 2007; Hill, Waldfogel, Brooks-Gunn, & Han, 2005). Overall, the combination of MI and PSM represents a dynamic field of ongoing scientific exploration, and applied education researchers may benefit from the use of these advanced methods.

Summary of Missing Data Methods

This section has provided an overview of traditional and modern missing data methods. Generally, these methods represent a progression from naïve approaches, such as case reduction and deterministic methods, to more principled ones, such as EM and MI. While simplistic traditional methods such as mean imputation can be adequate for simple missing data problems (Binder, 1996), stochastic approaches such as EM and MI generally yield better estimation performance. Due to its flexibility and general-purpose nature, MI is particularly well-suited for situations in which propensity score matching is applied to mixed multivariate data. Table 2.2 briefly summarizes the various missing data techniques, and the next section of this chapter provides a detailed description of propensity score matching.

Table 2.2
Summary of Missing Data Methods

Method	*Approach*	*Advantages*	*Disadvantages*
Complete case analysis	Discards all cases that exhibit missing data on one or more variables	Easy to implement; no computations necessary	Potentially massive case loss; loss of statistical power; requires data to be MCAR; sensitive to producing biased parameter estimates
Complete variables analysis	Discards all cases that exhibit missing data on particular variable pairs	Easy to implement; reduced case loss; slightly more accurate than complete case analysis	Case loss; loss of statistical power; requires data to be MCAR; sensitive to producing biased parameter estimates
Mean substitution	Replaces missing data with the variable arithmetic mean	Easy to implement; preserves sample size	Can lead to considerable bias by diminishing the relationships between variables; requires data to be MCAR
Regression imputation	Replaces missing data with predicted values from a linear regression	Easy to implement; preserves sample size	Can lead to considerable bias by artificially inflating correlations between variables; requires data to be MCAR
Hot-deck imputation	Replaces missing data with actual values from similar donor cases in the same dataset	Replaces missing data with realistic, actual values from within the dataset	Can lead to biased estimates of correlations and regression weights even when data are MCAR; requires a similar donor case to be present in the dataset
Expectation maximization	Arrives at missing data estimates through an iterative approximation process based on maximum-likelihood	Yields relatively unbiased estimates and small standard errors; performs well when data are MCAR or MAR	Computationally intensive; mostly limited to imputing continuous data
Multiple imputation	Uses a Monte Carlo approach to compute $m > 1$ complete datasets for analysis before pooling parameter estimates and standard errors	Provides highly accurate imputation results for a broad range of statistical analyses; can be used to impute both continuous and categorical data; performs well when data are MCAR or MAR	Computationally intensive

Propensity Score Matching

This study examined the effects of CTE and college-preparatory high school curricula on secondary and postsecondary educational attainment. Outcomes for participants in each of these specialized curriculum types (i.e., the treatment groups) were compared separately to those for individuals who completed a general high school curriculum (i.e., the control group). The availability of equivalent comparison groups is a fundamental prerequisite for the unbiased estimation of causal treatment effects. Depending on the particular research context, equivalent comparison groups can be established either through randomized experiments or covariate matching techniques.

Using Randomized Experiments to Determine Causal Treatment Effects

Over 80 years ago, Fisher's (1928) seminal work on statistical methods for researchers established the use of randomized experiments as the epitome of scientifically-based inquiry. Since then, randomized experiments have been widely accepted to be the most rigorous method, or *gold standard*, for the estimation of causal treatment effects (Burtless, 1995; Campbell & Stanley, 1963; Holland, 1986; Slavin, 2002). The Institute of Education Sciences' (2006) evidence standards for scientifically-based research provide a de facto federal endorsement of randomized experiments as the preferred research paradigm in education (Gemici & Rojewski, 2007). Randomized experiments draw a sample from a specific population of interest and randomly assign individuals from that sample to either a treatment or control group (Torgerson & Torgerson, 2001). The unique advantage of random assignment lies in its ability to overcome the problem of experimental bias by balancing, on average, all observed and unobserved characteristics of participants evenly across treatment and control groups

(Agodini & Dynarski, 2004; Fisher, 1935; Schmidt, Baltussen, & Sauerborn, 2000). Eliminating bias strengthens the internal validity of an experimental study, which denotes the design's ability to attribute observed differences between groups exclusively to a given treatment (Bauman, 2006). The American Psychological Association (2002) noted that "randomized controlled experiments represent a more stringent way to evaluate treatment efficacy because they are the most effective way to rule out threats to internal validity in a single experiment" (p. 1054).

Selection bias is a particularly grave threat to internal validity (Torgerson & Roberts, 1999). Selection bias occurs when, in the absence of randomization, individuals either self-select into treatment (Bryson et al., 2002), or are subject to endogenous assignment by the researcher based on some underlying rationale (Dehejia & Wahba, 2002). Isolating causal treatment effects is not viable under this scenario due to the potential existence of systematic pre-treatment differences between participants (Chase, 2002). Randomized experiments overcome such systematic pre-treatment differences because variables that influence self-selection into treatment are, on average, evenly distributed across both treatment and control groups (Baker, 2000; Gall, Gall, & Borg, 2007). Therefore, exogenous selection into treatment allows for deriving unbiased estimates of causal treatment effects based on observed differences between groups (Rosenbaum, 1995).

Notwithstanding the ability of randomized experiments to eliminate selection bias, randomization can be problematic in social science research. Ethical issues may arise from randomized experimental designs where a potentially beneficial treatment is deliberately denied to participants in the control group (De Anda, 2007; Kirkwood, Cousens, Victora, & De Zoysa, 1997). When control group participants perceive the withheld treatment as desirable they may become demotivated or decide to drop out of the study entirely, resulting in internal validity issues due to experimental mortality (Gall et al., 2007). In survey research, post-assignment attrition due to the perceived withholding of beneficial treatments has been shown to represent a major validity threat (White & Lakey, 1992). Comparison group substitution can become a further source of randomization bias when control group participants look for alternative ways to obtain the perceived benefits from withheld treatment (Heckman & Smith, 1995). Ultimately, in many areas of social science research carrying out randomized experiments is often unfeasible due to excessive costs and logistical barriers (Moore, Graham, & Diamond, 2003; Titus, 2007).

Using Covariate Matching Techniques to Determine Causal Treatment Effects

In situations where ex ante random assignment is impractical researchers may consider the use of post-intervention observational data in causal-comparative designs. Yet, as discussed earlier, the issue of selection bias negates the determination of causal treatment effects in the absence of randomized treatment and control groups, and real effects from the intervention are either overestimated or underestimated (Aakvik, 2001; Monette, Sullivan, & DeJong, 2004). Against this backdrop, covariate matching techniques have been developed to statistically model equivalent comparison groups by balancing a given analysis sample on observable background characteristics (Heckman & Navarro-Lozano, 2004; McCaffrey, Ridgeway, & Morral, 2004).

Propensity score matching (PSM) is among the most prominent covariate matching techniques used to estimate causal treatment effects post hoc from observational data (Caliendo & Kopeinig, 2005; Hahs-Vaughn & Onwuegbuzie, 2006). The application of PSM spans the literature in statistics (Rubin, 2006), econometrics (Abadie & Imbens, 2006), medicine (Christakis & Iwashyna, 2003), sociology (DiPrete & Engelhardt, 2004), political science (Imai, 2005), and education (Saiz & Zoido, 2005). Introduced by Rosenbaum and Rubin (1983), the propensity score is a single variable score that is defined as the probability of participating in a treatment or intervention based on a number of observable variables, or covariates. Formally, the propensity score is expressed as

$$\mathrm{e}(x) = \mathrm{pr}(z = 1 \mid x)$$

where x denotes the specified vector of covariates for the propensity score model, and the binary variable z indicates exposure to treatment (Rosenbaum & Rubin, 1985). The propensity e(x) for each individual is estimated through logistic regression of z on x, where z equals *1* for treatment group participants and *0* for control group participants (Rosenbaum, 1998).

Within the context of this study, effects of CTE and college-preparatory high school curricula on educational attainment can be credibly estimated only if participants in these respective specialized curricula are compared to participants in a general curriculum who exhibit similar background characteristics, or covariate values. The propensity score expresses a participant's probability of observing either a CTE, college-preparatory, or general curriculum based on his or her observable covariates. Matching treatment and control group participants on the propensity score then permits an estimation of causal treatment effects without overt bias. The key benefit of the nonparametric propensity score lies in its ability to sidestep the challenges inherent in parametric matching procedures, which require exact matches on every single covariate in a model (Leow, Marcus, Zanutto, & Boruch, 2004). In an influential article on designing observational studies, Cochran and Chambers (1965) illustrated early on how augmenting the number of covariates in a multivariate model causes an exponential increase in the number of possible matches, thus greatly exacerbating the challenge of finding equivalent controls. For instance, if a given multivariate matching model includes 15 dichotomous observable covariates, then 2^{15} (or 32,768) different matches of covariates would be possible. Finding exact matches for treatment participants on all 2^{15} covariates would be an insoluble dilemma that is known in the literature as the *curse of dimensionality* (Augurzky & Schmidt, 2001; Caliendo & Kopeinig, 2005; Dehejia & Wahba, 1999). The propensity score overcomes the dimensionality issue by collapsing a large number of observable covariates into a scalar variable between *0* and *1* (Luellen, Shadish, & Clark, 2005). Matching then occurs on the scalar as the primary reference variable (Aiken, Smith, & Lake, 1994; Rosenbaum, 1986; Rubin & Thomas, 1996), and effects can be determined by comparing the respective outcomes of treatment and control group participants who are highly similar (Becker & Ichino, 2002).

Traditional parametric estimators such as multiple regression analysis assume linear or logistic covariate distributions that may not accurately reflect actual covariate distributions, which can differ considerably between treatment and control groups. As a nonparametric procedure, PSM can be performed free of the functional form requirements of parametric procedures and is, therefore, more efficient in scenarios where covariate distribution is heterogeneous. This robustness represents an important advantage of PSM, since heterogeneous covariate distributions are a common occurrence in many datasets. A further benefit of the propensity score emanates from its use of observational data, which may allow for increases in sample size at low statistical cost (Schmidt et al., 2000).

Since the propensity score is an estimate as opposed to a factual outcome score, it is only as reliable as the quality of the assumptions on which it is based (Lechner, 1999a). Analyses using PSM rest on two cardinal assumptions. First, the *stable unit treatment value assumption* (SUTVA) requires the absence of treatment interference across units, meaning that the treatment effect on one individual must be independent of the treatment participation of other individuals in the treatment (Harknett, 2006; Morgan & Harding, 2006). A violation of the SUTVA would occur if peer effects influenced an individual's treatment status (Titus, 2007). Second, the *conditional independence assumption* (CIA) entails that an individual's decision to participate in a given treatment is based exclusively on a set of observable covariates (Rubin, 1977). Whereas randomized controlled experiments effectively balance both observable and unobservable covariates evenly across groups (Schmidt et al., 2000), PSM presumes that any remaining *unobservable* covariates are irrelevant for selection into treatment. The absence of a correlation between unobserved covariates and the participation decision is known as *strong ignorability* (Rosenbaum, 1998; Rosenbaum & Rubin, 1983) and ensures that every participant in the sample has a positive probability of being assigned either into the treatment or control group (Anstrom & Tsiatis, 2001). In this case, the selection equation can successfully remove bias and create comparable treatment and control groups for the estimation of causal effects.

Whether a given PSM model achieves conditional independence is untestable, since accounting for potentially influential variables is not viable if they cannot be observed or measured (Leow et al., 2004; Zhao, 2003). Clearly, where unobservable covariates impact the selection decision the CIA is violated and matching may produce biased estimates of the treatment effect (Behrman, Cheng, & Todd, 2004). Because PSM models are not resistant to the existence of endonegeity bias due to unobserved, causally-relevant covariates (Reiter, 2000), the CIA is a strong assumption and can be upheld only "where there is a firm understanding, based on theory and past empirical evidence, of determinants of programme participation and the outcomes of interest" (Bryson et al., 2002, p. 19). However, Bryson et al. also emphasized that violations of the CIA or the SUTVA do not necessarily discredit PSM or other matching approaches as long as researchers are aware of the magnitude of the violation and the nature and direction of the resulting bias.

Various replications of randomized benchmark studies have been conducted to assess the accuracy of PSM-based effect estimates. Such reproductions of experimental designs have yielded mixed outcomes. While numerous researchers have successfully replicated experiment-based effect estimates via PSM (Dehejia & Wahba, 2002; see also Heckman, Ichimura, & Todd, 1997; Hotz, Imbens, & Klerman, 2000; Newman, Pradhan, Rawlings, Ridder, Coa, & Evia, 2002), some have fallen short of generating satisfactory results (Agodini & Dynarski, 2004; see also Bloom, Michalopoulos, Hill, & Lei, 2002; Smith & Todd, 2005; Wilde & Hollister, 2002). Yet, failure to replicate experimental benchmarks does not necessarily indicate intrinsic deficiencies in propensity score modeling, but may be triggered by imperfect data or technical issues related to the matching process. As Glazerman, Levy, and Myers (2002) pointed out, for non-experimental estimators such as PSM to generate valid results data quality and methodical accuracy are critical. The following section provides a description of the PSM process in order to facilitate a deeper understanding of the method's complexities and decision points.

The Process of Propensity Score Matching

The fundamental purpose of calculating the propensity score is to identify those individuals from within a large pool of control group participants who exhibit covariate attributes that are similar to those of treatment group participants. Once comparable treatment and control group participants have been matched based on the propensity score, adequate statistical analyses can be conducted that determine causal treatment effects (Heckman, Ichimura, & Todd, 1998). The PSM process itself can be partitioned into the following steps:

1. Choosing the regression model.
2. Selecting covariates for the model.
3. Choosing a matching algorithm.
4. Checking post-matching covariate balance
5. Analyzing sensitivity to unobserved covariates.

The following sections provide details for each step.

Choosing the regression model. The choice of the regression model depends on the number of available treatments. In the binary treatment case, binomial logit or probit models can be used for estimation, whereas the multiple treatment case calls for either multinomial logit or probit models, or the sequential use of binomial models (Caliendo &

Kopeinig, 2005; Imbens, 2000). In a multiple treatment case the sequential use of binomial models is considered less complex and may result in a more robust model (Lechner, 1999b). Rosenbaum and Rubin (1985) based their original logistic regression model for propensity score estimation on the following logit model by Cox (1970),

$$q(x) \equiv \log\left[\frac{\left(1 - \mathrm{e}(x)\right)}{\mathrm{e}(x)}\right] = \alpha + \beta^{\mathrm{T}} f(x)$$

where α and β are parameters to be estimated, q(x) is the log odds against partaking in the treatment, and $f(x)$ is a specified function of observable covariates. The propensity score logit model "weighs the significance of each variable accordingly through the coefficients, thereby giving the important predictors of treatment implicit priority in the matching process" (Leow et al., 2004, p. 468). Since binomial logit and probit models produce similar estimation results, the choice of a particular approach over another has no practical importance.

Selecting covariates for the model. Once a model has been chosen, selecting the *right* covariates for inclusion in the model becomes a critical task. Variables that neither influence selection into treatment nor cause differences in the outcome of interest are irrelevant and should be excluded from the propensity score model (Bryson et al., 2002). Beyond the exclusion of clearly irrelevant covariates, the literature is divided over the most effective covariate selection strategy. One school of thought encourages the use of parsimonious covariate specifications for the propensity score model. Especially in situations where treatment and control groups differ considerably, including only those covariates that have a strong impact on treatment selection and outcome reduces standard errors and increases the amount of *stochastic noise*, or randomness in the participation decision (Augurzky & Schmidt, 2001). Stochastic noise is a prerequisite for the ability to match, and higher degrees of randomness result in higher probabilities of finding matches for treatment participants with high propensity scores for whom the identification of adequate controls would otherwise be challenging (Austin, Grootendorst, & Anderson, 2007; Heckman, Ichimura, & Todd, 1998). Several statistical tests are available to determine relevant variables and avoid overparameterization. *Statistical significance* (Caliendo & Kopeinig, 2005) is a straightforward approach that begins with a minimal initial model specification and successively adds only those covariates to the model that are statistically significant. The *prediction rate metric* (Breiman, Friedman, Olsen, & Stone, 1984; Smith & Todd, 2005) bases covariate choice on maximizing the within-sample correct prediction rates. The method assigns a value of *1* to observations where the estimated propensity score exceeds the sample proportion of treatment participants, and a value of *0* for the converse case. This method maximizes the overall prediction rate of selection into treatment for the sample. The prediction rate metric can be combined with statistical significance so that a covariate is included in the model specification if it contributes positively to prediction rates and is statistically significant (Heckman, Ichimura, Smith, & Todd, 1998). Finally, *leave-one-out cross validation* (Black & Smith, 2004) starts out with an extremely parsimonious model specification and, by successively adding more covariates to the model, compares the resulting mean squared errors of the predictions made with respective covariates.

Notwithstanding the potential advantages of parsimonious models, several researchers have advocated for including all available covariates that somehow influence treatment participation and outcome, regardless of statistical significance (see Hahs-Vaughn & Onwuegbuzie, 2006; Leow et al., 2004). In an important paper on the practical implementation of PSM theory, Rubin and Thomas (1996) asserted that "unless a variable can be excluded because there is consensus that it is unrelated to the outcome variables or not a proper covariate, it is advisable to include it in the propensity score model even if it is not statistically significant" (p. 253). Rosenbaum (2002) corroborated this stance by noting that statistical significance was a function of sample size and did not guarantee practical relevance. Overall, the choice of the *right* covariates for a given research context should be guided largely by theory and prior empirical research (Sianesi, 2004; Smith & Todd, 2005).

Choosing a matching algorithm. After specifying a vector of covariates and estimating propensity scores via logistic regression, one or more algorithms need to be selected to match treatment participants with suitable controls. Different matching algorithms exist, yet all methods seek to strike a reasonable balance between estimator variance and matching quality. Importantly, all matching algorithms should yield equivalent outcomes with increasing sample size (Bryson et al., 2002; Caliendo & Kopeinig, 2005). In a large-sample study applying various different matching procedures, Smith and Todd (2005) concluded that different algorithms did "not have strong or consistent effects on the estimated biases" (p. 350). In the case of small samples, however, the choice of one particular algorithm over another can lead to greater variation in matching results (Heckman et al., 1997). Therefore, PSM should be implemented using more than one single algorithm to verify the consistency of effect estimates (Zhao, 2003). The following section outlines four commonly used matching algorithms.

Nearest neighbor matching. Nearest neighbor matching is among the most commonly used algorithms and can be carried out with or without replacement. Nearest neighbor matching without replacement matches a given treatment group participant with a participant from the control group based on the proximity of their propensity scores (Caliendo & Kopeinig, 2005). Any control group participant can be matched with only one member from the treatment group such that the smallest propensity score distances between treatment and control observations are achieved across groups. Problems with this approach occur when the propensity score distributions of treatment and control groups differ considerably. This situation is commonly referred to as *strong selection into treatment* (Augurzky & Schmidt, 2001). Where strong selection into treatment requires many treated individuals with high propensity scores to be matched with few comparable controls, the number of adequate controls becomes exhausted quickly, leaving only low-scoring controls to be matched with high-scoring treatment participants. Under such a scenario, nearest neighbor matching without replacement is inefficient and yields low-quality matches. One way to overcome this problem is to allow for the possibility of replacement, whereby one control group member can be matched to several different treatment participants. Ratio settings specify the maximum number of treatment cases to which a single control case is allowed to be matched. Ratios of up to four or five control cases per treatment case have been found to yield efficient matching results, whereas higher ratios may produce negligible efficiency gains (Haviland, Nagin, & Rosenbaum, 2007). Moreover, the use of very high control-to-treatment case ratios makes it increasingly difficult to find good matches, therefore elevating the risk of biased parameter estimates (Smith, 1997). Overall, where selection into treatment is strong the use of proper replacement procedures is beneficial, for it leads to reduced bias through improvements in matching quality (Smith & Todd, 2005).

Stratification matching. Stratification matching divides the range of matches into separate layers and balances the propensity score within each stratum such that treatment and control participants have approximately the same probability of selection. Treatment effects are determined by the difference in mean outcomes between treatment recipients and controls (Aakvik, 2001). The advantage of stratification, or subclassification, is that it "involves direct comparisons of ostensibly comparable groups of units within each subclass and therefore can be both understandable and persuasive to an audience with limited statistical training" (Rosenbaum & Rubin, 1983, p. 51). Division into five strata has been shown to reduce bias by approximately 90 percent for many continuous distributions (Cochran, 1968). In cases where no propensity score balance within a particular stratum can be achieved, the stratum needs to be split in order to reduce its size (Caliendo & Kopeinig, 2005). If the covariates within a stratum remain unbalanced after splitting, higher-order or interaction terms can be included in the model specification (Dehejia & Wahba, 1999).

Genetic matching. Genetic matching is a generalization of propensity score matching. An evolutionary search algorithm is used to compute a specific weight for each covariate such that an optimal balance is achieved for a given sample across the entire vector of covariates (Diamond & Sekhon, 2008). While computationally intensive, genetic matching has been demonstrated to be highly effective in balancing covariates across treatment and control groups (see Sekhon & Grieve, 2009).

Full matching. Full matching is an approach that partitions a sample into non-overlapping subclasses within which close treatment and control units are matched (Rosenbaum, 1991). Each subclass contains a matched set of cases, whereby a treatment case can be matched to several control cases, or vice-versa. The principal advantages of full matching are the method's high efficiency and sample size preservation, since all treatment and control units are matched (Hansen, 2004; Hill et al., 2005).

Restrictions for improving matching quality. Most available matching algorithms allow for the use of certain restrictions to avoid the possibility of distant (i.e., low quality) matches. Such measures include the enforcement of common support and the use of calipers. The *region of common support* is the area where the density masses of the propensity score distributions for treatment and control participants overlap. Under perfect conditions, a successful matching process would yield the same propensity score distributions for both groups. In practice this outcome is unlikely, and discrepancies arise when no comparable control group participants can be identified. Whereas randomized experiments ensure common support throughout the entire sample, the use of post hoc covariate matching methods such as PSM permits a reliable estimation of treatment effects only for those treatment participants for whom suitable controls have been identified during matching (Smith & Todd, 2005). For observations falling outside of the common support region selection bias is undefined (Heckman, Ichimura, Smith et al., 1998). Heckman et al. (1997) pointed out that the use of common support restrictions "requires that the parameter of interest…be redefined as the mean impact over the common support region" (p. 632).

A visual inspection of covariate distributions is the most basic method to verify whether the common support condition has been met (Lechner, 2000). A more accurate alternative is the *minima-maxima comparison* that excludes all observations from analysis for which the propensity score intervals of treatment and control groups do not overlap. For instance, if propensity scores fall in the range [0.14, 0.94] for the treatment group and [0.09, 0.78] for the control group, the region of common support under the minima-maxima comparison is defined as the interval [0.14, 0.78]. All observations that lie outside of lower and upper bounds of the common support interval are discarded because no comparable match is available for these cases.

Ignoring a potential common support problem can lead to inaccuracies in the estimation of causal treatment effects. Where the number of discarded observations is small, the common support problem is negligible (Aakvik, 2001). However, when numerous observations fall outside the region of common support the estimated effect only partially represents the original sample. In this case, the distribution of individual background characteristics for the discarded cases should be examined for systematic differences between treatment recipients and controls (Heckman, Ichimura, Smith et al., 1998). Lechner (2000) proposed a procedure for nonparametric bounds analysis to adjust estimated treatment effects in the absence of common support. Instead of calculating effects only for the subpopulation within the common support region, bounds analysis includes all individuals in the sample and measures the impact of observations outside the common support on the effect estimate for the subgroup that meets the common support condition. This is accomplished by comparing the proportions of within common support participants relative to the total number of participants and individuals from the treatment group outside the common support.

The use of *calipers* is a further restrictive measure to ensure high quality matches. Introduced by Cochran and Rubin (1973), a caliper defines a bound around a matched observation with the objective to reduce bias. When used for PSM, a caliper delimits a maximum acceptable propensity score distance so that controls will be matched with treatment recipients only if their respective propensity scores fall within the determined caliper bounds. Rosenbaum and Rubin (1985) suggested that the caliper size equal .25 times the standard deviation of the propensity scores to achieve high quality matching results. While the reduced number of possible matches resulting from caliper restrictions can lead to reductions in sample size and increases in estimator variance, overall matching quality is improved by avoiding matches with overly distant propensity scores.

Checking post-matching covariate balance. The matching process can be considered successful if treatment and control group participants are balanced across the entire vector of covariates. Several approaches have traditionally been used to verify covariate balance after matching. The pseudo R-square method is one possible approach to checking covariate balance after matching. In contrast to ordinary least-squares (OLS) regression, the logistic regression procedure used to calculate propensity scores produces maximum-likelihood estimates that are not aimed at minimizing variance. Consequently, an equivalent to the OLS R-square statistic, which indicates the goodness-of-fit for an OLS regression model, is not readily available in logistic regression. Pseudo R-square measures, such as the Cox and Snell (1989) or Nagelkerke (1991) tests, have been developed in order to facilitate an assessment of validity for logistic regression and thus emulate the goodness-of-fit notion available in OLS regression (Meyers, Gamst, & Guarino, 2005). The pseudo R-square value itself emanates from a re-estimation of the propensity score on the *already matched* sample in order to detect differences in covariate distributions across treatment and control groups. A low pseudo R-square value indicates a successful match (Caliendo & Kopeinig, 2005).

While the pseudo R-square approach provides a certain indication of covariate balance, formal hypothesis tests have been the method of choice used in the overwhelming majority of applied PSM studies (see, e.g., Caliendo, Hujer, & Thomsen, 2005; D'Agostino, 1998; Leow et al., 2004). Hypothesis tests based on *t*-tests (for continuous covariates) and chi-square analysis (for dichotomous covariates) reveal statistically significant differences between treatment and control groups on one or more covariates in the PSM model. If no statistically significant differences at conventional alpha levels remain on any covariate after matching, the estimates of causal treatment effects based on the matched sample are considered free of overt bias. Notwithstanding the overwhelming popularity of conducting hypothesis tests to establish covariate balance, recent developments in PSM theory have highlighted several issues regarding this approach (see Ho, Imai, King, & Stuart, 2007a; Imai, King, & Stuart, 2008). Specifically, these studies described the *balance test fallacy* of using hypothesis tests to evaluate covariate balance by showing that (a) *t* and chi-square statistics depend on the number of control cases in a given sample, (b) the power of *t* and chi-square statistics decreases as more cases are dropped during the matching process, (c) hypothesis tests are irrelevant from a theoretical standpoint, for covariate balance is specific to a given sample instead of some hypothetical population, and (d) there is no threshold such as conventional significance levels below which the extent of remaining covariate imbalance is always acceptable. Visual checks of propensity score distributions and covariate balance using jitter and quantile-quantile (QQ) plots are considered preferred alternatives to traditional hypothesis tests. Whereas jitter plots graphically display the propensity scores for matched and unmatched treatment and control cases, QQ-plots allow researchers to visually compare the empirical distribution of individual covariates before and after the matching process.

As a complement to visual balance checks, standardized covariate mean differences between treatment and control groups are examined before and after matching. Pre and post-matching changes in the standardized mean difference are expressed in percent and provide information about the extent to which covariate balance has either improved or deteriorated as a result of the matching process. Technically, the percent change in standardized mean difference is expressed as

$$\frac{100(\bar{x}_1 - \bar{x}_0)}{\left[\frac{(s_1^2 + s_0^2)}{2}\right]^{\frac{1}{2}}}$$

where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985). While it is not readily visible whether bias reduction has been sufficient relative to some arbitrary benchmark (Caliendo & Kopeinig, 2005), knowledge of the magnitude of balance improvement may nonetheless provide an indication of matching success in combination with visual assessment measures, especially for those covariates with apparent pre-matching differences.

Analyzing sensitivity to unobserved covariates. After estimating causal treatment effects based on the matched sample of treatment and control group participants, the robustness of effect estimates needs to be determined. This is necessary because PSM is limited to removing overt bias, which is bias resulting from a set of *observed* covariates. Yet, hidden bias emanating from the existence of relevant *unobserved* covariates may well impact the selection decision even after successfully balancing on observables. Rosenbaum (1998) recognized that “propensity scores do little or nothing about hidden biases, which must be addressed by other means” (p. 1). Conceptually, sensitivity analysis assumes the existence of an influential, yet unobserved, covariate *U* and posits “various associations between *U* and the outcomes, and between *U* and treatment assignment” (Rubin, 2006, p. 168). The existence of *U* affects an individual’s odds of receiving treatment at different levels, depending on the assumed strength of the impact of *U*. The estimated treatment effect is considered robust if it remains measurable despite adjustments for the impact of *U*. To implement this concept with continuous outcome variables, Rosenbaum (2002) proposed a bounding approach to measure the extent to which endonegeity bias impacts the selection decision when strong ignorability cannot be reasonably assumed. Sensitivity analysis is based on the Wilcoxon (1945) signed rank test for continuous outcome variables, and on the Mantel and Haenszel (1959) test statistic for dichotomous outcome variables. In either case, different levels of hidden bias are expressed by the parameter γ, and where $\gamma = 1$, no hidden bias is assumed to be present. Increasing values of γ reflect increasing uncertainty about the impact of an unobserved covariate *U* on the parameter estimate. Overall, while bounding does not per se test for the existence of unobserved covariates, it allows researchers to gauge the degree to which effect estimates may be undermined by potential endonegeity bias (Becker & Caliendo, 2007).

Summary

This section discussed the PSM process and outlined the method’s potential benefits and limitations. Given appropriate data and justifiable assumptions, causal treatment effects can be estimated credibly from observational data using PSM. On the surface, PSM appears to be a straightforward, albeit lengthy, process that can be executed in an almost mechanical fashion. However, researchers are faced with numerous intricacies and decision points during implementation. Making the *right* choices that produce meaningful results demands a profound understanding of the assumptions inherent in the matching procedure, as well as the nature of the underlying observational data set. While PSM should not be understood as a panacea for the limitations of observational studies (Rubin, 2006), it is a powerful approach to determine causal treatment effects from observational data in the absence of randomized experiments.

CHAPTER 3

METHOD

This chapter illustrates the methodological approaches used to conduct the present study. Included herein are descriptions of the (a) research design, (b) implementation of multiple imputation, (c) implementation of propensity score matching, and (d) post-matching data analysis.

Research Design

This study examined the effects of CTE and college-preparatory high school curricula on secondary and postsecondary educational attainment. Given ongoing debates over the resource intensity of CTE (Cavanagh, 2005; Gray, 2004), outcomes for CTE concentrators were of particular interest. Data from the National Longitudinal Survey of Youth 1997 (NLSY97, U.S. Bureau of Labor Statistics, 2009a) were used to determine potential curriculum effects. NLSY97 is a large-scale dataset designed to provide information about educational and labor market experiences of a nationally-representative group of men and women in the United States. When using large-scale observational datasets researchers need to address two critical issues. For one, almost all datasets contain missing data and require the implementation of appropriate strategies to minimize bias. While a number of statistical alternatives to dealing with missing data exist, many studies in educational research either ignore the issue or use overly simplistic approaches that exacerbate, rather than mitigate, bias. This study illustrates a best-practice approach to treating missing data by using multiple imputation (MI) as one of several modern imputation techniques.

The second important issue concerning causal-comparative analyses using complex observational datasets arises from the fact that causal inferences can be made only in the absence of systematic, or qualitative, differences between treatment and control groups. Propensity score matching (PSM) is a powerful approach for creating, post hoc, equivalent treatment and control groups by balancing a given sample on observable background variables, or covariates. The advantage of using PSM for non-parametric preprocessing of a sample lies in the ability to control for observable covariates so that causal inferences can be derived free of overt bias. The method's principal limitation revolves around the potential existence of hidden bias due to unobserved, causally relevant concomitants. By applying PSM to the sample from the NLSY97, two pairs of comparable treatment and control groups (i.e., CTE and general-track students; college-preparatory and general-track students) were created as a basis for estimating curriculum effects. Subsequently, chi-square analysis was used to determine variations in secondary and postsecondary educational attainment.

Sample

Data from the NLSY97 were used to determine the potential effects of different high school curricula on secondary and postsecondary educational achievement. Sponsored and managed by the U.S. Bureau of Labor Statistics (2005), NLSY97 is an ongoing annual survey that provides data to examine the transition process of secondary students into postsecondary education and/or the workplace. The NLSY97 sample was selected to "represent the civilian, noninstitutional population of the United States within the eligible age range - 12 to 16 years of age as of December 31, 1996 - with oversamples of Hispanics and non-Hispanic blacks" (Moore, Pedlow, Krishnamurty, & Wolter, 2000, p. 11). To collect data for the 1997 base year, over 90,000 U.S. Census-based, randomly selected housing units were initially screened to identify household residents in the eligible age range. These household residents were asked to participate in the survey interview using a computer-assisted personal interviewing system. Separate interviews were conducted with one resident parent or parent figure to gather additional information. While youth questionnaires collected detailed information on school experiences, employment activities, family background, social behavior, health status, and financial characteristics, parent questionnaires focused on family life and other aspects related to their NLSY97-eligible children's lives. The availability of a wide variety of student, family, and education-related background variables has made NLSY97 particularly useful for the exploration of high school curriculum effects.

The comprehensive NLSY97 dataset comprises an original sample of 8,984 respondents that is representative of those U.S. residents in 1997 who were born between 1980 and 1984. Follow-up survey rounds have been conducted annually in order to track the original youth sample throughout secondary schooling and into post-high school transition and adulthood. Of the original sample, 1,852 students were enrolled in the ninth grade of a regular secondary program during the 1997 school year. This cohort of 1997 ninth graders was used as the base sample for the present study. Transcript information was available for 1,199 individuals from the base sample. While the use of an alternative self-report high school program variable would have resulted in a larger sample, such self-report information is considered less accurate than transcript information (National

Center for Education Statistics, 2009). Since curriculum type represented the treatment condition in the present causal-comparative design, the sample was restricted to those individuals for whom transcript information on curriculum type was available. A further 84 individuals were removed from the sample because they were classified as participants in a combined CTE/college-preparatory curriculum. Even though an examination of curriculum effects for this combined category would have been desirable, the available sample size was too small for the specific data analytic procedures used in this study.

In NLSY97, as in most other large-scale datasets, missing values that occur because a respondent is not supposed to answer a certain question (e.g., a question that applies only to respondents of a specific gender, ethnicity, or age range) are coded as *legitimate item skips*. Such skips are qualitatively different from regular missing values and should not be imputed, for they are intended to be missing. One hundred eighty-nine cases containing legitimate item skips were deleted listwise (see the Multiple Imputation section for detailed information about deleted cases). A final analysis sample emerged containing 926 individuals who were in ninth grade during the NLSY97 base year. Of this final sample, 262 individuals were enrolled in a CTE curriculum, 204 individuals were enrolled in a college-preparatory curriculum, and 460 individuals were classified as general-track students (see the Measures section for detailed information about curriculum classifications). Table 3.1 provides select descriptive data for the final sample by curriculum type.

Table 3.1
Select Descriptive Data for the Final Analysis Sample

		Curriculum type					
		CTE		*College-preparatory*		*General*	
Variables	*Categories*	*n*	*%*	*n*	*%*	*n*	*%*
Gender	Male	164	35.5	73	15.8	225	48.7
	Female	98	21.1	131	28.2	235	50.6
Race/ethnicity	Black	57	24.1	35	14.8	145	61.2
	Hispanic	43	24.6	30	17.1	102	58.3
	Non-Black/Non-Hispanic	162	31.5	139	27.0	213	41.4
School type	Public	253	29.5	172	20.0	434	50.5
	Private and other	9	13.4	32	47.8	26	38.8
Ever suspended from school	No	184	28.4	189	29.2	274	42.3
	Yes	78	28.0	15	5.4	186	66.7

Note. Descriptive data are pre-imputation and pre-matching.

Measures

Propensity score matching (PSM) was used to create comparable and equivalent treatment and control groups as a basis for estimating causal effects. Conceptually, a series of background variables, or covariates, available in NLSY97 were selected to predict participation in different curriculum types or treatments. Under the assumption of *strong ignorability* (Rosenbaum & Rubin, 1983), which holds that an individual's selection into treatment is unaffected by unobserved covariates, curriculum effects on secondary and postsecondary educational achievement were estimated. The conceptual model underlying data analysis is depicted in Figure 1.1.

Covariates

PSM uses observable covariates to model, via logistic regression, a student's propensity of receiving a treatment or intervention. For the propensity score model to be efficient, it is essential that (a) only those covariates be included in the model specification that influence treatment participation and outcome (Caliendo & Kopeinig, 2005; McCaffrey et al., 2004), and (b) the selection of covariates be guided by theory and prior empirical research (Bryson, Dorsett, & Purdon, 2002; Sianesi, 2004; Smith & Todd, 2005).

A plethora of literature exists about predictors of high school track placement (e.g., Agodini et al., 2004; Jones et al., 1995; Oakes et al., 1992; Stone & Aliaga, 2005) and educational attainment (Hossler & Stage, 1992; Jimerson et al., 2000; Kao & Tienda, 1998; Lee & Staff, 2007; Plank et al., 2008, Rylance, 1997; Swanson, 2004. Consistently, such predictors have included gender, race/ethnicity, socioeconomic status, urbanicity, academic achievement (especially in eighth grade), participation in work-based learning, special needs status, English language learner status, behavior in school, school affluence and control, and parental education. In a longitudinal analysis of high school transcripts, Levesque (2003) provided specific background data for students whose course-taking patterns during high school identified them as CTE concentrators (see Table 3.2).

Table 3.2
Characteristics of Public High School Graduates Identified as CTE Concentrators in Percent (taken from Levesque, 2003)

	1994	*1998*
Sex		
Male	55.4	58.6
Female	44.6	41.4
Race/ethnicity		
American Indian/Alaska Native	.7	.5
Asian/Pacific Islander	2.0	2.4
Black/Non-Hispanic	14.4	14.8
Hispanic	7.9	10.5
White, non-Hispanic	75.0	71.7
School urbanicity		
Urban	–	26.2
Suburban	–	33.8
Rural	–	40.1
School poverty level		
High	9.9	10.7
Middle	62.2	64.2
Low	10.6	10.2
Not reported	17.3	14.9
Grade 9 mathematics level		
High	7.0	14.3
Medium	59.9	65.5
Low	33.1	20.2
Grade-point average		
Above 3.5	7.0	10.5
Between 2.0 and 3.5	75.8	79.5
Below 2.0	17.3	10.0
Academic coursework completed	5.1	8.5
All high-level	90.5	87.8
Mid-level or mixed	4.3	3.7
All low		
English proficiency		
Limited	.6	.2
Proficient	99.4	99.8
Disability status		
Has disability	6.0	4.2
No indicated disability	94.0	95.8

- Not available

Note. CTE concentrators earned 3.0 or more credits in one of the following 10 broad occupational program areas: agriculture, business, marketing, health care, protective services, technology, trade and industry, food service and hospitality, child care and education, and personal and other services. Percentages may not add to 100.0 due to rounding.

Levesque's (2003) data showed that in the period between 1994 and 1998, CTE concentrators were predominantly male, low-achieving, of minority status, English proficient, disabled, and from rural areas. Table 3.3 illustrates the specific vector of covariates used in the present study to determine a given student's propensity for participating in a CTE, college-preparatory, or general high school curriculum.

Table 3.3

NLSY97 Covariates Selected for Propensity Score Estimation

Variable	*Variable designation*	*Variable levels*
Survey weight	CUSTOM_WEIGHT	Continuous
Gender	KEY!SEX	1=Male 2=Female
Race/ethnicity (recoded, dummy-coded)	KEY!RACE_ETHNICITY	1=Black 2=Hispanic 3=Non-Black/Non-Hispanic
Urbanicity (recoded)	CV_URBAN-RURAL	0=Rural 1=Urban
Household poverty ratio (square root)	CV_HH_POV_RATIO	Composite
Grades received in eighth grade (recoded)	YSCH-6800	Continuous
PIAT math standard score[a]	CV_PIAT_STANDARD_UPD	Continuous
Work-based learning (composite)	WBL	0=No 1=Yes
Remedial English and/or math (composite)	REMEDIAL	0=No 1=Yes
ESL and/or bilingual program (composite)	ESL	0=No 1=Yes
Educational and/or physical handicap (composite)	HANDICAP	0=No 1=Yes
Attitudes toward school (composite)	ATS	Continuous
Number of days absent from school	YSCH-36300	Continuous
Ever suspended from school	YSCH-5800	0=No 1=Yes
School type (recoded)	CV_SCHOOL_TYPE	1=Public 2=Private and other
Student-teacher ratio[b]	CV_STUDENT_TEACHER_RATIO	1=<14 2=14 to <18 3= 18 to <22 4=22+
Percent peers college-bound	YPRS-1100	1=Almost none (less than 10%) 2=About 25% 3=About half (50%) 4=About 75% 5=Almost all (more than 90%)

[a] The Peabody Individual Achievement Test (PIAT) is a widely used brief assessment of academic achievement for children ages five and over. The PIAT mathematics assessment was given to all NLSY97 respondents who were in ninth grade or lower during the base year administration.

[b] Student-teacher ratio was used as a proxy for school resources (Balfanz, 2009).

Prior to estimating propensity scores for treatment participation through logistic regression, covariates used in my analysis sample were manipulated as follows:

Weighting. Large-scale datasets use complex, multi-stage sampling schemes that require weighting to ensure that parameter estimates are representative of the target population. The literature provides no clear guidance on the use of survey weights in combination with nonparametric matching methods such as PSM. King (2009) noted that separate weighting, either on pre-matching or post-matching samples, should not be performed because survey weights have no effect on bias when estimating a single constant treatment effect. Instead, survey weights should be included in the propensity score model as a covariate to account for the complexities of multi-stage sampling. Following this recommendation, a custom survey weight was created through the NLSY97 web interface (U.S. Bureau of Labor Statistics, 2009b) to be included as a covariate in the propensity score model. The creation of a custom survey weight was necessary due to the use of cross-sectional NLSY97 data from the 1997 base year and select follow-up administrations, including the survey years 1999 for the curriculum type treatment variable and 2007 for the educational attainment outcome variable. Appropriate custom weights for the required years are automatically calculated through the web interface. No specific information is available about the calculation routines used in the creation of custom weights.

Recoding. Several variables, including race/ethnicity, urbanicity, school type, and grades received in eighth grade, were recoded to reduce the number of distinct variable levels. While such recoding may have reduced some variance in the data, this step was necessary because one or more of the original variable levels contained very low case counts. The existence of variable levels with very low case counts severely complicates the establishment of covariate balance during propensity score matching.

Dummy coding. Logistic regression models can be specified using continuous and categorical predictor variables. However, the use of multi-level categorical variables requires dummy coding, whereby each individual variable level is coded as a separate binary variable. In my vector of covariates, race/ethnicity represented a multi-level categorical variable and was dummy-coded for use in logistic regression. The non-Black/non-Hispanic race/ethnicity category represented the majority of cases and was used as the reference group for dummy coding. The Black and Hispanic categories were used as separate subgroups.

Transformations. Even though multiple imputation (MI) and PSM are fairly robust to deviations from multivariate normality, estimates are most stable when predictors are normally distributed. When predictors deviate from normality, several mathematical transformations, including log, square root, and reciprocal transformation can be applied (Field, 2005). The original household poverty ratio variable was positively skewed and leptokurtic (skewness = 2.45; kurtosis = 10.75). From within the available transformation options, square root transformation was applied to the household poverty ratio variable because it yielded the closest approximation to normality (skewness = .46; kurtosis = 1.34).

Attitudes toward school. Students' attitudes toward teachers and the school environment were captured in the NLSY97 base year interview by using the following Likert-type items:

1. Discipline is fair.
2. Disruptions by other students get in the way of my learning.
3. I feel safe at this school.
4. Students are fairly graded.
5. Teachers are good.
6. Teachers are interested in the students.
7. There is a lot of cheating on tests and assignments.

Response categories included 1 (*strongly agree*), 2 (*agree*), 3 (*disagree*), and 4 (*strongly disagree*). Since items 2 and 7 were reverse-coded in the original scale, these items were recoded into a continuous *attitudes toward school* composite variable with scores ranging from 7 (*very positive attitudes toward school*) to 28 (*very negative attitudes toward school*).

NLSY97 documentation did not provide validity or reliability information for the attitudes toward school scale. A reliability coefficient alpha of .621 was calculated for the specific analysis sample used in this study. The obtained coefficient alpha value was relatively low, meaning that not all of the seven items measured the attitudes toward school construct equally well. However, the low magnitude of this alpha value was not altogether unexpected given the limited number of items and nature of the construct being examined. The attitude toward school composite was retained as a covariate given Cronbach and Shavelson’s (2004) caution against using coefficient alpha as the best or sole way to judge instrument reliability.

Work-based learning. A dichotomous composite variable was created for participation in one or more work-based learning interventions, including cooperative education, tech-prep, and internship/apprenticeship.

Additional variable composites. Linear and logistic regression models that are overfitted (i.e., specified using too many independent variables) will produce overly optimistic results that will fail to replicate, thus producing unreliable outcomes and spurious conclusions (Babyak, 2004). To avoid overfitting the logistic regression model for propensity score estimation, several additional covariates were grouped into dichotomous composite variables. Specifically, two individual variables that captured remedial English and mathematics participation separately were collapsed into a single variable for remedial English and/or mathematics participation. The same procedure was applied to create composite variables for ESL and/or bilingual program participation, as well as educational and/or physical handicap status.

Interaction terms. Rosenbaum and Rubin (1984) demonstrated that the inclusion of relevant interaction terms can improve the estimation quality of the propensity score. Given the existence of interaction effects between gender and math achievement (Linver, Davis-Kean, & Eccles, 2002), and between socioeconomic status and academic achievement (Jones et al., 1995; Sirin, 2005), two interaction terms were included in the covariate specification. Interaction terms consisted of (a) gender by PIAT math standard scores, and (b) household poverty ratio (square root) by grades received in eighth grade.

Parental education. Even though data on parental education were available in NLSY97, these variables contained a high number of legitimate item skips that would not have been eligible for imputation. In the interest of preserving sample size, parental education was excluded from the propensity score model. This exclusion was deemed justifiable given the relationship between education and socioeconomic status (Mueller & Parcel, 1981; Oakes & Rossi, 2003), which was captured by the household poverty ratio variable.

Treatment and Control Conditions

In NLSY97, transcript data for each student are categorized to reflect a student's full course-taking behavior in high school as career-technical, college-preparatory, combined career-technical and college-preparatory, or general. TRANS_SCH_PGM is a NLSY97-generated curriculum type variable that was used as the treatment/control indicator in this study to compare CTE and college-preparatory against general-track students. TRANS_SCH_PGM is a composite of four curriculum type variables, as outlined in Table 3.4.

Table 3.4
NLSY97 Description of Transcript-based Curriculum Type Variables

Variable name	*Variable specification*
TRANS_ACAD_SPEC Academic specialist	Student earned: at least 4 credits in English at least 3 credits in mathematics at the Algebra 1 level or higher at least 2 credits in biology, chemistry, or physics at least 2 credits in social studies with at least 1 credit in US or world history at least 2 credits in a single foreign language
TRANS_ACAD_CONC Academic concentrator	Student earned: at least 4 credits in English at least 3 credits in mathematics at least 3 credits in science at least 3 credits in social studies
TRANS_VOC_SPEC Vocational specialist	Student earned: at least 4 credits in a single Specific Labor Market Preparation (SMLP) vocational area, with at least 2 of these credits in that SLMP's 2nd-level or higher courses or co-op/work experience coursework
TRANS_VOC_CONC Vocational concentrator	Student earned: at least 3 credits total in a single Specific Labor Market Preparation (SLMP) vocational area
TRANS_SCH_PGM School program (composite)	This variable combines the information from the four variables above and is coded as follows: Academic specialist (and not vocational concentrator)[a] Vocational concentrator (and not academic specialist)[b] Both academic specialist and vocational concentrator Neither academic specialist nor vocational concentrator[c]

[a] Referred to as *college-preparatory* in this study.
[b] Referred to as *career-technical education (CTE)* in this study.
[c] Referred to as *general* in this study.
Note. The combined academic specialist/vocational concentrator category was excluded from analysis in this study due to insufficient sample size. Given that credit systems vary considerably across schools, the credits indicated reflect the standardized credit system put forth the U.S. Department of Education’s (1999) secondary school taxonomy.

Outcome

This study examined effects of high school curriculum type on secondary and postsecondary educational attainment. CVC_HIGHEST_DEGREE_EVER_2007 is a variable that captures the highest level of formal education attained by an individual in the NLSY97 dataset as of 2007. The original variable coding consisted of seven levels, including 0 (*no diploma or degree*), 1 (*GED*), 2 (*regular high school diploma*), 3 (*associate/junior college degree*), 4 (*Bachelor's degree*), 5 (*Master's degree*), 6 (*PhD*), and 7 (*professional degree*). Given the very small number of individuals with graduate and professional degrees, these categories were combined with bachelor's degree holders. The recoded outcome variable for educational attainment consisted of five levels, including 0 (*no high school diploma or GED*), 1 (*GED*), 2 (*regular high school diploma*), 3 (*two-year college degree*), and 4 (*four-year college degree*).

Multiple Imputation

The handling of missing data was a particularly important aspect of this study because the use of PSM requires a complete data matrix. Multiple imputation (MI) was used to impute missing data in the present study's NLSY97 analysis sample. Similar to other modern imputation methods, MI is a *stochastic* (i.e., considering randomness) approach aimed at providing valid inferences for statistical estimates from incomplete data. Specifically, MI creates *m* complete datasets by generating synthetic data points drawn from a distribution of plausible values from a set of predictor variables. Each complete dataset is analyzed separately before the *m* analysis results are pooled into final point and variance estimates that incorporate the uncertainty inherent in the missing data (Enders, 2001; Schafer, 1999a). Calculating the efficiency of an imputation estimate by using

$$\left(1 + \frac{\gamma}{m}\right)^{-1}$$

where γ is the rate of missing information and *m* denotes the number of complete imputed datasets, Rubin (1987) demonstrated that as little as $m = 3$ imputations may suffice to achieve efficient results with 20 percent missing values in a given dataset. Unless the rate of missing data is very high, the marginal utility of producing a large number of imputed datasets is small, and $m = 5$ imputations is considered to yield robust imputation results under most circumstances (see, e.g., Schafer, 1997; Van Buuren et al., 1999).

MI can be carried out using specialized software applications such as MICE (Van Buuren & Oudshoorn, 2000) or NORM (Schafer, 1999b). Recently, the method has been incorporated into mainstream statistical software packages such as SPSS 17 or STATA 11. In this study, MI was carried out in SPSS 17, which uses the fully conditional specification (FCS) method. FCS is an imputation algorithm whereby missing values are imputed on a variable-by-variable basis (Van Buuren, 2007). A separate imputation model is specified for each variable with missing values, and imputations are carried out in sequential iterations until missing data points have been replaced in all variables. For *continuous* variables, sample mean and standard deviation estimates of the non-missing values are determined. Then, missing values are replaced with random draws from a normal distribution with mean and standard deviation equal to the sample values, but within the range of the observed minimum and maximum values. For *categorical* variables with missing values, the observed proportion of each category for the non-missing values is determined. Subsequently, missing values are replaced with random draws from a multinomial distribution with category probabilities equal to the observed category proportions. Several theoretical studies have demonstrated the ability of FCS to yield unbiased imputation estimates (e.g., Raghunathan et al., 2001; Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006).

Before proceeding with the imputation process, the specific missing data mechanism underlying the analysis sample had to be determined. Missing data mechanisms "offer different explanations for how missingness is probabilistically related to the values of variables in the dataset" (Enders, 2006, p. 315). Data can be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR; see Chapter 2 for a detailed discussion of these diverse mechanisms and their implications). Little's (1988a) MCAR test was used to determine whether data were missing completely at random. If Little's chi-square test statistic is statistically significant at conventional levels the assumption of data missing completely at random cannot be upheld. In such a case data can be either MAR or MNAR. Since no statistical test is available to further determine whether data are MAR or MNAR (Schafer & Graham, 2002), MI operates under the assumption that data are MAR (Schafer, 1999a). Results from Little's MCAR test showed that data in the sample were not MCAR, $\chi^2 = 47.771$, $p = .028$, $df = 31$. Given that traditional missing data methods, such as listwise deletion, pairwise deletion, or mean imputation provide unbiased parameter estimates only when data are MCAR, such methods could not be used in the present scenario. Analysis proceeded with MI, which provides unbiased estimates even when data are MAR.

For the present study, $m = 5$ complete datasets were imputed. Each of the five datasets was analyzed separately, meaning that each individual dataset underwent PSM and post-matching data analytic procedures as if it had been the only existing dataset in the study. Once all five imputed datasets were analyzed, results were pooled into a final set of parameter estimates and standard errors following procedures proposed by Rubin (1987). Based on Enders (2006), Table 3.5 outlines the specific computational steps involved in obtaining pooled parameter estimates and standard errors.

Table 3.5
Computational Steps for Pooling Multiple Parameter Estimates and Standard Errors

Statistic	*Formula*	
1. Parameter estimate	$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m} \hat{Q}_i$	where *m* is the number of imputations and $\hat{Q}_i$ is the parameter estimate from the *i*th imputed dataset
2. Standard error		
a. Within-imputation variance	$\bar{U} = \frac{1}{m}\sum_{i=1}^{m} \hat{U}_i$	where $\hat{U}_i$ is the variance estimate from the *i*th imputed dataset, and *m* is the number of imputations
b. Between-imputation variance	$B = \frac{1}{m}\sum_{i-1}^{m} \left(\hat{Q}_i - \bar{Q}\right)^2$	
c. Total imputation variance	$T = \bar{U} + \left(1 + \frac{1}{m}\right) B$	

In NLSY97, as in most other large-scale datasets, missing values that occur because a respondent is not supposed to answer a certain question (e.g., a question that applies only to respondents of a specific gender, race/ethnicity, or age) are coded as *legitimate item skips*. Such skips are qualitatively different from regular missing values, for they are intended to be missing. Thus, legitimate item skips may not be replaced with imputed data. Before conducting MI, 189 cases that contained legitimate item skips on background variables were deleted listwise from the initial analysis sample. Table 3.6 provides an overview of variables with missing data and their respective treatment.

Table 3.6
Deleted and Imputed Cases by Variable

Variable	*Listwise deleted cases*[a]		*Imputed cases*[b]	
	n	*%*[c]	*n*	*%*[d]
Urbanicity	-	-	37	4.0
Household poverty ratio	123	11.0	169	18.3
Grades received in eighth grade	1	.1	15	1.6
PIAT math standard score	7	.6	30	3.2
Work-based learning (cooperative education)	-	-	3	.3
Work-based learning (tech-prep)	-	-	3	.3
Work-based learning (internship)	-	-	3	.3
Remedial English	49	4.4	-	-
Attitudes toward school (teacher interest)			1	.1
Attitudes toward school (grade fairly)	-	-	1	.1
Attitudes toward school (cheat on test)	-	-	3	.3
Attitudes toward school (discipline is fair)	-	-	1	.1
Number of days absent from school	-	-	18	1.9
Student-teacher ratio	9	.8	37	4.0
Percent peers college-bound	-	-	6	.6
Total	189	17.0	327	35.3

[a] Cases containing legitimate item skips were deleted listwise.
[b] Cases containing missing values other than legitimate item skips were imputed.
[c] Based on the pre-listwise deletion sample of 1,115 cases.
[d] Based on the post-listwise deletion sample of 926 cases.

When specifying a multiple imputation model, constraints may be imposed on each variable in order to specify its particular role in the estimation procedure. Any variable can be specified as a (a) predictor variable for imputing missing values only, (b) dependent variable for which missing values are to be imputed only, or (c) both a predictor and dependent variable. Complete variables are typically used as predictors only, whereas variables with a low-to-moderate percentage of missing values are specified as both predictors and dependent variables. Variables that contain a high number of missing values are used as dependent variables only. For continuous variables, a minimum-maximum range of plausible values needs to be specified. The specific imputation constraints used for this study are outlined in Table 3.7.

Table 3.7
Multiple Imputation Constraints

	Role in imputation		*Imputed values*		
	Dependent	*Predictor*	*Minimum*	*Maximum*	*Rounding*
Survey weight	No	Yes			
Gender	No	Yes	-	-	-
Race/ethnicity	No	Yes	-	-	-
Urbanicity	Yes	Yes	-	-	-
Household poverty ratio (square root)	Yes	Yes	1	41	Integer
Grades received in eighth grade	Yes	Yes	1	8	Integer
PIAT math standard score	Yes	Yes	50	149	Integer
Work-based learning (coop)	Yes	Yes	-	-	-
Work-based learning (tech prep)	Yes	Yes	-	-	-
Work-based learning (internship)	Yes	Yes	-	-	-
Remedial English	No	Yes	-	-	-
Remedial math	No	Yes	-	-	-
ESL program	No	Yes	-	-	-
Bilingual program	No	Yes	-	-	-
Educational handicap	No	Yes	-	-	-
Physical handicap	No	Yes	-	-	-
Attitudes toward school (1)	No	Yes	-	-	-
Attitudes toward school (2)	Yes	Yes	-	-	-
Attitudes toward school (3)	No	Yes	-	-	-
Attitudes toward school (4)	Yes	Yes	-	-	-
Attitudes toward school (5)	Yes	Yes	-	-	-
Attitudes toward school (6)	Yes	Yes	-	-	-
Attitudes toward school (7)	No	Yes	-	-	-
Number of days absent from school	Yes	Yes	0	50	Integer
Ever suspended from school	No	Yes	-	-	-
School type	No	Yes	-	-	-
Student-teacher ratio	Yes	Yes	-	-	-
Percent peers college-bound	Yes	Yes	-	-	-
High school program of study	No	Yes	-	-	-
Highest degree ever received	No	Yes	-	-	-

Note. Minimum and maximum values apply to continuous variables used in the imputation model. Attitudes toward school items include *discipline is fair*, *disruptions by other students get in the way of my learning*, *I feel safe at this school*, *students are fairly graded*, *teachers are good*, *teachers are interested in students*, and *there is a lot of cheating on tests and assignments*.

Propensity Score Matching

The principal obstacle to using observational datasets for causal-comparative research lies in the biased nature of such data due to the lack of random assignment. The PSM process mitigates this limitation by balancing treatment and control groups, post hoc, on observable background variables. Large-scale observational datasets such as NLSY97 lend themselves particularly well to the application of PSM. NLSY97 collects information on a wide variety of variables, or covariates, for inclusion into the PSM model. Availability of a large pool of covariates from which to specify a PSM model is important because the better the model covers relevant factors for selection into treatment, the smaller the risk of hidden bias due to unobserved, potentially relevant concomitants. A further advantage of NLSY97 lies in the dataset's large sample size, which is a prerequisite for the use of PSM. A sufficiently large sample facilitates the identification of a suitable control group and renders the choice of a particular matching algorithm less critical, for all algorithms should yield equivalent outcomes with increasing sample size (Bryson et al., 2002; Caliendo & Kopeinig, 2005; Smith & Todd, 2005). In this study, a five-step process was followed to implement PSM:

1. Choosing the regression model.
2. Selecting covariates for the model.
3. Choosing a matching algorithm and performing the match.
4. Checking post-matching covariate balance.
5. Analyzing sensitivity to unobserved covariates.

The remainder of this section describes how each of these steps was implemented.

Choosing the Regression Model

PSM can be applied to single treatment or multiple treatment scenarios. Binomial logit or probit models are used for single treatment cases, whereas multiple treatment cases can be approached through either multinomial probit models or an independently carried-out sequence of binomial probit or logit models (Caliendo & Kopeinig, 2005). This study represented a multinomial treatment case because it involved two mutually exclusive treatment conditions, consisting of participation in either a CTE or college-preparatory high school curriculum. Each treatment condition was compared separately against the control condition (i.e., participation in a general high school curriculum). Students enrolled in a combined CTE/college-preparatory high school curriculum were not considered in this study because the number of these students (n = 74) was too small to allow for meaningful post-matching statistical procedures. Also, no inter-treatment comparison (i.e., a direct comparison between CTE and college-preparatory curricula) was performed because both groups were of similar size in the base sample. Since PSM requires a minimum control-to-treatment group ratio of approximately two-to-one (Stuart, 2007) to achieve balance across a given vector of covariates, only general-track students offered a sufficiently large sample size to function as an adequate control group.

Propensity scores were estimated for each individual treatment-control comparison through sequential application of a binomial logit model using the *MatchIt* (Ho, Imai, King, & Stuart, 2007b) module for *R packages*. The sequential application of binomial logit models was chosen because this approach is considered less complex and more robust than the use of multinomial probit regression (Lechner, 1999b). The particular binomial model choice (i.e., logit or probit) was immaterial because both approaches yield comparable results.

Selecting Covariates for the Model

The literature provides no clear guidance on selecting covariates for propensity score estimation. Recommendations run the gamut from including all available covariates (Dehejia & Wahba, 1999; Hahs-Vaughn & Onwuegbuzie, 2006; Leow et al., 2004; Rubin & Thomas, 1996) to specifying parsimonious models that exclude statistically insignificant covariates from the selection equation to increase the probability of successful matches (Augurzky & Schmidt, 2001). Whatever the chosen strategy, it is essential that (a) only those covariates be included in the model specification that influence treatment participation and outcome (Caliendo & Kopeinig, 2005; McCaffrey et al., 2004), and (b) the selection of covariates be guided by theory and prior empirical research (Bryson et al., 2002; Sianesi, 2004; Smith & Todd, 2005).

The selection of covariates for this study was guided by the (a) extant literature on factors that have been found to influence high school curriculum participation and educational attainment, and (b) availability of covariates in the NLSY97 dataset. Multicollinearity occurs when two or more predictor variables are highly correlated (Graham, 2003). High correlations among predictors inflate standard errors of the parameter estimate and may produce erroneous conclusions about the respective predictors' influence on the outcome variable. However, multicollinearity is not problematic when regression is used for prediction (Lipovetsky & Conklin, 2001), which is the fundamental purpose of conducting logistic regression in PSM. King (2007) noted that, unlike controlling for variables in linear regression, the inclusion of more highly collinear variables in the context of PSM does not reduce the efficiency of a given analysis due to relatively small effects on standard errors. The specific covariates selected for the PSM model are described in the Research Design section.

Choosing a Matching Algorithm and Performing the Match

Once a model has been specified with an appropriate vector of covariates, propensity scores are estimated and the match between suitable treatment and control cases is performed. Propensity scores are estimated through logistic regression. The estimation procedure results in a propensity score between *0* and *1* that expresses each participant's probability of selection into treatment given the covariate values from the underlying PSM model. Several algorithms, such as nearest-neighbor, full, exact, optimal, genetic, and stratification matching are available to match treatment group participants with one or more controls. With increasing sample size, all estimators should yield similar results (Bryson et al., 2002). However, in the case of smaller samples the choice of one particular matching algorithm over another can lead to greater variation in matching results (Heckman, Ichimura, & Todd, 1997). Zhao (2003) recommended the use of more than one matching method to verify that results remain consistent, independent of the use of a particular algorithm. Following Zhao's advice, nearest-neighbor and full matching were used. All matching procedures were conducted in *MatchIt* (Ho et al., 2007b).

Due to its efficiency and relative ease of implementation, nearest-neighbor matching is the most commonly used algorithm. When implemented *without* replacement, a given treatment participant is matched with his or her closest neighbor from within the control group based on the propensity score (Caliendo & Kopeinig, 2005). However, matching quality becomes an issue when considerable differences exist in the propensity score distributions of treatment and control groups. Nearest-neighbor matching *with* replacement overcomes this problem by allowing the same control group member to be matched to more than one treatment recipient (Smith & Todd, 2005). Additional measures aimed at increasing matching quality include the enforcement of common support restrictions and the use of calipers. The *region of common support* is the area where the density masses of the propensity score distributions overlap for participants (treatment) and non-participants (control). A high amount of overlap indicates a successful matching process. In all likelihood, not all treatment group participants will find suitable matches from within the pool of controls. Observations falling outside of the common support region are discarded because selection bias for such cases is undefined (Heckman, Ichimura, Smith, & Todd, 1998). A *caliper* delimits a maximum acceptable propensity score distance so that controls are matched with treatment recipients only if their respective propensity scores fall within defined caliper bounds. In the interest of preserving sample size, nearest-neighbor matching was conducted with replacement using a 5:1 control-to-treatment matching ratio. Furthermore, common support restrictions were enforced and caliper bounds were used with a caliper size of .25 times the standard deviation of the propensity scores (as recommended by Rosenbaum & Rubin, 1985). Detailed information about case loss due to common support and caliper restrictions used for nearest-neighbor matching is provided in Appendix C for the sample of CTE and general-track curricula, and in Appendix G for the sample of college-preparatory and general-track curricula.

Full matching is an approach that partitions a sample into non-overlapping subclasses within which close treatment and control units are matched (Rosenbaum, 1991). Each subclass contains a matched set of cases, whereby a treatment case can be matched to several control cases, or vice-versa. Full matching was chosen due to its high efficiency and limited case loss, since all treatment and control units are matched within the region of common support (Hansen, 2004; Hill et al., 2005). Full matching does not support the use of ratios and calipers. However, common support restrictions were enforced. Detailed information about case loss due to common support and caliper restrictions used for full matching is provided in Appendix C for the sample of CTE and general-track curricula, and in Appendix G for the sample of college-preparatory and general-track curricula.

Checking Post-matching Covariate Balance

A successful matching process occurs when the distribution of all observable covariates is balanced across the sample of treatment and control group participants. Traditionally, covariate balance has been assessed using *t*-tests for continuous and chi-square analysis for categorical covariates. However, recent landmark studies have highlighted several issues regarding this approach (see Ho et al., 2007a; Imai et al., 2008). Specifically, these studies described the *balance test fallacy* of using hypothesis tests to evaluate covariate balance by showing that (a) *t* and chi-square statistics depend on the number of control cases in a given sample, (b) the power of *t* and chi-square statistics decreases as more cases are dropped during the matching process, (c) hypothesis tests are irrelevant from a theoretical standpoint, for covariate balance is specific to a given sample instead of some hypothetical population, and (d) there is no threshold such as conventional significance levels below which the extent of remaining covariate imbalance is always acceptable. Visual checks of propensity score distributions and covariate balance using jitter and quantile-quantile (QQ) plots are considered preferred alternatives to traditional hypothesis tests. Jitter plots graphically display the propensity scores for matched and unmatched treatment and control cases. QQ-plots allow researchers to visually compare the empirical distribution of individual covariates before and after the matching process.

Even though visual covariate balance checks represent a state-of-the-art approach, the use of hypothesis tests continues to be prevalent in the overwhelming majority of PSM-based research studies. One likely reason might be that while visual checks are methodologically superior, the use of t and chi-square statistics is more readily understood by many applied researchers. Against this backdrop, post-matching covariate balance checks were conducted using both formal hypothesis tests and visual jitter/QQ-plots.

Analyzing Sensitivity to Unobserved Covariates

PSM operates under the assumption that bias can be removed and groups can be balanced based on a vector of *observable* covariates. However, propensity scores cannot account for hidden bias resulting from the effect of *unobservable*, yet causally-relevant, covariates (Reiter, 2000; Rosenbaum, 1998). The exclusion of influential covariates may lead to hidden bias in the propensity score estimation, for two individuals with the same observed covariate values will have unequal *actual* probabilities of receiving treatment (Heckman et al., 1997; Rosenbaum & Rubin, 1983, 1985). With continuous and dichotomous outcome variables, sensitivity analysis can be conducted as a final step in the PSM process to test the robustness of a given model against the potential impact of unobserved covariates. Rosenbaum (2002) proposed a bounding approach to assess the impact of unobserved covariates on treatment effects by calculating upper and lower bounds around a given test statistic. While bounding does not, per se, test for the existence of unobserved covariates, it allows researchers to gauge the degree to which effect estimates may be undermined by potential hidden bias (Becker & Caliendo, 2007). Sensitivity analysis for continuous outcome variables is based on the Wilcoxon (1945) signed-rank test, while the sensitivity of binary outcomes is assessed using the Mantel and Haenszel (1959) test statistic. In both cases, the strength of hidden bias is captured by the parameter γ, and where $\gamma = 1$, no hidden bias is assumed to be present. Increasing values of γ reflect increasing uncertainty about the impact of an unobserved covariate on the parameter estimate.

No designated test statistic or approach exists for multi-level categorical outcome variables. Thus, the present study's five-level educational attainment outcome measure was treated as a continuous variable for the purpose of sensitivity analysis. In the absence of methodological alternatives, this approach allowed for at least an approximate assessment of sensitivity to unobserved covariates. Sensitivity analysis was conducted using the *psmatch2* (Leuven & Sianesi, 2003) module for STATA 10. However, limitations in the software restricted sensitivity analysis to those analysis samples that were based on 5:1 nearest-neighbor matching.

Post-matching Analysis

The basic tenet of this study was to compare each of two specialized high school curricula (i.e., CTE, college-preparatory) to a general curriculum with regard to secondary and postsecondary educational attainment. The analysis strategy is illustrated in Table 3.8.

Table 3.8
Analysis Strategy

Analysis focus	*Treatment variable*	*Outcome variable*	*Procedure*
Determining curriculum effects on secondary and postsecondary educational achievement for CTE and general-track students	High school curriculum 0 = General 1 = CTE	Highest degree received as of 2007 (recoded) 0 = No high school diploma or GED 1 = GED 2 = Regular high school diploma 3 = Two-year college degree 4 = Four-year college degree	Chi-square analysis for main effects; adjusted standardized residuals analysis for post hoc test
Determining curriculum effects on secondary and postsecondary educational achievement for academic specialists and general track students	High school curriculum 0 = General 1 = Academic	Highest degree received as of 2007 (recoded) 0 = No high school diploma or GED 1 = GED 2 = Regular high school diploma 3 = Two-year college degree 4 = Four-year college degree	Chi-square analysis for main effects; adjusted standardized residuals analysis for post hoc test

Curriculum effects on secondary and postsecondary educational attainment were determined using chi-square analysis. Chi-square is the most widely used parametric statistic for the analysis of categorical data (Slavin, 1992). Categorical data include nominal and ordinal scales. Nominal data group observations into unordered, mutually exclusive categories, whereas ordinal data refer to rank-ordered observations that express degrees or incremental categories (Cohen, Manion, & Morrison, 2007). Levels of educational attainment were treated as ordinal data for the purpose of chi-square analysis, as they reflected a logical progression from not obtaining a high school diploma or GED to obtaining a four-year college degree.

In chi-square analysis, a set of observed frequencies in a two-dimensional table is compared to "a corresponding set of expected frequencies that are calculated by estimating cell values under the null hypothesis" (Rojewski & Bakeman, 1997, p. 171). Such two-dimensional frequency tables are commonly referred to as contingency tables (Wiersma, 1991). The null hypothesis holds that observed cell frequencies in a given contingency table do not deviate significantly from expected cell frequencies. The basic formula for chi-square analysis is expressed as

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where *O* represents observed cell frequencies and *E* represents expected cell frequencies (Cohen et al., 2007). As a nonparametric procedure, chi-square analysis does not require normally-distributed data (Mertler & Charles, 2005). Nonetheless, the approach rests on two critical assumptions, whereby (a) each observation can be associated with only a single cell in a given two-dimensional data matrix, and (b) no more than 20 percent of cells in a given contingency table should have an expected frequency count of less than five observations (Cohen et al., 2007; Field, 2005).

The omnibus chi-square hypothesis test can be used in two different ways. When applied as a test of homogeneity, it is used to assess the goodness of fit of a given variable relative to some theoretical distribution (Rojewski & Bakeman, 1997). When used as a test of independence, the omnibus test of main effects determines whether a relationship exists between categorical variables in a two-dimensional data matrix. In the latter case, the chi-square statistic is a measure of statistical significance that does not provide any information about the strength of association between two variables. Thus, a measure of effect size is needed to determine the practical importance of a statistically significant relationship (Kirk, 1996). In parametric analyses, Cohen's *d* is often used to assess the practical importance of a statistically significant omnibus test (Vacha-Haase &

Thompson, 2004). For nonparametric chi-square analysis, coefficient *phi* and Cramer's (1946) coefficient *V* represent appropriate statistics for measuring the strength of association between categorical variables (Wickens, 1989).

While the use of coefficient *phi* is limited to four-fold (2 x 2) contingency tables, Cramer's coefficient *V* extends to situations in which row and/or column variables have more than two levels (Field, 2005). In the present study, secondary and postsecondary educational attainment was determined using a 5 x 2 contingency table, whereby row values represented five levels of educational achievement (no high school diploma or GED, GED, regular high school diploma, two-year college degree, four-year college degree) and column values represented any one of two dichotomous curriculum comparisons (CTE and general; college-preparatory and general).

Cramer's coefficient *V* was used as a determinant of effect size. Cramer's coefficient *V* is calculated as

$$V = \sqrt{\frac{\chi^2}{n(M-1)}}$$

where χ^2 denotes the chi-square parameter estimate, *n* denotes sample size, and *M* equals the minimum number of rows or columns. The interpretation of Cramer's coefficient *V* is straightforward, for it is expressed as a scalar value between *0* and *1*. Cohen (1988) provided a rough classification scheme for Cramer's coefficient *V*, whereby for a five-row contingency table *V* values of equal or greater than .250 indicate a large effect size, *V* values between .150 and .249 indicate a moderate effect size, and *V* values below .150 indicate a small effect size.

Chi-square analysis determines whether the null hypothesis of no association between two categorical variables should be accepted or rejected. As such, the omnibus hypothesis test is limited to detecting overall significant deviations of observed cell frequencies from expected frequencies. No information is offered, however, about which individual cells contribute to a significant omnibus test. A frequent misuse of chi-square analysis lies in the fact that many researchers attempt to determine individual cell contributions through a subjective and unreliable inspection of individual cell frequencies (Thompson, 1988). Instead, formal post-hoc tests should be used to determine whether a deviation from expected frequencies in any given cell contributes to an overall significant omnibus chi-square test result.

Two frequently conducted post-hoc tests for chi-square analysis are based on Haberman's (1973) examination of residual values. Residuals are deviations between observed and expected frequency counts for each cell. In situations where row and/or column variables have multiple levels, cell-wise comparisons need to be carried out to evaluate individual cell contributions using either standardized residuals or adjusted standardized residuals. MacDonald and Gardner (2000) recommended the use of adjusted standardized residuals because the distributional characteristics of the statistic are unit normal. While the post-hoc analysis of residual values does not constitute a formal significance test, the meaningfulness of deviations is determined through the establishment of critical values (Rojewski & Bakeman, 1997). When post-hoc tests are conducted based on individual cells, appropriate adjustments are necessary to maintain the experimentwise Type I error rate (Beasley & Schumacker, 1995). Type I error denotes the possibility of rejecting a null hypothesis when the null hypothesis is in fact true (Wiersma, 1991). Type I error rates are determined *testwise*, meaning that significance levels are set for one hypothesis test only. In practice, most studies involve more than one test, which requires the establishment of an *experimentwise* error rate. Conducting multiple hypothesis tests raises the problem of simultaneous inference, which means that experimentwise Type I error rates are almost always higher than testwise Type I error rates (Wickens, 1989). To avoid experimentwise Type I error inflation, significance levels for multiple post-hoc tests should be adjusted by using the formula

$$\alpha_{adj} = 1 - (1 - \alpha)^{1/c}$$

where α denotes the chosen significance level and c refers to the total number of cells within the contingency table (Sidak, 1967). Adjusted significance levels are then used to calculate critical values for the evaluation of individual cell contributions. In the present study, chi-square post-hoc tests were carried out using the adjusted standardized residuals method as described by Gardner (2001). A testwise alpha level of .05 was adjusted using the Sidak method to maintain the experimentwise Type I error rate. Since the 5 x 2 contingency table for secondary and postsecondary educational achievement consisted of 10 cells, the Sidak procedure resulted in an adjusted alpha level of .005. Using a standard normal distribution table (McNemar, 1969), the resulting two-tailed critical value was determined as $z = \pm 2.80$. Any adjusted standardized residual exceeding this critical value in a given cell was considered significant.

CHAPTER 4

DATA ANALYSIS

The purpose of this study was to determine if differences existed in secondary and postsecondary educational attainment as a result of participation in different high school curricula. Specifically, my study addressed the following research questions:

1. Is there a statistically significant difference between CTE and general track students on secondary and postsecondary educational attainment?
2. Is there a statistically significant difference between college-preparatory and general track students on secondary and postsecondary educational attainment?

Verification of Covariate Balance

CTE and General-track Students

Multiple imputation of missing values resulted in the creation of five complete datasets for comparing CTE and general-track students. Each complete dataset was balanced on a defined vector of covariates using nearest-neighbor and full matching. Thus, the post-matching balance properties of 10 complete datasets had to be verified. Covariate balance checks were carried out using formal hypothesis tests and visual assessments based on jitter and quantile-quantile (QQ) plots. In each of the five imputation cycles, the pre-matching sample for CTE and general-track students exhibited statistically significant differences at the specified .05 alpha level on seven covariates, including survey weight, gender, race/ethnicity, attitudes toward school, number of school absences, number of school suspensions, and student-teacher ratio. Hypothesis test statistics for pre-matching differences by imputation cycle and matching algorithm for CTE and general-track students are provided in Appendix A.

Nearest-neighbor and full matching were highly efficient in balancing analysis samples across the entire vector of covariates. Hypothesis tests on the post-matching samples showed no statistically significant covariate differences between CTE (treatment) and general-track students (control). Table 4.1 illustrates the pooled percent improvement in the standardized mean difference across all five imputation cycles for the seven covariates with pre-matching significant differences. This value captures the percent improvement in the standardized mean difference on a given covariate between treatment and control groups before and after matching. The matching process yielded a balance improvement of between 77 and 90 percent across all covariates with significant pre-matching differences. A detailed overview of percent covariate balance improvement in standardized mean differences for each individual imputation cycle and matching method for CTE and general-track students is provided in Appendix B.

Table 4.1
Pooled Percent Standardized Mean Difference[a] Improvement for Covariates with Pre-matching Significant Differences for CTE and General-track Students by Matching Algorithm

	Matching algorithm	
Variable	*Nearest-neighbor*	*Full*
Survey weight	83.47	86.90
Gender	90.17	85.53
Race/ethnicity (dummy 1)	90.52	88.32
Attitudes toward school	82.72	90.47
Number of days absent from school	89.91	87.82
Ever suspended from school	82.09	78.95
Student-teacher ratio	84.62	77.05

Note. Race/ethnicity is a multi-level categorical variable that was dummy coded into two distinct covariates for propensity score estimation and matching. The significant pre-matching difference occurred for the first dummy-coded race/ethnicity covariate.
[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

The visual inspection of jitter and QQ plots corroborated the efficiency of both matching methods in drastically improving balance for covariates with significant pre-matching differences across treatment and control groups (see Appendix C for jitter plots for CTE and general-track students, and Appendix D for QQ plots for CTE and general-track students). The matching process was successful in producing comparable and equivalent treatment and control groups as a prerequisite for determining differences in secondary and postsecondary educational achievement between CTE and general-track students.

College-preparatory and General-track Students

Multiple imputation of missing values resulted in the creation of five complete datasets for the comparison of college-preparatory and general-track students. Each complete dataset was balanced on a defined vector of covariates using nearest-neighbor and full matching. Thus, the post-matching balance properties of 10 complete datasets had to be verified. Balance checks were carried out using formal hypothesis tests and visual assessments based on jitter and quantile-quantile (QQ) plots. In each of the five imputation cycles, the pre-matching sample for college-preparatory and general-track students exhibited statistically significant differences at the specified .05 alpha level on 15 covariates, including survey weight, gender, race/ethnicity, household poverty ratio, grades received in eighth grade, PIAT math standard scores, work-based learning participation, remedial English and/or mathematics participation, educational and/or physical handicap status, attitudes toward school, number of school absences, number of school suspensions, school type, and percentage of college-bound peers. Hypothesis test statistics for pre-matching differences by imputation cycle for college-preparatory and general-track students are provided in Appendix E.

Despite the very high number of background variables with significant pre-matching differences between CTE and general-track students, both nearest-neighbor and full matching successfully balanced the analysis sample across the entire vector of covariates. Hypothesis tests on the post-matching samples showed no statistically significant covariate differences between college-preparatory (treatment) and general-track students (control). Table 4.2 illustrates the pooled percent improvement in the standardized mean difference across all five imputation cycles for the 15 covariates with pre-matching significant differences. This value captures the percent improvement in the standardized mean difference on a given covariate between treatment and control groups before and after matching. The matching process yielded a balance improvement of between 70 and 98 percent across all covariates with significant pre-matching differences. A detailed overview of percent covariate balance improvements in standardized mean differences for each individual imputation cycle and matching method for college-preparatory and general-track students is provided in Appendix F.

Table 4.2
Pooled Percent Standardized Mean Difference[a] Improvement for Covariates with Pre-matching Significant Differences for College-preparatory and General-track Students by Matching Algorithm

	Matching algorithm	
Variable	*Nearest-neighbor*	*Full*
Survey weight	93.49	85.45
Gender	86.28	87.03
Race/ethnicity (dummy 1)	93.13	83.63
Race/ethnicity (dummy 2)	85.54	70.95
Household poverty ratio (square root)	89.82	86.41
Grades received in eighth grade	95.01	95.98
PIAT math standard score	95.37	97.15
Work-based learning participation	84.15	80.91
Remedial English and/or math	87.24	90.06
Educational and/or physical handicap	91.25	94.95
Attitudes toward school	93.27	82.93
Number of days absent from school	93.49	93.14
Ever suspended from school	98.09	98.54
School type	78.48	72.52
Percent peers college-bound	90.58	81.60

Note. Race/ethnicity is a multi-level categorical variable that had to be dummy coded into two distinct covariates for propensity score estimation and matching. Significant pre-matching differences occurred on both dummy-coded race/ethnicity covariates.
[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

The visual inspection of jitter and QQ plots corroborated the efficiency of both matching methods in drastically improving balance for covariates with significant pre-matching differences across treatment and control groups (see Appendix G for jitter plots for college-preparatory and general-track students, and Appendix H for QQ plots for college-preparatory and general-track students). The matching process was successful in producing comparable and equivalent treatment and control groups as a prerequisite for determining differences in secondary and postsecondary educational attainment between college-preparatory and general-track students.

Post-matching Analysis

CTE and General-track Students

Chi-square analysis was used to determine overall differences in secondary and postsecondary educational attainment between CTE and general-track students. After conducting a separate omnibus test for each of the five imputed datasets, pooled chi-square statistics, multiple imputation standard errors, and V effect size coefficients were calculated. The pooled chi-square parameter estimate indicated a significant association between high school curriculum type and overall educational attainment with a large effect size (see Table 4.3). Results were consistent across all imputation cycles and matching methods.

Table 4.3
Chi-square Omnibus Test for CTE and General-track Students by Imputation Cycle

Matching algorithm	*Imputation*	*n*	χ^2	*MI standard error*	*df*	*Cramer's V*
Nearest-neighbor	1	639	59.907^{***}		4	.306
	2	642	55.701^{***}		4	.295
	3	623	53.208^{***}		4	.292
	4	627	68.602^{***}		4	.331
	5	634	66.861^{***}		4	.325
	Pooled	633	60.856^{***}	6.625	4	.310
Full	1	700	62.337^{***}		4	.298
	2	706	45.638^{***}		4	.254
	3	703	54.594^{***}		4	.279
	4	701	79.324^{***}		4	.336
	5	702	53.779^{***}		4	.277
	Pooled	702	59.134^{***}	12.491	4	.289

Note. Sample size variations result from the use of matching ratios and corresponding ratio weights, as well as common support/caliper restrictions. Pooled sample sizes are rounded to the nearest integer.
MI standard errors are calculated using $\sqrt{\bar{U}+(1+1/m)B}$, where $\bar{U}$ denotes the within-imputation variance, m denotes the total number of imputation cycles, and B denotes the between-imputation variance.
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

While results led to the rejection of the null hypothesis of no association between high school curriculum type and overall educational attainment, the omnibus chi-square test did not provide any information about the particular educational attainment level(s) that may have contributed to the overall statistically significant result. Adjusted standardized residuals were examined in cell-wise post-hoc tests to determine which of the five educational attainment levels (no high school diploma or GED, GED, regular high school diploma, two-year college degree, four-year college degree) exhibited observed frequencies that deviated significantly from expected frequencies. Any adjusted standardized residuals that exceeded an adjusted critical value of $z = \pm 2.80$ within the 5 x 2 chi-square contingency table indicated a significant deviation from expected cell frequencies. A Sidak (1967) critical value adjustment was necessary to avoid experimentwise Type I error inflation through multiple post-hoc tests. Pooled results from cell-wise post-hoc tests are illustrated in Table 4.4, and detailed chi-square contingency tables by imputation cycle and matching algorithm for CTE and general-track students are provided in Appendix I.

Table 4.4

Cell-wise Post-hoc Test for CTE and General-track Students Using Pooled Adjusted Standardized Residuals

Matching algorithm	*Educational achievement*	*Curriculum type*	
		CTE	*General*
Nearest-neighbor	No HS diploma or GED	-4.57	4.57
	GED	-5.68	5.68
	Regular HS diploma	5.16	-5.16
	Two-year college degree	1.38	-1.38
	Four-year college degree	.84	-.84
Full	No HS diploma or GED	-4.42	4.42
	GED	-5.62	5.62
	Regular HS diploma	4.88	-4.88
	Two-year college degree	1.60	-1.60
	Four-year college degree	1.07	-1.07

Note. Pooled adjusted standardized residuals are obtained by calculating the arithmetic average of individual adjusted residual values. Adjusted standardized residual values exceeding an adjusted critical value of $z = \pm 2.80$ are considered significant.

The examination of adjusted standardized residuals indicated significant differences on all three levels of secondary educational attainment. A significantly greater number of general-track students never completed their secondary education when compared to CTE students. CTE students obtained a regular high school diploma at significantly higher rates than general-track students, who had significantly higher rates of earning a GED. Overall, these results translate into CTE students having significantly lower drop-out rates when compared to their general-track peers. Curriculum type had no significant effect on any level of postsecondary educational attainment. Results from post-hoc tests were consistent across both matching methods.

College-preparatory and General-track Students

Chi-square analysis was used to determine overall differences in secondary and postsecondary educational attainment between college-preparatory and general-track students. After conducting a separate omnibus test for each of the five imputed datasets, pooled chi-square statistics, multiple imputation standard errors, and *V* effect size coefficients were calculated. The pooled chi-square parameter estimate indicated a significant association between high school curriculum type and overall educational attainment with a medium effect size (see Table 4.5). Results were consistent across all imputation cycles and matching methods.

Table 4.5
Chi-square Omnibus Test for CTE and General-track Students by Imputation Cycle

Matching algorithm	*Imputation*	*n*	χ^2	*MI standard error*	*df*	*Cramer's V*
Nearest-neighbor	1	420	16.045**		4	.195
	2	425	18.266**		4	.207
	3	423	19.543**		4	.215
	4	424	13.077*		4	.176
	5	422	15.809**		4	.194
	Pooled	423	16.548**	2.485	4	.197
Full	1	620	17.306**		4	.167
	2	612	13.967**		4	.151
	3	610	17.442**		4	.169
	4	615	14.057**		4	.151
	5	617	14.912**		4	.155
	Pooled	615	15.537**	1.745	4	.159

Note. Sample size variations result from the use of matching ratios and corresponding ratio weights. Pooled parameter estimates are obtained by calculating the arithmetic average of individual parameter estimates. Pooled sample sizes are rounded to the nearest integer.
MI standard errors are calculated using $\sqrt{\bar{U}+(1+1/m)B}$, where $\bar{U}$ denotes the within-imputation variance, *m* denotes the total number of imputation cycles, and *B* denotes the between-imputation variance.
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

While results led to the rejection of the null hypothesis of no association between high school curriculum type and overall educational attainment, the omnibus chi-square test did not provide any information about the particular education level(s) that may have contributed to the overall statistically significant result. Adjusted standardized residuals were examined in cell-wise post-hoc tests to determine which of the five educational achievement levels (no high school diploma or GED, GED, regular high school diploma, two-year college degree, four-year college degree) exhibited observed frequencies that deviated significantly from expected frequencies. Any adjusted standardized residual that exceeded an adjusted critical value of $z = \pm 2.80$ within the 5 x 2 chi-square contingency table indicated a significant deviation from expected cell frequencies. A Sidak (1967) critical value adjustment was necessary to avoid experimentwise Type I error inflation through multiple post-hoc tests. Pooled results from cell-wise post-hoc tests are illustrated in Table 4.6, and detailed chi-square contingency tables by imputation cycle and matching algorithm for college-preparatory and general-track students are provided in Appendix J.

Table 4.6
Cell-wise Post-hoc Test for College-preparatory and General-track Students Using Pooled Adjusted Standardized Residuals

Matching algorithm	*Educational achievement*	*Curriculum type*	
		College-preparatory	*General*
Nearest-neighbor	No HS diploma or GED	-1.35	1.35
	GED	-2.31	2.31
	Regular HS diploma	-2.39	2.39
	Two-year college degree	.73	-.73
	Four-year college degree	3.19	-3.19
Full	No HS diploma or GED	-1.38	1.38
	GED	-2.26	2.26
	Regular HS diploma	-2.10	2.10
	Two-year college degree	.82	-.82
	Four-year college degree	2.96	-2.96

Note. Pooled adjusted standardized residuals are obtained by calculating the arithmetic average of individual adjusted standardized residual values. Adjusted standardized residual values exceeding an adjusted critical value of
$z = \pm 2.80$ are considered significant.

The examination of adjusted standardized residuals indicated a significant difference on the highest level of postsecondary attainment. The number of four-year college degrees completed by college-preparatory students was significantly higher when compared to their general-track counterparts. Curriculum type had no significant effect on educational attainment at any other level.

Sensitivity Analysis

While many influential predictors for curriculum choice and educational attainment were included in the present study's research design, additional relevant background factors may exist that were not captured in the propensity score model. Sensitivity analysis was conducted to determine the potential impact of unobserved covariates. The *psmatch2* (Leuven & Sianesi, 2003) module for STATA 10 was used to calculate Rosenbaum (2002) bounds for overall educational attainment. Due to software limitations, sensitivity analysis could only be conducted for those imputed analysis samples that were matched using 5:1 nearest-neighbor matching.

Sensitivity analysis corroborated the robustness of results for all five imputed samples of CTE and general-track students. Results were insensitive to hidden bias until the pooled odds of differential treatment assignment increased to 2.00 times, at which point the influence of CTE curricula on secondary and postsecondary academic attainment may have been overestimated. Results for the five imputed samples of college-preparatory and general-track students were also robust. Results were insensitive to hidden bias until the pooled odds of differential treatment assignment increased to 1.94 times, at which point the influence of college-preparatory curricula on secondary and postsecondary educational attainment may have been overestimated. Detailed sensitivity analysis statistics by imputation cycle are provided in Appendix K for CTE and general-track students, and in Appendix L for college-preparatory and general-track students.

CHAPTER 5

DISCUSSION

This chapter comprises three parts. The first part briefly reviews the study, including purpose and method. The second part discusses and relates findings to the extant literature. The third part describes implications, along with recommendations for future research.

Review

Introduction

Global competition is a cause for concern over current levels of secondary and postsecondary educational attainment in the U.S. Policymakers have issued severe warnings regarding the relative decline in domestic educational attainment levels when compared to other industrialized nations and newly emerging economies (see National Center on Education and the Economy, 2007; National Research Council, 2001). At the high school level, the dropout problem has grown into a *silent epidemic* (Bridgelan, DiIulio, & Morison, 2006), whereby roughly one-third of all secondary students leave school without a high school diploma (Barton, 2005). At the college level, the percentage of U.S. bachelor's degree holders is far below that of comparable nations, especially in the critical areas of mathematics and science (Douglass, 2006; OECD, 2007b; Snyder, Dillow, & Hoffman, 2009). Collectively, these attainment gaps threaten the competitiveness of the American workforce and highlight the need to actively promote, at a minimum, some form of postsecondary education and training (Rojewski, 2002).

A successful high school experience sets the course for subsequent postsecondary educational attainment. The nature and type of secondary curriculum is an important determinant of the high school experience. Most American high schools follow a three-tiered curriculum structure consisting of college-preparatory, career-technical education (CTE), and general tracks (Stone & Aliaga, 2005). This traditional tracking system is further divided into sub-groups that follow an ability continuum ranging from basic and regular to honors and advanced placement courses (Argys, Rees, & Brewer, 1996; Dornbusch, Glasgow, & Lin, 1996; Hallinan, 1994, 2004; Oakes, 1985, 1992). Even though curriculum tracking has been severely criticized for perpetuating racial and socioeconomic stratification (Burris & Welner, 2005; Burris, Wiley, Welner, & Murphy, 2008; Hoffman, 2003; Lucas, 1999; Rubin, 2003, 2006), the practice has remained widely in place due to support from teachers (Lee, Dedrick, & Smith, 1991; Riehl & Sipple, 1996) and parents of high-achieving students (Oakes & Guiton, 1995) who argue that instruction in homogeneous classes is more efficient and yields higher learning outcomes for all.

A plethora of studies have examined the effects of differential curriculum placement on educational outcomes, mostly focusing on academic achievement, high school completion, and postsecondary educational attainment. Based on the more stimulating academic climate and better school resources typically associated with college-preparatory curricula (Coleman, 1995; Crosby & Owens, 1993; Hallinan, 2003, 2004; Marsh & Raywid, 1994), students in such programs have consistently experienced more desirable educational outcomes than their peers in other tracks (Broussard & Joseph, 1998; Gamoran & Mare, 1989; Lee, Burkam, Chow-Hoy, Smerdon, & Geverdt, 1998; Lee, Croninger, & Smith, 1997; Natriello, Pallas, & Alexander, 1989). Much more controversy has arisen over the potential effects of curricula with a CTE focus. Positive outcomes, especially those related to secondary and postsecondary educational attainment, have been proposed by a number of researchers (e.g., Kulik, 1998; Plank, 2001; Plank, DeLuca, & Estacion, 2008; Rojewski, 1997; Stone & Aliaga, 2005), whereas no CTE curriculum effects have been ascertained by others (e.g., Agodini & Deke, 2004; Pittman, 1991).

The inadequate treatment of selection bias in the design of many studies is considered a fundamental cause for the divergent conclusions drawn from previous studies on high school curriculum effects (Lee & Ready, 2009). Against this backdrop, there is a strong need for studies that effectively control for selection bias and, therefore, allow a more rigorous re-assessment of curriculum effects on secondary and postsecondary educational attainment. The present study set out to meet this need using data from the National Longitudinal Survey of Youth 1997 (NLSY97, U.S. Bureau of Labor Statistics, 2009a).

The purpose of this study was to examine the effects of CTE and college-preparatory high school curricula on secondary and postsecondary educational attainment. Outcomes for participants in each of these specialized curriculum types were compared separately to those for individuals who completed a general high school curriculum. Educational attainment was defined as the highest level of formal education completed by an individual in the NLSY97 dataset as of 2007, the most recent year for which NLSY97 data were released at the time of writing. Given recurring debates over the resource intensity of secondary CTE (Cavanagh, 2005; Gray, 2004), educational attainment outcomes for individuals enrolled in CTE curricula were of particular interest.

Method

Using data from the NLSY97, this study analyzed a nationally representative cohort of students who were in ninth grade during the 1996/97 school year. The NLSY97 is a large-scale longitudinal dataset that provides a variety of student, family, and education-related background variables, as well as transcript-based high school curriculum information. As with other large-scale observational datasets, the analysis of NLSY97 data demanded that careful attention be given to the issues of nonresponse and selection bias. The presence of missing values due to nonresponse can result in reduced statistical power, difficulties in data analytic procedures that require complete data matrices, and biased analysis results due to the potential existence of systematic differences between missing and observed data (Barnard & Meng, 1999; Roth, Switzer, & Switzer, 1999; Schafer, 1997). Selection bias threatens the internal validity of causal-comparative designs and occurs when external factors affect selection into treatment and control conditions (Bryson, Dorsett, & Purdon, 2002; Torgerson & Roberts, 1999). Isolating causal treatment effects is not viable under either nonresponse or selection bias.

Of the final sample of 926 ninth graders, 327 (35.3%) exhibited missing values on one or more variables. Deleting over one-third of the sample would have resulted in a drastic loss of statistical power and the introduction of nonresponse bias into the analysis. Multiple imputation (MI; Rubin, 1987) was used to sidestep these issues. Five complete datasets were created in the MI process, where originally missing observations were replaced with slightly different plausible values in each individual complete dataset.

These plausible values are derived from a *Monte Carlo* approach, which is a general term for computational techniques that generate statistical results by repeating an artificially created chance process using random numbers (Barreto & Howland, 2006; Mooney, 1997).

Propensity score matching (PSM) was applied to each complete dataset in order to address the issue of selection bias. The PSM process creates comparable and equivalent treatment and control groups by balancing a given sample, post hoc, on observable background variables that influence both selection into treatment and the outcome of interest. The application of PSM to the present study resulted in different pairs of curriculum comparison groups (college-preparatory and general-track; CTE and general-track) that were balanced on all observable covariates. Successful balance across the entire vector of covariates was verified using formal hypothesis tests as well as visual assessments of quantile-quantile and jitter plots. The matching process was a prerequisite for the estimation of curriculum effects on secondary and postsecondary educational attainment without overt selection bias. Post-matching curriculum effects were determined using chi-square analysis in combination with appropriate post-hoc tests.

Discussion of Findings

CTE and General-track Curricula

Secondary and postsecondary educational attainment of individuals who were enrolled in a CTE high school curriculum was compared to those for individuals who were enrolled in a general high school curriculum. Members of the CTE group obtained regular high school diplomas at significantly higher rates when compared to their peers in the general-track group. General-track students, in turn, obtained GEDs at significantly higher rates when compared to their CTE counterparts. Data analysis determined a large effect size, indicating that attainment differences at the regular high school diploma and GED levels had practical relevance beyond mere statistical significance. However, the use of PSM rendered an explicit quantification of the curriculum effect in terms of percentages or absolute numbers unfeasible, since PSM was implemented using varying treatment-to-control group ratios. No attainment differences were found at two-year and four-year postsecondary levels.

Several studies have ascertained positive effects of CTE curricula on high school completion (Kulik, 1998; Plank, 2001), whereas others (Agodini & Deke, 2004; Pittman, 1991) have found no such effects for average students. The present study substantiates the notion that participation in CTE curricula has a substantial positive impact on regular high school diploma attainment. The issue of student engagement appears to be an important factor, for it is a critical determinant of high school completion and related educational outcomes (for a detailed review see Fredricks, Blumenfeld, & Paris, 2004). Since the application-focused nature of CTE offers increased opportunities for contextual learning (Advisory Committee for the National Assessment of Vocational Education, 2003), career-focused programs may have a higher potential than general curricula to keep students engaged in school by providing them with a sense of purpose throughout their high school experience. The present study provides evidence corroborating this notion, especially since (a) the impact of observable confounders was controlled by the research design, and (b) sensitivity analysis supported the robustness of the results with regard to unmeasured concomitants.

Individuals enrolled in a general curriculum obtained GEDs at significantly higher rates than their CTE counterparts. This finding is congruent with the higher dropout rate of general-track students. Obtaining a GED is a preferred outcome for individuals who had previously dropped out of high school because GED holders earn higher wages (Murnane, Willett, & Boudett, 1995; Tyler, Murnane, & Willett, 2000) and exhibit a higher likelihood to enroll in postsecondary education or training when compared to non-GED dropouts (Garet, Jing, & Kutner, 1996; Kroll & Qi, 1995). At the same time, those GED holders who pursue postsecondary education opportunities are much less likely to *finish* their degree when compared to regular high school completers (Cameron & Heckman, 1993). Besides lower postsecondary persistence, GED completers earn lower average incomes than regular high school graduates (Sum, 1996). It can thus be concluded that a GED is beneficial for dropouts, yet is overall less desirable when compared to completing a regular high school program (Ou, 2008).

The lack of an attainment difference between CTE and general-track students at the two and four-year college levels deserves closer attention. The Carl D. Perkins Vocational and Applied Technology Act of 1990 (hereafter referred to as Perkins II) was a major legislative initiative aimed at preparing students for both labor market entry *and* postsecondary education or training (Threeton, 2007). Perkins II resulted in specific career-technical initiatives, such as tech-prep and career academies. Tech-prep is based on articulation agreements between career-focused high school programs and community or vocational colleges whereby students can begin to earn credit toward a four-year college degree in a particular occupational area during the last two years of their high school career (National Center for Education Statistics, 2000). Career academies follow a school-in-school model that offers an in-depth career-focused curriculum with the objective to increase student achievement and promote postsecondary education (Cannon & Reed, 1999; Orr, 2005). Prior research has found positive effects of tech-prep (Bragg, Loeb, Gong, Deng, Yoo, & Hill, 2002; Riegg Cellini, 2006) and career-academies (Maxwell, 2001) on two and four-year postsecondary enrollment. While the present study neither measured initial college enrollment nor differentiated between various CTE interventions, overall results indicate that CTE concentrators did not have a postsecondary attainment advantage relative to their general-track peers. This outcome, which is congruent with existing research (see Silverberg, Warner, Fong, & Goodwin, 2004), points at policy-relevant issues regarding the implementation of Perkins II. The following section further explores these issues.

College-preparatory and General-track Curricula

Secondary and postsecondary educational attainment of individuals who were enrolled in a college-preparatory high school curriculum was compared to those for individuals who were enrolled in a general high school curriculum. The college-preparatory group exhibited significantly higher postsecondary attainment rates at the four-year college degree level when compared to the general curriculum group. Data analysis determined a medium effect size, indicating that attainment differences at the bachelor’s level were not just statistically significant but also practically relevant. However, the use of PSM rendered an explicit quantification of the curriculum effect in terms of percentages or absolute numbers unfeasible, since PSM was implemented using varying treatment-to-control group ratios. No attainment differences were found at two-year postsecondary, GED, or regular high school diploma levels.

This study provides evidence that, for this particular sample, focused college-preparatory high school curricula prepared students for successful four-year college careers at higher rates when compared to general curricula, which usually consist of a random amalgamation of *pseudo-academic* (Stone & Aliaga, 2005) courses. When considered from an input-output perspective, this outcome is unsurprising. It is well-established that the allocation of educational inputs, such as better teachers, more academic role models, and more engaging curriculum, distinctly favors students in college-preparatory programs (Coleman, 1995; Crosby & Owens, 1993; Hallinan, 2003, 2004; Marsh & Raywid, 1994; Oakes, 1992, 2008). While general-track curricula can produce college graduates as well, this study provides evidence for the notion that higher-quality inputs yield higher-quality outputs in terms of four-year college completion. As such, it is in line with findings from prior investigations on curriculum effects (Kulik, 1998; Vanfossen, Jones, & Spade, 1987).

The present findings may lead to premature conclusions about the positive educational attainment effects of college-preparatory high school curricula on academically-inclined students *in general*. Such conclusions, however, need to be qualified in light of one unanticipated outcome of the present analysis: the lack of positive college-preparatory curriculum effects at the regular high school diploma level. Students in college-preparatory curricula dropped out of high school at roughly the same rate as did students in general-track programs. Prima facie, this finding is somewhat startling given that the former group sported a significantly higher four-year college attainment rate. It is reasonable to expect that the same track-based resource allocation mechanisms that allegedly foster attainment at the bachelor’s degree level would provide similar advantages to college-preparatory students at the regular high school diploma level. Moreover, previous studies have linked participation in college-preparatory high school curricula to reduced dropout rates (see Bryk & Thum, 1989; Gamoran & Mare, 1989; Nyberg, McMillan, O'Neill-Rood, & Florence, 1997; Weber, 1988).

Issues related to research design and methodology may offer one possible explanation for the outcome differences between the present and prior studies. Except for Nyberg et al. (1997), whose longitudinal study roughly approximated what can be called an experimental design, prior investigations addressed nonresponse and selection bias to varying (and often inadequate) degrees. Insufficient consideration of analysis bias leads to (a) ignoring the ways in which nonresponse affects parameter estimates, and (b) disregarding the impact of unmeasured concomitants that distort analysis and ensuing conclusions. The strength of the present study lies in its principled approach to the treatment of nonresponse and selection bias. Based on this strength, the absence of positive college-preparatory curriculum effects on high school dropout should be considered a comparatively robust finding.

Besides differences in methodology, inefficient tracking mechanisms offer another possible explanation for the present findings. College-preparatory curricula are geared toward fostering scholastic achievement of academically able students. By the same token, students with applied talents, as well as those who may be at-risk of dropping out, may benefit less from the academically challenging nature of college-preparatory courses. While curriculum classifications in the NLSY97 dataset were based on students' course-taking patterns, myriad factors beyond academic achievement and standardized test scores influence track assignment, including teacher recommendations

(Oakes & Guiton, 1995), personal intentional choice (Delany, 1991), parental and peer influences (Kilgore, 1991; Useem, 1991), as well as organizational and logistical exigencies at the local school level (Garet & DeLany, 1988; Useem, 1992). Consequently, there is a real possibility that some low-achieving students within the sample who were enrolled in college-preparatory curricula may simply have been unable to keep up academically, leading to frustration and eventual dropout.

Finally, the connection between student engagement and dropout (see Ekstrom, Goertz, Pollack, & Rock, 1986; Manlove, 1998), which is a critical factor in interpreting the absence of curriculum effects on high school completion, requires closer examination. In a recent study (Bridgeland et al., 2006), *high-GPA* dropouts reported the perceived irrelevance of classes and the resulting disengagement from school as their primary reason for leaving school. The authors of the report concluded that even high-achieving students needed stronger support "to connect what they are learning in the classroom to the skills they will need in the workforce" (p. 4). Results from the present study substantiate this conclusion, which contradicts commonly-held beliefs about the positive engagement potential of college-preparatory tracks. School disengagement mechanisms seem to affect college-preparatory and general-track students to the same degree, regardless of differences in resource allocation and academic rigor. While college-preparatory tracks may stimulate some students, they alienate other potential high-achievers to the point of dropping out. Only those students who, in fact, persist during high school and pursue a four-year college career can reap the benefits of college-preparatory curricula when compared to their general-track peers. Even for those college-preparatory students who persist throughout high school it remains unclear whether they do so in an engaged manner, or whether they are simply more resilient in coping with a curriculum that is perceived by many as theoretical and quixotic. Overall, with regard to regular secondary educational attainment, findings from the present study qualify widely-held blanket assumptions about the advantages of college-preparatory curricula for *all* students.

Implications for Policy and Practice

Policy

CTE has long been afflicted with a stigma of educational inferiority due to a curriculum that is generally perceived as stunted and non-academic (Cohen & Besharov, 2002; Lynch, 2000). While this image is slowly changing as a result of major reform efforts, career-focused programs continue to be viewed by many as “a place for other people’s children” (Elliot, 2007, p. 5). One major goal of Perkins II legislation, whose provisions are directly relevant to this study’s sample, entailed improving the public image of career-oriented programs through the integration of secondary CTE and core academic courses. Another goal aimed at positioning secondary CTE more clearly as a direct pathway to postsecondary education and training. Results from the present study have implications for both goals.

Improving the public image of CTE. This study applied state-of-the-art methodology to examine the causal link between high school curriculum type and educational attainment. CTE emerged from this rigorous investigation as an effective high school curriculum with regard to student persistence. This positive result debunks longstanding myths that portray career-oriented high school programs as dumping grounds for the unmotivated and academically challenged. It is possible that the application-focused and context-based nature commonly associated with CTE curricula (see Advisory Committee for the National Assessment of Vocational Education, 2003) is more successful at keeping students in school by bestowing a sense of real-world relevance on the education process. The present study did not examine the specific degree to which Perkins II legislation was responsible for bringing about this positive outcome. What has become clear, however, is that continued debates over the resource intensity of secondary CTE programs (see Cavanagh, 2005; Gray, 2004) seem misplaced. Current and future policymakers would be well-advised to (a) rethink the role of CTE in U.S. public education, (b) provide the field with adequate resources for continuous improvement, and (c) use CTE as a strategic asset to boost high school completion rates as a critical precursor to postsecondary educational attainment.

CTE as a pathway to postsecondary education. While the positive impact of CTE on high school completion is highly encouraging, the absence of postsecondary curriculum effects is a cause for concern. Actual two and four-year college attainment for this study's sample of 1996/97 ninth graders does not correspond with the stated goals of Perkins II for promoting CTE as a pathway to postsecondary education. This recognition raises doubts over the successful implementation of Perkins II at state and local levels. In fact, the failure of Perkins II to produce increased postsecondary attainment among CTE students has direct policy implications for the subsequent Carl D. Perkins Vocational and Technical Education Act of 1998 (hereafter referred to as Perkins III) and the Carl D. Perkins Career and Technical Education Act of 2006 (hereafter referred to as Perkins IV). Since Perkins III and IV further intensified the notion of a seamless transition from secondary CTE to postsecondary education and training, future high school cohorts enrolled in CTE curricula can be reasonably expected to have positive educational attainment outcomes beyond high school completion. Yet, given findings from the present investigation it appears as if policymakers need to more carefully monitor the implementation and outcomes of Perkins III and IV legislation. An important part of such monitoring consists of regular assessments of curriculum effects that use rigorous, scientifically-based evaluation designs. The present study represents one example of such an assessment.

Practice

Teachers and guidance counselors. General-track students had the weakest educational attainment outcomes of the three curricula under consideration in this study. General-track students dropped out at significantly higher rates than CTE students, and obtained four-year college degrees at significantly lower rates than their peers in college-preparatory programs. Against this backdrop, one important implication for teachers and guidance counselors should be to steer students away from unstructured course-taking patterns and toward a CTE, college-preparatory, or dual concentration. Such an approach would likely yield higher overall attainment, for it would keep students from meandering through their high school careers without a clear objective or sense of purpose. Awareness of the curriculum-attainment link is particularly important for practitioners who have considerable influence on track placement decisions.

Students and parents. The most important implications of this study on high school curriculum effects are those for students themselves. Hopefully, cognizance of the ways in which different high school curricula impact secondary and postsecondary educational attainment will prompt students (and their parents) to actively participate in decisions over track placement. For college-bound students who are persistent enough to stay in school, college-preparatory curricula continue to be the best choice in terms of obtaining a bachelor's degree. For non-college-bound students, or those who are undecided about their postsecondary plans, CTE curricula lead to significantly higher rates of regular high school completion as an important precursor to further education and training. CTE+ curricula, although not examined in this study due to methodological/sample size constraints, hold the promise of providing students with improved postsecondary attainment opportunities for both two and four-year college degrees by combining academic rigor with engaging real-world applications of curriculum content. Overall, students who follow a structured high school career and keep clear of smorgasbord general-track curricula improve their chances of attaining the levels of education necessary to be competitive in a globalized labor market.

Recommendations for Future Research

Since the overwhelming majority of 1996/97 ninth graders in the present study graduated from high school during the 1999/2000 school year, it was not possible to determine the effects of CTE curricula beyond the provisions introduced by Perkins II. Future research should re-examine the curriculum-attainment link with regard to Perkins III and IV. The major objective of Perkins III was the preparation of students for postsecondary education by even more strongly focusing on the integration of core academic courses in CTE curricula. There is evidence in support of the notion that integrated CTE curricula, which are known as *CTE+* or *dual concentration*, are highly effective in bringing about positive educational attainment outcomes (DeLuca et al., 2006; Plank et al., 2008). CTE+ or dual concentrations are "comprised of students who follow both a rigorous academic sequence of courses and a rigorous sequence of CTE courses" (Stone & Aliaga, 2005, p. 127). In fact, recent studies found students enrolled in CTE+ curricula to be 2.7 times more likely to complete high school when compared to general-track students (Fletcher, 2009), and 1.7 times more likely to obtain a college degree when compared to regular CTE students (Novel, 2009).

The particular ratio between CTE and college-preparatory courses appears to play a critical role in secondary and postsecondary education outcomes. A survival analysis recently conducted on a survey of high school students suggested the existence of a curvilinear relationship between CTE participation and high school completion, whereby a 2:1 ratio of core academic-to-CTE courses was associated with minimizing the risk of dropping out (Plank et al., 2008). In contrast, CTE students with four or more CTE credits have a reduced likelihood of enrolling in college when compared to students without any CTE credits (Levesque, Laird, Hensley, Choy, Cataldi, & Hudson, 2008). Overall, the promising effects of integrated CTE+ models on educational attainment underscore the necessity to further evaluate the effectiveness of curriculum integration policies. If recent legislative initiatives that focus on secondary curriculum integration (as with Perkins III) and articulation-based secondary-to-postsecondary transition (as with Perkins IV) work as intended, future investigations of curriculum effects should be expected to ascertain measurable improvements in educational attainment levels.

A final, yet important, recommendation for future research entails the examination of particular programs and interventions within secondary CTE. Although the present study provides clear evidence for the overall positive impact of CTE curricula on high school completion, no data about specific occupational program areas were examined. Moreover, the present investigation did not evaluate the potential impact of substantive interventions within CTE, such as career academies, tech-prep, or work-based learning programs. It is possible that certain occupational program areas or substantive interventions produce more desirable educational attainment outcomes than others. Even though the literature features a variety of studies that assess such substantive interventions in a singular manner (see Kemple, 2001, 2004; Kemple & Snipes, 2000; Maxwell, 2001; Orr, Bailey, Hughes, Karp, & Keinzl, 2004 for career academies; Karp, Calcagno, Hughes, Jeong, & Bailey, 2007; Riegg Cellini, 2006 for tech-prep; Bennett, 2007; Brown, 2003; Gemici & Rojewski, 2010; Hughes, Bailey, & Mechur, 2001 for work-based learning programs), few investigations provide more comprehensive and unified evaluations (DeLuca et al., 2006; Neumark & Rothstein, 2006 are notable exceptions). Additional investigation would be useful to address this issue.

A Note to Researchers Analyzing Large-scale Datasets

Two methodological approaches, including multiple imputation (MI) and propensity score matching (PSM), were combined in the present study to mitigate the potential effects of missing data and selection bias. MI is a modern missing data imputation method, whereby missing observations are replaced with different plausible values from an assumed distribution (see Chapter 2 for detailed discussion of MI). Conceptually, MI deals with the missing data problem by creating a number of complete datasets that can be analyzed separately before parameter estimates and standard errors are pooled. While modern software packages automate the analysis and final pooling of results for many commonly used statistical procedures, this automation is not currently available when MI is used in combination with PSM. The lack of automation tremendously complicated the implementation of the present study's research design, for the complex PSM and post-matching procedures had to be carried out on a total of 20 individual datasets (two curriculum comparisons by five complete datasets by two matching algorithms) before pooling results and standard errors. While MI is a principled and highly efficient approach to dealing with nonresponse bias, researchers planning to use the method in combination with PSM should be aware of the extensive workload involved. Such researchers should contemplate using the expectation maximization (EM; Dempster, Laird, & Rubin, 1977) method as a viable alternative, as it results in one single complete dataset for analysis.

PSM is a highly efficient method for addressing selection bias by creating, post-hoc, comparable treatment and control groups based on observable background variables. PSM appears straightforward from a conceptual standpoint, yet its practical implementation is highly complex and presents researchers with numerous intricacies and decision points, including the treatment of sampling weights and the selection of particular matching algorithms. The process of preparing the raw NLSY97 dataset to the point where it could be used for PSM involved a great amount of recoding and related data manipulation. Another issue was the lack of one integrated statistical software suite that would perform all necessary procedures for MI, PSM, and post-matching analysis in a satisfactory manner. For example, data had to be transferred from SPSS to the *MatchIt* (Ho, Imai, King, & Stuart, 2007b) module in *R packages* to carry out the matching process, and from *MatchIt* to STATA to carry out sensitivity analysis as the final step of the PSM process. While such movement of data may be feasible with one dataset, handling 20 datasets proved to be less than practical. Overall, while combining MI and

PSM is a state-of-the-art approach in terms of addressing analysis bias, it is currently rather costly to implement in terms of time and labor. Given the speed of technological development, however, researchers should expect to see the seamless and comprehensive integration of these complex methods into standard statistical software packages hopefully sooner rather than later.

REFERENCES

Aakvik, A. (2001). Bounding a matching estimator: The case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics, 63*, 115-143.

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica, 74*, 235-267.

Abraham, W. T., & Russell, D. W. (2004). Missing data: A review of current methods and applications in epidemiological research. *Current Opinion in Psychiatry, 17,* 315-321.

Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation, 33*, 50-68.

Acemoglu, D. (1998). Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics, 113*, 1055-90.

Adler, M. J. (1940). *How to read a book: The art of getting a liberal education.* New York, NY: Simon and Schuster.

Advisory Committee for the National Assessment of Vocational Education. (2003). *Report of the advisory committee for the national assessment of vocational education*. Washington, DC: U.S. Department of Education.

Agodini, R., & Deke, J. (2004). *The relationship between high school vocational education and dropping out*. Princeton, NJ: Mathematica Policy Research.

Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. *Review of Economics and Statistics, 86*, 180-194.

Agodini, R., Uhl, S., & Novak, T. (2002). *Medium-run outcomes of secondary vocational students.* Washington, DC: U.S. Department of Education, National Assessment of Vocational Education.

Agodini, R., Uhl, S., & Novak, T. (2004). *Factors that influence participation in secondary vocational education*. Princeton, NJ: Mathematica Policy Research.

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, NJ: Prentice Hall.

Aiken, L. H., Smith, H. L., & Lake, E. T. (1994). Lower Medicare mortality among a set of hospitals known for good nursing care. *Medical Care, 32*, 771-787.

Alfeld, C., Hansen, D. M., Aragon, S. R., & Stone, J. R., (2006). Inside the black box: Exploring the value added by career and technical student organizations to students' high school experience. *Career and Technical Education Research*, *31*, 121-155.

Allison, P. D. (2002). *Missing data.* Thousand Oaks, CA: Sage.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology,* 112, 545-557.

Alvarez, D., & Mehan, H. (2007). Whole school de-tracking: A strategy for equity and excellence. *Theory into Practice, 45*, 82-89.

American Council on Education. (2010). *The content area GED tests.* Retrieved from http://www.acenet.edu/Content/NavigationMenu/ged/test/prep/The_Content_Area_Tes.htm

American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist, 57*, 1052-1059.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivational processes. *Journal of Educational Psychology, 80*, 260-267.

An, C. B., Haveman, R. H., & Wolfe, B. L. (1993). Teen out-of-wedlock births and welfare receipt: The role of childhood events and economic circumstances. *Review of Economics and Statistics, 75*, 195-208.

Angrist, J., & Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics, 106*, 979-1014.

Anstrom, K. J., & Tsiatis, A. A. (2001). Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data. *Biometrics, 57*, 1207-1218.

Anyon, J. (1988). Schools as agencies of social legitimization. In W. F. Pinar (Ed.), *Contemporary curriculum discourses* (pp. 175-200). Scottsdale, AZ: Gorush Scarisbrick.

Apple, M. W. 1982. *Education and power*. Boston, MA: Routledge & Kegan Paul.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243-277). Mahwah, NJ: Erlbaum.

Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management, 15*, 623-645.

Arum, R., & Shavit, Y. (1995). Secondary vocational education and the transition from school to work. *Sociology of Education, 68*, 187-204.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York, NY: Longmann.

Atkin, J. M., & Black, P. (1997). Policy perils of international comparisons. *Phi Delta Kappan, 79*, 22-28.

Augurzky, B., & Schmidt, C. M. (2001). *The propensity score: A means to an end* [Working Paper No. 271]. Heidelberg, Germany: University of Heidelberg, Institute for the Study of Labor.

Austin, P. C., Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. *Statistics in Medicine, 26*, 734-753.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*, 411-421.

Baker, D. P., & Wiseman, A. W. (2005). The worldwide explosion of internationalized education policy. In D. P. Baker & A. W. Wiseman (Eds.), *Global trends in educational policy* (pp. 1-22). Oxford, UK: Elsevier.

Baker, J. L. (2000). *Evaluating the impact of development projects on poverty: A handbook for practitioners.* Washington, DC: World Bank.

Balfanz, R. (2009). Can the American high school become an avenue of advancement for all? *The Future of Children, 19*, 17-36.

Barnard, J., & Meng, X. L. (1999). Applications of multiple imputation in medical studies: From AIDS to NHANES. *Statistical Methods in Medical Research, 8*, 7-36.

Barreto, H., & Howland, F. M. (2006). *Introductory econometrics: Using Monte Carlo simulation with Microsoft Excel*. New York, NY: Cambridge University Press.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics, 106*, 407-443.

Barro, R. J. (2001). Human capital and growth. *American Economic Review, 91*, 12-17.

Barro, R. J., & Lee, J. (2001). International data on educational attainment: Updates and implications. *Oxford Economic Papers, 53*, 541-563.

Barton, P. E. (2005). *One-third of a nation: Rising dropout rates and declining opportunities.* Princeton, NJ: Educational Testing Service, Policy Information Center.

Bauman, S. (2006). Using comparison groups in school counseling research: A primer. *Professional School Counseling, 9*, 357-366.

Beasley, T. M., & Schumacker, R. E. (1995). Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. *Journal of Experimental Education, 64*, 79-93.

Becker, G. S. (1964). *Human capital*. New York, NY: National Bureau of Economic Research.

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis with special reference to education*. Chicago, IL: University of Chicago Press.

Becker, S. O., & Caliendo, M. (2007). *Mhbounds: Sensitivity analysis for average treatment effects* [Discussion Paper No. 659]. Berlin, Germany: German Institute for Economic Research.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal, 2*, 358-377.

Behrman, J. R., Cheng, Y., & Todd, P. E. (2004). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *Review of Economics and Statistics, 86*, 108-132.

Benhabib, J., & Spiegel, M. M. (1994). The role of human capital in economic development: Evidence from aggregate cross-country data. *Journal of Monetary Economics, 34*, 143-173.

Benhabib, J., & Spiegel, M. M. (2005). Human capital and technology diffusion. In P. Aghion & S. N. Durlaud (Eds.), *Handbook of economic growth* (pp. 935-966). Amsterdam, The Netherlands: Elsevier.

Bennett, J. V. (2007). Work-based learning and social support: Relative influences on high school seniors' occupational engagement orientations. *Career and Technical Education Research, 32*, 187-214.

Berends, M. (1995). Educational stratification and students' social bonding to school. *British Journal of Sociology of Education, 16*, 327-351.

Berktold, J., Geis, S., & Kaufman, P. (1998). *Subsequent educational attainment of high school dropouts*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Binder, D. A. (1996). On variance estimation with imputed survey data: Comment. *Journal of the American Statistical Association, 91*, 510-512.

Birch, S., & Ladd, G. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61-79.

Black, D. A., & Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics, 121*, 99-124.

Bloom, H., Michalopoulos, C., Hill, C., & Lei, Y. (2002). *Can non-experimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* New York, NY: Manpower Demonstration Research Corporation.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside school. *Teachers College Record,110*, 608-645.

Bobbitt, F. (1918). *The curriculum*. Boston, MA: Riverside Press.

Boesel, D., Alsalam, N., & Smith, T. M. (1998). *Educational and labor market performance of GED recipients.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Boesel, D., Hudson, L., Deich, S., & Masten, C. (1994). *National assessment of vocational education: Final report to Congress.* Washington, DC: U.S. Department of Education.

Bradburn, E. M. (2002). *Short-term enrollment in postsecondary education: Student background and institutional differences in reasons for early departure, 1996-98*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Bradbury, K. L. (2002). Education and wages in the 1980s and 1990s: Are all groups moving up together? *New England Economic Review, Q1*, 19-46.

Braddock, J. H. (1990). *Tracking: Implications for student race-ethnic subgroups.* Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students.

Bragg, D., Loeb, J. W., Gong, Y., Deng, C., Yoo, J., & Hill, J. L. (2002). *Transition from high school to college and work for tech prep participants in eight selected consortia* (Report No. V051A990006). Washington, DC: Office of Vocational and Adult Education.

Breiman, L., Friedman, J., Olsen, R., & Stone, C. (1984). *Classification and regression trees.* Pacific Grove, CA: Wadsworth.

Bridgeland, J. M., DiIulio, J. J., & Morison, K. B. (2006). *The silent epidemic: Perspectives of high school dropouts.* Washington, DC: Civic Enterprises.

Broussard, A. C., & Joseph, A. L. (1998). Tracking: A form of educational neglect? *Social Work in Education, 20*, 110-120.

Brown, B. L. (2003). *CTE and work-based learning*. Columbus, OH: Clearinghouse on Adult Career and Vocational Education.

Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling, 4*, 287-316.

Brownstone, D., & Valletta, R. G. (1996). Modeling earnings measurement error: A multiple imputation approach. *The Review of Economics and Statistics, 78*, 705-717.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy, 52*, 781-861.

Bryk, A. S., & Thum, Y. M. (1989). The effects of school organization on dropping out: An exploratory investigation. *American Educational Research Journal, 26*, 353-383.

Bryson, A., Dorsett, R., & Purdon, S. (2002). *The use of propensity score matching in the evaluation of active labour market policies* [Working Paper No. 4]. London, England: Policy Studies Institute, National Centre for Social Research. Retrieved from http://www.dwp.gov.uk/asd/asd5/WP4.pdf

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: An examination of mediating process. *Developmental Psychology, 37*, 550-560.

Burris, C. C., & Welner, K. G. (2005). Closing the achievement gap by detracking. *Phi Delta Kappan, 86*, 594-598.

Burris, C. C., Wiley, E., Welner, K., & Murphy, J. (2008). Accountability, rigor, and detracking: Achievement effects of embracing a challenging curriculum as a universal good for all students. *Teachers College Record, 110*, 571-607.

Burtless, G. (1995). The case for randomized field trials in economic and policy research. *Journal of Economic Perspectives, 9*, 63-84.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Caliendo, M., Hujer, R., & Thomsen, S. (2005). *The employment effects of job creation schemes in Germany: A microeconometric evaluation* [Discussion Paper No. 1512]. Bonn, Germany: University of Bonn, Institute for the Study of Labor (IZA).

Caliendo, M., & Kopeinig, S. (2005). *Some practical guidance for the implementation of propensity score matching* [Discussion Paper No. 1588]. Bonn, Germany: University of Bonn, Institute for the Study of Labor.

Cameron, S. V., & Heckman, J. J. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics, 11*, 1-47.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago, IL: Rand McNally.

Cannon, D. G., & Reed, B. (1999). Career academies: Teaming with a focus. *Contemporary Education, 70*, 48-51.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1801-1863). Amsterdam, The Netherlands: Elsevier.

Carl D. Perkins Career and Technical Education Improvement Act, 20 U.S.C. § 2301 (2006).

Carl D. Perkins Vocational and Applied Technology Education Act, 20 U.S.C. § 2301 (1990).

Carl D. Perkins Vocational and Technical Education Act, 20 U.S.C. § 2301 (1998).

Carter, R. L. (2006). Solutions for missing data in structural equation modeling. *Research and Practice in Assessment, 1*, 1-6.

Castellano, M., Stringfield, S., & Stone III, J. R. (2003). Secondary career and technical education and comprehensive school reform: Implications for research and practice. *Review of Educational Research, 73*, 231-272.

Cataldi, E. F., Laird, J., & KewalRamani, A. (2009). *High school dropout and completion rates in the United States: 2007* (NCES 2009-064). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Cavanagh, S. (2005a). Finnish students are at the top of the world class. *Education Week, 24*, 8-8.

Cavanagh, S. (2005b). Vocational education's new job: Defend thyself. *Education Week, 24*, 1-18.

Chaplin, D. (1999). *GEDs for teenagers: Are there unintended consequences?* Washington, DC: Urban Institute.

Chase, R. S. (2002). Supporting communities in transition: The impact of the Armenian social investment fund. *World Bank Economic Review, 16*, 219-240.

Chaudhuri, A., & Stenger, H. (1992). *Survey sampling: Theory and methods*. New York, NY: Dekker.

Chen, J., & Shao, J. (2000). Nearest-neighbor imputation for survey data. *Journal of Official Statistics, 16*, 113-131.

Chen, J., & Shao, J. (2001). Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association, 96*, 260-269.

Christakis, N. A., & Iwashyna, T. J. (2003). The health impact of health care on families: A matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses. *Social Science and Medicine, 57*, 465-475.

Clark, M. A., & Jaeger, D. A. (2006). Natives, the foreign-born and high school equivalents: New evidence on the returns to the GED. *Journal of Population Economics, 19*, 769-793.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24*, 295-313.

Cochran, W. G., & Chambers, S. P. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society - Series A (General), 128*, 234-266.

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies. *Sankyha, 35*, 417-446.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York, NY: Routledge.

Cohen, M., & Besharov, D. J. (2002). *The role of career and technical education: Implications for the federal government.* Report prepared for the U.S. Department of Education, Office of Vocational and Adult Education. Retrieved from https://www.ed.gov/about/offices/list/ovae/pi/hs/besharov.doc

Coleman, J. (1995). Achievement-oriented school design. In M. T. Hallinan (Ed.), Restructuring schools: Promising practices and policies (pp. 11-30). New York, NY: Plenum.

Conant, J. (1959). *The transformation of the American high school*. Cambridge, MA: Harvard University Press.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Minnesota symposium on child psychology* (pp. 43-77). Chicago, IL: University of Chicago Press.

Cooper, H. M., Burger, J. M., & Seymour, G. E. (1979). Classroom context and student ability influences on teacher perceptions of classroom control. *American Educational Research Journal, 16*, 189-196.

Cox, D. R. (1970). *The analysis of binary data*. London, England: Methuen.

Cox, D. R., & Snell, D. J. (1989). *The analysis of binary data.* London, England: Chapman & Hall.

Crain, R., Allen, A., Thaler, R., Sullivan, D., Zellman, G., Little, J., & Quigley, D. (1999). *The effects of academic career magnet education on high schools and their graduates.* Berkeley, CA: University of California, National Center for Research in Vocational Education.

Cramer, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Crissey, S. R. (2009). *Educational attainment in the United States: 2007*. Washington, DC: U.S. Census Bureau.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391-418.

Crosby, M., & Owens, E. (1993). The disadvantages of tracking and ability grouping: A look at cooperative learning. Retrieved from Clemson University, National Dropout Prevention Center website: http://www.dropoutprevention.org/pubs/pdfs/SS05.pdf

Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics, 118*, 1495-1532.

D'Agostino, R. B. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*, 2265-2281.

De Anda, D. (2007). Intervention research and program evaluation in the school setting: Issues and alternative research designs. *Children and Schools, 29*, 87-94.

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94*, 1053-1062.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*, 151-161.

Delany, B. (1991). Allocation, choice, and stratification within high schools: How the sorting machine copes. *American Journal of Education, 99*, 181-207.

DeLuca, S., Plank, S., & Estacion, A. (2006). *Does career and technical education affect college enrollment?* St. Paul, MN: University of Minnesota, National Research Center for Career and Technical Education.

Demirtas, H. (2004). Modeling incomplete longitudinal data. *Journal of Modern Applied Statistical Methods,* 3, 305-321.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, 1-38.

Dewey, J. (1916). *Democracy and education.* New York, NY: Macmillan.

Dewey, J. (1948). *Reconstruction in philosophy.* Boston, MA: Beacon.

Diamond, A., & Sekhon, J. S. (2008). *Genetic matching for estimating causal effects: A general multivariate method for achieving balance in observational studies*. Retrieved from http://sekhon.berkeley.edu/papers/GenMatch.pdf

DiPrete, T. A., & Engelhardt, H. (2004). Estimating causal effects with matching methods in the presence and absence of bias cancellation. *Sociological Methods and Research, 32*, 501-528.

Dornbusch, S. M., Glasgow, K. L., & Lin, I. (1996). The social structure of schooling. *Annual Review of Psychology, 47*, 401-429.

Douglass, J. A. (2006). *The waning of America's higher education advantage: International competitors are no longer number two and have big plans in the global economy*. Berkeley, CA: University of California, Berkeley, Center for Studies in Higher Education.

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *Journal of General Psychology, 125*, 175-191.

Drost, W. H. (1967). *David Snedden and education for social efficiency*. Madison, WI: University of Wisconsin Press.

Du Bois, W. E. B. (1903). *The talented tenth.* Retrieved from http://teachingamericanhistory.org/library/index.asp?document=174

Eisner, E. (1974). *Conflicting conceptions of curriculum.* Berkeley, CA: McCutchan.

Ekstrom, R. B., Goertz, M. E., & Rock, D. A. (1988). *Education and American youth.* New York, NY: Falmer.

Ekstrom, R. B., Goertz, M. E., Pollack, J. M., & Rock, D. A. (1986). Who drops out of high school and why? Findings from a national study. *Teachers College Record, 87*, 356-373.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice*. Newbury Park, CA: Sage.

Elliot, J. (2007). The 2006 ACTER presidential address: The premier educational delivery system. *Career and Technical Education Research, 32*, 3-7.

Elliott, M. N., Edwards, C., Angeles, J., Habarsoomians, K., & Hays, R. D. (2005). Patterns of unit and item nonresponse in the CAHPS hospital survey. *Health Services Research, 40*, 2096-2119.

Ellis, A. K. (2004). *Exemplars of curriculum theory.* Larchmont, NY: Eye on Education.

Ellwood, D. T. (2001). *The sputtering labor force of the 21st century: Can social policy help?* Cambridge, MA: National Bureau of Economic Research.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, 8*, 128-141.

Enders, C. K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 313-342). Greenwich, CT: Information Age.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457.

Fenstermacher, G. D., & Soltis, J. F. (1992). *Approaches to teaching.* New York, NY: Teachers College Press.

Field, A. (2005). *Discovering statistics using SPSS*. London, England: Sage.

Fielding, S., Fayers, P. M., & Ramsay, C. R. (2009). Investigating the missing data mechanism in quality of life outcomes: A comparison of approaches. *Health and Quality of Life Outcomes, 7*, 1-10.

Finn, J. D. (1993). *School engagement and students at risk.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *Elementary School Journal, 95*, 421-454.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology, 82*, 221-234.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fisher, R. A. (1928). *Statistical methods for research workers*. London, England: Oliver & Boyd.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh, Scotland: Oliver & Boyd.

Fitzpatrick, K. M., & Yoels, W. C. (1992). Policy, school structure, and sociodemographic effects on statewide high school dropout rates. *Sociology of Education, 65*, 76-93.

Fletcher, E. C. (2009). *The relationship of high school curriculum tracks to degree attainment and occupational earnings* (Doctoral dissertation). Retrieved from http://etd.ohiolink.edu

Foster, J. C. (2007). Assessment as a tool to evaluate the benefits of CTE. *International Journal of Vocational Education and Training, 15*, 55-62.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109.

Freeman, C. E. (2004). *Trends in educational equity of girls and women: 2004.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Fuller, W. C., Manski, C. F., & Wise, D. A. (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources, 17*, 477-498.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction*. New York: Allyn and Bacon.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94*, 1146-1183.

Gamoran, A., & Weinstein, M. (1998). Differentiation and opportunity in restructured schools. *American Journal of Education, 106*, 385-415.

Gandini, L. (2004). Foundations of the Reggio Emilia approach. In J. Hendrick (Ed.), *Next steps toward teaching the Reggio way: Accepting the challenge to change* (pp. 13-26). Upper Saddle River, NJ: Pearson.

Gardner, R. C. (2001). *Psychological statistics using SPSS for Windows*. Upper Saddle River, NJ: Prentice Hall.

Garet, M. S., & DeLany, B. (1988). Students, courses, and stratification. *Sociology of Education, 61*, 61-77.

Garet, M. S., Jing, Z., & Kutner, M. (1996). *The labor market effects of completing the GED: Asking the right questions*. Washington, DC: American Institutes for Research, Pelavin Research Center.

Gartin, B. C., Murdick, N. L., Imbeau, M., & Perner, D. E. (2002). *How to use differentiated instruction with students with developmental disabilities in the general education program.* Arlington, VA: Council for Exceptional Children.

Gemici, S., & Rojewski, J. W. (2007). Examining research in career and technical education against scientifically-based research standards. *Career and Technical Education Research, 32*, 143-159.

Gemici, S., & Rojewski, J. W. (in press). *Contributions of cooperative education in preparing at-risk students for post-high school transition.*

Gemmel, N. (1996). Evaluating the impacts of human capital stocks and accumulation on economic growth: Some new evidence. *Oxford Bulletin of Economics and Statistics, 58*, 9-28.

Germann, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching, 25*, 689-703.

Giroux, H. A. (2006). *America on the edge.* New York, NY: Palgrave Macmillan.

Given, B. A., Keilman, L. J., Collins, C., & Given, C. W. (1990). Strategies to minimize attrition in longitudinal studies. *Nursing Research, 39*, 184-187.

Glazerman, S., Levy, D. M., & Myers, D. (2002). *Nonexperimental replications of social experiments: A systematic review* [Discussion Paper No. 8813-300]. Washington, DC: Corporation for the Advancement of Policy Evaluation.

Gleason, P., & Dynarski, M. (1998). *Do we know whom to serve? Issues in using risk factors to identify dropouts.* Princeton, NJ: Mathematica Policy Research.

Goldberg Dey, J., & Hill, C. (2007). *Behind the pay gap.* Washington, DC: American Association of University Women Educational Foundation.

Gordon, H. (2008). *The history and growth of career and technical education in America.* Long Grove, IL: Waveland.

Graham, J. W., & Donaldson, S. W. (1993). Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology*, *78*, 119-128.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31,* 197-218.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1-29). Thousand Oaks, CA: Sage.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology, 84*, 2809-2815.

Gray, K. (1991). Vocational education in high school: A modern phoenix? *Phi Delta Kappan, 71*, 437-445.

Gray, K. (2004). Is high school career and technical education obsolete? *Phi Delta Kappan, 86*, 128-134.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York, NY: Wiley.

Haberman, S. J. (1973). The analysis of residuals in cross-classified tables. *Biometrics, 29*, 205-222.

Hahs-Vaughn, D. L., & Onwuegbuzie, A. J. (2006). Estimating and using propensity score analysis with complex samples. *Journal of Experimental Education, 75*, 31-65.

Hallinan, M. T. (1994). Tracking: From theory to practice. *Sociology of Education, 67*, 79-84.

Hallinan, M. T. (2003). Ability grouping and student learning. *Brookings Papers on Education Policy, 1*, 95-124.

Hallinan, M. T. (2004). The detracking movement. *Education Next, 4*, 72-76.

Hammond, C. (2003). How education makes us healthy. *London Review of Education, 1*, 61-78.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association, 99*, 609-618.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*, 527-544.

Hanushek, E. A., & Woessmann, L. (2009). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation.* Cambridge, MA: National Bureau of Economic Research.

Hargreaves, D. H. (1967). *Social relations in a secondary school*. London, England: Routledge & Kegan Paul.

Harknett, K. (2006). Does receiving an earnings supplement affect union formation? Estimating effects for program participants using propensity score matching. *Evaluation Review, 30*, 741-778.

Harnish, D., & Lynch, R. L. (2005). Secondary to postsecondary technical education transitions: An exploratory study of dual enrollment in Georgia. *Career and Technical Information Research, 30*, 169-188.

Harriger, G. N. (1948). The college-preparatory curriculum as an instrument of educational guidance. *School Review, 56*, 163-167.

Haveman, R., Wolfe, B., & Spaulding, J. (1991). Childhood events and circumstances: Influencing high school completion. *Demography, 28*, 133-157.

Haviland, A., Nagin, D. S., & Rosenbaum, P. R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods, 12*, 247-267.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica, 66*, 1017-1098.

Heckman, J. J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies, 64*, 605-654.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies, 65*, 261-294.

Heckman, J. J., & LaFontaine, P. A. (2006). Bias corrected estimates of GED returns. *Journal of Labor Economics, 24*, 661-700.

Heckman, J. J., LaFontaine, P. A., & Rodriguez, P. L. (2008). *Taking the easy way out: How the GED testing program induces students to drop out.* Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics, 86*, 30-57.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives, 9*, 85-110.

Heubert, J. P., & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*, 151-179.

Hill, J. L. (2004, January). *Reducing bias in treatment effect estimation in observational studies suffering from missing data* [Working Paper]. New York, NY: Columbia University, School of International and Public Affairs, Institute for Social and

Economic Research and Policy. Retrieved from http://www.iserp.columbia.edu/research/working_papers/downloads/2004_01.pdf

Hill, J. L., Reiter, J. P., & Zanutto, E. L. (2007). A comparison of experimental and observational data analyses. In A. Gelman & X. L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete-data perspectives* (pp. 49-60). New York, NY: Wiley.

Hill, J. L., Waldfogel, J., Brooks-Gunn, J., & Han, W. J. (2005). Maternal employment and child development: A fresh look using newer methods. *Developmental Psychology, 41*, 833-850.

Hirsch, E. D. (1987). *Cultural literacy.* Boston, MA: Houghton Mifflin.

Hirschi, T. (1969). *The causes of delinquency*. Berkeley, CA: University of California Press.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007a). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Policy Analysis, 15*, 199-236.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007b). Matchit (2.4-11) [computer software]. Retrieved from http://gking.harvard.edu/matchit

Ho, M. S., & Jorgenson, D. W. (1995). *The quality of the U.S. work force, 1948-95*. Cambridge, MA: Harvard University, Kennedy School of Government. Retrieved from http://www.hks.harvard.edu/m-rcbg/ptep/laborjbes.pdf

Hoffman, N. (2003). College credit in high school: Increasing college attainment rates for underrepresented students. *Change, 35*, 43-48.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945-960.

Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies, 40*, 417-456.

Horton, N. J., & Lipsitz, S. R. (2001). Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician, 55,* 244-254.

Hossler, D., & Stage, F. K. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. *American Educational Research Journal, 29*, 425-451.

Hotz, V. J., Imbens, G. W., & Klerman, J. (2000). *The long-term gains from GAIN: A re-analysis of the impacts of the California GAIN program* [NBER Working Paper 8007]. Cambridge, MA: National Bureau of Economic Research.

Hughes, K. L., Bailey, T. R., & Mechur, M. J. (2001). *School-to-work: Making a difference in education. A research report to America.* New York, NY: Columbia University, Institute on Education and the Economy.

Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity, 34*, 331-351.

Hyland, T. (1993). Vocational reconstruction and Dewey's instrumentalism. *Oxford Review of Education, 19*, 89-100.

Hyslop-Margison, E. J. (2001). An assessment of the historical arguments in vocational education reform. *Journal of Career and Technical Education, 17*, 23-30.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association, 85*, 765-769.

Ibrahim, J. G., Chen, M., Lipsitz, S. R. (1999). *Monte Carlo EM for missing covariates in parametric regression models. Biometrics, 55*, 591-596.

Imai, K. (2005). Do get-out-the-vote calls reduce turnout? The importance of statistical methods for field experiments. *American Political Science Review, 99*, 283-300.

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society - Series A, 171*, 481-502.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*, 706-710.

Institute of Education Sciences. (2006). *What works clearinghouse evidence standards for reviewing studies*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/study_standards_final.pdf

Jarolimek, J. (1981). *The schools in contemporary society: An analysis of social currents, issues, and forces*. New York, NY: Macmillan.

Jimerson, S., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology, 38*, 525-549.

Jones, J. D., Vanfossen, B. E., & Ensminger, M. E. (1995). Individual and organizational predictors of high school track placement. *Sociology of Education, 68*, 287-300.

Jorgenson, D. W., & Fraumeni, B. M. (1992). Investment in education and U.S. economic growth. *Scandinavian Journal of Economics, 94*, 51-70.

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2002). *Information technology, education, and the sources of economic growth across U.S. industries.* Cambridge, MA: Harvard University, Kennedy School of Government.

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2003). Lessons from the U.S. growth resurgence. *Journal of Policy Modeling, 25*, 453-470.

Kablaoui, B. N., & Paulter, A. J., (1991). The effects of part-time work experience on high school students. *Journal of Career Development, 17*, 195-211.

Kalton, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics, 2*, 303-314.

Kamens, D. H., & McNeely, C. L. (2009). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review, 54*, 5-25.

Kao, G., & Tienda, M. (1998). Educational aspirations of minority youth. *American Journal of Education, 106*, 349-384.

Karp, M. M., Calcagno, J. C., Hughes, K. L., Jeong, D. W., & Bailey, T. R. (2007). *The postsecondary achievement of participants in dual enrollment: An analysis of student outcomes in two states*. St. Paul, MN: University of Minnesota, National Research Center for Career and Technical Education.

Kaufman, P., & Bradbury, D. (1992). *Characteristics of at-risk students in NELS:88*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Kelly, S., & Price, H. (2009). Vocational education: A clean slate for disengaged students? *Social Science Research, 38*, 810-825.

Kemple, J. J. (2001). *Career academies: Impacts on students' initial transitions to postsecondary education and employment*. New York, NY: Manpower Demonstration Research Corporation (MDRC).

Kemple, J. J. (2004, March). *Career academies: Impacts on labor market outcomes and educational attainment*. New York, NY: Manpower Demonstration Research Corporation.

Kemple, J. J., & Scott-Clayton, J. (2004). *Career academies: Impacts on labor market outcomes and educational attainment.* New York, NY: Manpower Demonstration Research Corporation.

Kemple, J. J., & Snipes, J. (2000). *Career academies: Impacts on students' engagement and performance in high school*. New York, NY: Manpower Demonstration Research Corporation.

Kenkel, D. (1991). Health behavior, health knowledge, and schooling. *Journal of Political Economy, 99*, 287-305.

KewalRamani, A., Gilbertson, L., Fox, M., & Provasnik, S. (2007). *Status and trends in the education of racial and ethnic minorities* (NCES 2007-039). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Kilgore, S. B. (1991). The organizational context of tracking in schools. *American Sociological Review, 56*, 189-203.

King, G. (2007, October 31). Re: Which variables, and how many, to match on? [Electronic mailing list message]. Retrieved from http://lists.hmdc.harvard.edu/lists/matchit/2007_10/msg00009.html

King, G. (2009, October 18). Re: Clarification on including survey weights [Electronic mailing list message]. Retrieved from http://lists.hmdc.harvard.edu/lists/matchit/2009_10/msg00002.html

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review, 95,* 49-69.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.

Kirkwood, B. R., Cousens, S. N., Victora, C. G., & De Zoysa, I. (1997). Issues in the design and interpretation of studies to evaluate the impact of community-based interventions. *Tropical Medicine and International Health, 2*, 1022-1029.

Kliebard, H. (1992). Constructing a history of the American curriculum. In P. W. Jackson (Ed.), *Handbook of research on curriculum: A project of the American Educational Research Association* (pp. 157-184). New York, NY: Macmillan.

Kliebard, H. (2004). *The struggle for the American curriculum: 1893-1958.* New York, NY: Taylor & Francis.

Kline, R. B. (1998). *Principles and practices of structural equation modeling.* New York, NY: Guilford.

Kmetic, A., Joseph, L., Berger, C., & Tenenhouse, A. (2002). Multiple imputation to account for missing data in a survey: Estimating the prevalence of osteoporosis. *Epidemiology, 13*, 437-444.

Kodrzycki, Y. K. (2002, June). *Educational attainment as a constraint on economic growth and social progress.* Paper presented at the Federal Reserve Bank of

Boston's 47th annual conference, Boston, MA. Retrieved from http://www.bos.frb.org/economic/conf/conf47/conf47d.pdf

Kofman, P., & Sharpe, I. G. (2003). Using multiple imputation in the analysis of incomplete observations in finance. *Journal of Financial Econometrics, 1*, 216-249.

Konstantopoulos, S. (2006). Trends of school effects on student achievement: Evidence from the NLS:72, HSB:82, and NELS:92. *Teachers College Record, 108*, 2550-2581.

Krathwohl, D. R., Bloom, B. S., & Masia, B. S. (1964). *Taxonomy of educational objectives. Handbook 2: Affective domain.* New York, NY: David McKay.

Kroll, B. S., & Qi, S. (1995). *A promise of empowerment: Results of the GED 1992 follow-up survey*. Washington, DC: American Council on Education, GED Testing Service.

Kulik, J. A. (1998). Curricular tracks and high school vocational education. In A. Gamoran (Ed.), *The quality of vocational education: Background papers from the 1994 national assessment of vocational education* (pp. 65-131). Washington, DC: U.S. Department of Education.

Laforgia, J. (1988). The affective domain related to science education and its evaluation. *Science Education, 72*, 407-421.

Laird, J., Cataldi, E. F., KewalRamani, A., & Chapman, C. (2008). *Dropout and completion rates in the United States: 2006*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Lakes, R. D. (1985). John Dewey's theory of occupations: Vocational education envisioned. *Journal of Vocational and Technical Education, 2*, 41-47.

Langdridge, D., Sheeran, P., & Connolly, K. J. (2007). Analyzing additional variables in the theory of reasoned action. *Journal of Applied Social Psychology, 37*, 1884-1913.

Laska, J. A. (1984). The relationship between instruction and curriculum: A conceptual clarification. *Instructional Science, 13*, 203-212.

Lechner, M. (1999a). Earnings and unemployment effects of continuous off-the-job training in East Germany after unification. *Journal of Business and Economic Statistics, 17*, 74-90.

Lechner, M. (1999b). *Identification and estimation of causal effects of multiple treatments under the conditional independence assumption* [Discussion Paper No. 91]. Bonn, Germany: University of Bonn, Institute for the Study of Labor.

Lechner, M. (2000). *A note on the common support problem in applied evaluation studies* [Discussion Paper No. 2001-01]. St. Gallen, Switzerland: University of St. Gallen, Forschungsgemeinschaft für Nationalökonomie an der Universität St. Gallen.

Lee, C. L., & Staff, J. (2007). When work matters: The varying impact of work intensity on high school dropout. *Sociology of Education, 80*, 158-178.

Lee, V. E., Burkam, D. T., Chow-Hoy, T., Smerdon, B. A., & Geverdt, D. (1998). *High school curriculum structure: Effects on coursetaking and achievement in mathematics for high school graduates* (Working Paper No. 98-09). Washington, DC: National Center for Education Statistics.

Lee, V. E., Croninger, R. G., & Smith, J. B. (1997). Course-taking, equity, and mathematics learning: Testing the constrained curriculum hypothesis in U. S. secondary schools. *Educational Evaluation and Policy Analysis, 19*, 99-121.

Lee, V. E., Dedrick, R. F., & Smith, J. B. (1991). The effect of the social organization of schools on teachers' efficacy and satisfaction. *Sociology of Education, 64*, 190-208.

Lee, V. E., & Ready, D. D. (2009). U.S. high school curriculum: Three phases of contemporary research and reform. *The Future of Children, 19*, 135-156.

Lee, V. E., & Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis, 19*, 205-227.

Legutko, R. (2008). A decade's difference: Research revisited on family influence of rural high school students' postsecondary decisions. *The Rural Educator, 29*, 4-7.

Leow, C., Marcus, S., Zanutto, E., & Boruch, R. (2004). Effects of advanced course-taking on math and science achievement: Addressing selection bias using propensity scores. *American Journal of Evaluation, 25*, 461-478.

LeTendre, G., Hofer, B., & Shimizu, H. (2003). What is tracking? Cultural expectations in the U.S., Germany, and Japan. *American Educational Research Journal, 40*, 43-89.

Leuven, E., & Sianesi, B. (2003). Psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. [computer software]. Boston, MA: Boston College.

Levesque, K. (2003). *Public high school graduates who participated in vocational/technical education: 1982-1998* (Report No. NCES 2003-24). Washington, DC: National Center for Education Statistics.

Levesque, K., Laird, J., Hensley, E., Choy, S. P., Cataldi, E. F., & Hudson, L. (2008). *Career and technical education in the United States: 1990 to 2005* (NCES 2008-035). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Lewis, T. (2007). Social inequality in education: A constraint on an American high-skills future. *Curriculum Inquiry, 37*, 329-349.

Li, Y., Whalley, J., Zhang, S., & Xiliang, Z. (2008). *The higher educational transformation of China and its global implications*. [NBER Working Paper W13849]. Cambridge, MA: National Bureau of Economic Research.

Lillard, D. R., & DeCicca, P. (2001). Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review, 20*, 459-473.

Linver, M. R., Davis-Kean, P., & Eccles, J. E. (2002, April). *Influences of gender on academic achievement.* Paper presented at the biennial meetings of the Society for Research on Adolescence, New Orleans, LA.

Lipovetsky, S., & Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry, 17*, 319-330.

Lipsitz, S. R., & Ibrahim, J. G. (1996). Using the EM algorithm for survival data with incomplete categorical covariates. *Lifetime Data Analysis, 2,* 5-14.

Little, R. J. (1988a). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198-1202.

Little, R. J. (1988b). Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics, 6*, 287-296.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York, NY: Wiley.

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research, 18*, 292-326.

Long, M. C. (2009). Changes in the returns to education and college quality. *Economics of Education Review*. Advance online publication. doi.org/10.1016/j.econedurev.2009.10.005

Loveless, T. (1999). Principles for future policy: The tracking and ability grouping debate. *Educational Horizons, 77*, 148-150.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*, 3-42.

Lucas, S. (1999). *Tracking inequality: Stratification and mobility in American high schools.* New York, NY: Teachers College.

Luellen, J., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review, 29*, 530-558.

Lynch, R. L. (2000). High school career and technical education for the first decade of the 21st century. *Journal of Vocational Education Research, 25*, 155-198.

MacDonald, P. L., & Gardner, R. C. (2000). Type I error rate comparisons of post hoc procedures for I x J chi-square tables. *Educational and Psychological Measurement, 60*, 735-754.

Mankiw, M. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics, 107*, 407-437.

Manlove, J. (1998). The influence of high school dropout and school disengagement on the risk of school-aged pregnancy. *Journal of Research on Adolescence, 82*, 187-220.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Marsh, R. S., & Raywid, M. A. (1994). How to make detracking work. *Phi Delta Kappan, 76*, 314-317.

Maxwell, N. L. (2001). Step to college: Moving from the high school career academy through the four-year university. *Evaluation Review, 25*, 619-654.

Mayer, A. (2008). Understanding how U.S. schools sort students for instructional purposes: Are all students being served equally? *American Secondary Education, 36*, 7-25.

Mayer, D. P., Mullens, J. E., & Moore, M. T. (2000). *Monitoring school quality: An indicators report.* Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9*, 403-425.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: Guilford.

McNeil, J. D. (1977). *Curriculum: A comprehensive introduction*. Boston, MA: Little and Brown.

McNemar, Q. (1969). *Psychological statistics*. New York, NY: Wiley.

Meeker, S. D. (2008). The voices of high school dropouts: Implications for research and practice. *The International Journal on School Disaffection, 6*, 40-52.

Mertler, C. A., & Charles, C. M. (2005). *Introduction to educational research*. Boston, MA: Allyn and Bacon.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2005). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.

Mickelson, R. A., & Velasco, A. E. (2006). Bring it on! Diverse responses to acting White among academically able Black adolescents. In E. McNamara Horvat & C. O'Connor (Eds.), *Beyond acting White: Reassessments and new directions in research on Black students and school success* (pp. 3-32). Boulder, CO: Rowman & Littlefield.

Minini, P., & Chavance, M. (2004). Sensitivity analysis of longitudinal binary data with non-monotone missing values. *Biostatistics, 5*, 531-544.

Molenberghs, G., Burzykowski, T., Michiels, B., & Kenward, M. G. (1999). Analysis of incomplete public health data. *Revue d'Epidemiologie et de Sante Publique, 47,* 499-514.

Monette, D. R., Sullivan, T. J., & DeJong, C. R. (2004). *Applied social research: A tool for the human services*. San Diego, CA: Harcourt College.

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Moore, L., Graham, A., & Diamond, I. (2003). On the feasibility of conducting randomized trials in education: Case study of a sex education intervention. *British Educational Research Journal, 29*, 673-689.

Moore, W., Pedlow, S., Krishnamurty, P., & Wolter, K. (2000). *National Longitudinal Survey of Youth 1997 (NLSY97): Technical sampling report*. Retrieved from http://www.nlsinfo.org/nlsy97/nlsdocs/nlsy97/TechnicalSamplingReport.pdf

Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research, 35*, 3-60.

Mosteller, F., Light, R. J., & Sachs, J. A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Education Review, 66*,797-842.

Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development, 52*, 13-30.

Murnane, R. J., Willett, J. B., & Boudett, K. P. (1995). Do high school dropouts benefit from obtaining a GED? *Educational Evaluation and Policy Analysis, 17*, 133-147.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 42*, 431-462.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692.

National Center for Education Statistics. (2000). *Vocational education in the United States: Toward the year 2000*. Retrieved from http://nces.ed.gov/pubs2000/2000029.pdf

National Center for Education Statistics. (2009). *High school transcript studies*. Retrieved from http://www.nces.ed.gov/surveys/hst/

National Center on Education and the Economy. (1990). *America's choice: High skills or low wages! The report of the commission on the skills of the American workforce*. Rochester, NY: Author.

National Center on Education and the Economy. (2007). *Tough choices or tough times: The report of the new commission on the skills of the American workforce*. San Francisco, CA: Jossey-Bass.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for reform*. Retrieved from http://www.ed.gov/pubs/NatAtRisk/risk.html

National Research Council. (2001). *Building a workforce for the information economy: Workforce needs in information technology*. Washington, DC: National Academies Press.

National Research Council and Institute of Medicine. (2004). *Engaging schools: Fostering high school students' motivation to learn.* Washington, DC: National Academies Press.

Natriello, G., McDill, E. L., & Pallas, A. M. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York, NY: Teachers College Press.

Natriello, G., Pallas, A. M., & Alexander, K. (1989). On the right track? Curriculum and academic achievement. *Sociology of Education, 62*, 109-118.

Nelson, R. R., & Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *American Economic Review, 56*, 69-75.

Neumark, D., & Rothstein, D. (2006). School-to-career programs and transitions to employment and higher education. *Economics of Education Review, 25*, 374-393.

Newman, J., Pradhan, M., Rawlings, L., Ridder, G., Coa, R., & Evia, J. L. (2002). An impact evaluation of education, health, and water supply investments by the Bolivian social investment fund. *World Bank Economic Review, 16*, 241-274.

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*, 280-312.

Newmann, F. M., & Wehlage, G. (1993). Five standards of authentic instruction. *Educational Leadership, 50*, 8-12.

Newmann, F., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11-39). New York, NY: Teachers College Press.

No Child Left Behind Act, 115 U.S.C. § 1425 (2001).

Novel, J. L. (2009). *Implementation of the Carl D. Perkins career-technical education reforms of the 1990s: Postsecondary education outcomes of students taking an enhanced vocational curriculum* (Doctoral dissertation). Retrieved from http://etd.ohiolonk.edu

Nuttall, J., Hollmen, L., & Staley, E. M. (2003). The effect of earning a GED on recidivism rates. *Journal of Correctional Education, 54*, 90-94.

Nyberg, K. L., McMillan, J. D., O'Neill-Rood, N., & Florence, J. M. (1997). Ethnic differences in academic retracking: A four-year longitudinal study. *Journal of Educational Research, 91*, 33-41.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and ability grouping on opportunities to learn mathematics and science.* Santa Monica, CA: Rand.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher, 21*, 12-21.

Oakes, J. (1994). More than misapplied technology: A normative and political response to Hallinan on tracking. *Sociology of Education, 67*, 84-89.

Oakes, J. (2008). Keeping track: Structuring equality and inequality in an era of accountability. *Teachers College Record, 110*, 700-712.

Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal, 32*, 3-33.

Oakes, J., & Lipton, M. (1992). Detracking schools: Early lessons from the field. *Phi Delta Kappan, 73*, 448-454.

Oakes, J., Selvin, M., Karoly, L., & Guiton, G. (1992). *Educational matchmaking: Academic and vocational tracking in comprehensive high schools*. Santa Monica, CA: Rand.

Oakes, J., & Wells, A. S. (1998). Detracking for high student achievement. *Educational Leadership, 55*, 38-41.

Oakes, J. M., & Rossi, P. H. (2003). The measurement of SES in health research: Current practice and steps toward a new approach. *Social Science and Medicine, 56*, 769-784.

Olaniyan, D. A., & Okemakinde, T. (2008). Human capital theory: Implications for educational development. *European Journal of Scientific Research, 24*, 157-162.

Oliner, S. D., & Sichel, D. E. (2000). The resurgence of growth in the late 1990s: Is information technology the story? *Journal of Economic Perspectives, 14*, 3-22.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research, 151*, 53-79.

Olson, J. M., & Zanna, M. P. (1993). Attitudes and attitude change. *Annual Review of Psychology*, *44*, 117-154.

Organization for Economic Cooperation and Development. (2003). *Education at a glance 2003.* Paris, France: Author.

Organization for Economic Cooperation and Development. (2007a). *PISA 2006 science competencies for tomorrow's world. Volume 1: Analysis.* Paris, France: Author.

Organization for Economic Cooperation and Development. (2007b). *Education at a glance 2007.* Paris, France: Author.

Organization for Economic Cooperation and Development. (2010). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes.* Paris, France: Author.

Orr, M. T. (2005). Career academies as a professionally engaging and supportive teaching experience. *Education and Urban Society, 37*, 453-489.

Orr, M. T., Bailey, T., Hughes, K. L., Karp, M. M., & Keinzl, G. S. (2004). *The National Academy Foundation's career academies: Shaping postsecondary transitions* (IEE Working Paper, No. 17). New York, NY: Institute for Education and the Economy.

Ou, S. (2008). Do GED recipients differ from graduates and school dropouts? Findings from an inner-city cohort. *Urban Education, 43*, 83-117.

Perez, A., Dennis, R. J., Gil, J. F. A., Rondon, M. A., & Lopez, A. (2002). Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia. *Statistics in Medicine, 21*, 3885-3896.

Perkins, R., Kleiner, B., Roey, S., & Brown, J. (2004). *The high school transcript study: A decade of change in curricula and achievement, 1990-2000.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*, 525-556.

Pintrich, P. R., & De Groot, E. (1990). Motivated and self-regulated learning components of academic performance. *Journal of Educational Psychology, 82*, 33-40.

Pittman, R. B. (1991). Social factors, enrollment in vocational/technical courses, and high school dropout rates. *Journal of Educational Research, 84*, 288-295.

Plank, S. B. (2001). A question of balance: CTE, academic courses, high school persistence, and student achievement. *Journal of Vocational Education Research, 26*, 279-327.

Plank, S. B., DeLuca, S., & Estacion, A. (2008). High school dropout and the role of career and technical education: A survival analysis of surviving high school. *Sociology of Education, 81*, 345-370.

Posner, G. J. (1992). *Analyzing the curriculum.* New York, NY: McGraw-Hill.

Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Boston, MA: Houghton Mifflin.

Prosser, C. A., & Allen, C. R. (1925). *Vocational education in a democracy*. New York, NY: Century.

Psacharopoulos, G., & Woodhall, M. (1997). *Education for development: An analysis of investment choice.* New York, NY: Oxford University Press.

Qian, Z., & Blair, S. (1999). Racial/ethnic differences in educational aspirations of high school seniors. *Sociological Perspectives, 42,* 605‑625.

Raaijmakers, Q. A. W. (1999). Effectiveness of different missing data treatments in surveys with Likert-type data: Introducing the relative mean substitution approach. *Educational and Psychological Measurement, 59*, 725-748.

Raghunathan, T. E. (2004). What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health, 25*, 99-117.

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27,* 85-95.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008 trends in academic progress.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Rao, J. N. K. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association, 91*, 499-506.

Rasinski, K. A., & Pedlow, S. (1994). Using transcripts to study the effectiveness of vocational education. *Journal of Vocational Education Research, 19*, 23-44.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education, 65*, 150-167.

Raymond, M. R. (1986). Missing data in evaluation research: *Evaluation and the Health Profession, 9*, 395-420.

Raymond, M. R., & Roberts, D. M. (1987). A comparison of methods for treating incomplete data in selection research. *Educational and Psychological Measurement, 47*, 13-26.

Reichel, M., & Ramey, M. A. (Eds.). (1987). *Conceptual frameworks for bibliographic education: Theory to practice.* Littleton, CO: Libraries Unlimited.

Reiter, J. (2000). Using statistics to determine causal relationships. *American Mathematical Monthly, 107*, 24-32.

Riegg Cellini, S. (2006). Smoothing the transition to college? The effect of Tech-Prep programs on educational attainment. *Economics of Education Review, 25*, 394-411.

Riehl, C., & Sipple, J. W. (1996). Making the most of time and talent: Secondary school organizational climates, teaching task environments, and teacher commitment. *American Educational Research Journal, 33*, 873-901.

Robeyns, I. (2006). Three models of education. *Theory and Research in Education, 4*, 69-84.

Rojewski, J. W. (1997). Effects of economic disadvantaged status and secondary vocational education on adolescent work experience and postsecondary aspirations. *Journal of Vocational and Technical Education, 14*, 43-59.

Rojewski, J. W. (2002). Preparing the workforce of tomorrow: A conceptual framework for career and technical education. *Journal of Vocational Education Research, 27*, 7-35.

Rojewski, J. W. (2009). A conceptual framework for technical and vocational education and training. In R. Maclean & D. N. Wilson (Eds.), *International handbook of education for the changing world of work* (pp. 19-39). New York, NY: Springer.

Rojewski, J. W., & Bakeman, R. (1997). Applying log-linear models to the study of career development and transition of individuals with disabilities. *Exceptionality, 7*, 169-186.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy, 98*, S71-S102.

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking.* New York, NY: Wiley.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics, 11*, 207-224.

Rosenbaum, P. R. (1991). A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society. Series B (Methodological), 53*, 597-610.

Rosenbaum, P. R. (1995). *Randomized experiments*. New York, NY: Springer.

Rosenbaum, P. R. (1998). Propensity score. In T. Colton & P. Armitage (Eds.), *Encyclopedia of biostatistics* (pp. 3551-3555). New York, NY: Wiley.

Rosenbaum, P. R. (2002). *Observational studies*. New York, NY: Springer.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 516-523.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*, 33-38.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*, 537-560.

Roth, P. L., & Switzer, F. S. (1995). A Monte Carlo analysis of missing data techniques in HRM settings. *Journal of Management, 21*,1003-1023.

Roth, P. L., Switzer, F. S., & Switzer, D. M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods, 2*, 211-232.

Rubin, B. (2003). Unpacking detracking: When progressive pedagogy meets students' social worlds. *American Educational Research Journal, 40*, 539-573.

Rubin, B. (2006). Tracking and detracking: Debates, evidence, and best practices for a heterogeneous world. *Theory into Practice, 45*, 4-14.

Rubin, B. (2008). Detracking in context: How local constructions of ability complicate equity-geared reform. *Teachers College Record, 110*, 646-699.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics, 2*, 1-26.

Rubin, D. B. (1978). Multiple imputation in sample surveys – a phenomenological Bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section*. American Statistical Association, 20-34.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91*, 473-489.

Rubin, D. B. (2006). *Matched sampling for causal effects*. New York, NY: Cambridge University Press.

Rubin, D. B., & Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association, 81*, 366-374.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics, 52*, 249-264.

Ruud, P. A. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics, 49*, 305-341.

Ryan, R. M., & Deci, E. L. (1996). When paradigms clash: Comments on Cameron and Pierce's claim that rewards do not undermine intrinsic motivation. *Review of Educational Research, 66*, 33-38.

Rylance, B. J. (1997). Predictors of high school graduation or dropping out for youths with severe emotional disturbances. *Behavioral Disorders, 23*, 5-17.

Saiz, A., & Zoido, E. (2005). Listening to what the world says: Bilingualism and earnings in the United States. *Review of Economics and Statistics, 87*, 523-538.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall.

Schafer, J. L. (1999a). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*, 3-15.

Schafer, J. L. (1999b). Norm (2.03) [computer software]. Retrieved from http://www.stat.psu.edu/~jls/misoftwa.html

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-571.

Schank, T., Schnabel, C., & Wagner, J. (2007). Do exporters really pay higher wages? First evidence from German linked employer–employee data. *Journal of International Economics, 72*, 52-74.

Schargel, F. P., & Smink, J. (2001). *Strategies to help solve our school dropout problem*. Larchmont, NY: Eye on Education.

Scheffler, I. (1995). John Dewey on work and education. In V. A. Howard & I. Scheffler (Eds.), *Work, education and leadership: Essays in the philosophy of education* (pp. 27-43). New York, NY: Peter Lang.

Schiro, M. S. (2008). *Curriculum theory: Conflicting visions and enduring concerns.* Thousand Oaks, CA: Sage.

Schmidt, C. M., Baltussen, R., & Sauerborn, R. (2000). *The evaluation of community-based interventions: Group-randomization, limits and alternatives* [Discussion Paper No. 206]. Bonn, Germany: University of Bonn, Institute for the Study of Labor (IZA).

Schubert, W. H. (1996). Perspectives on four curriculum traditions. *Educational Horizons, 74*, 169-176.

Schulte Nordholt, E. (1998). Imputation: Methods, simulation experiments and practical examples. *International Statistical Review, 66*, 157-180.

Schultz, T. W. (1963). *The economic value of education.* New York, NY: Columbia University Press.

Schultz, T. W. (1971). *Investment in human capital*. New York, NY: Free Press.

Schwartz, F. (1981). Supporting or subverting learning: Peer group patterns in four tracked schools. *Anthropology and Education Quarterly, 12*, 99-121.

Scott, J., & Sarkees-Wircenski, M. (2008). *Overview of career and technical education.* Homewood, IL: American Technical.

Seels, B., & Glasgow, Z. (1990). *Exercises in instructional design.* Columbus, OH: Merrill.

Sekhon, J. S., & Grieve, R. (2009). *A nonparametric matching method for covariate adjustment with application to economic evaluation.* Retrieved from http://sekhon.berkeley.edu/papers/GeneticMatching_SekhonGrieve.pdf

Shettle, C., Roey, S., Mordica, J., Perkins, R., Nord, C., Teodorovic, J., Brown, J., Lyons, M., Averett, C., & Kastberg, D. (2007). *The nation's report card: America's high* *school graduates*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics, 86*, 133-155.

Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*, 626-633.

Silverberg, M., Warner, E., Fong, M., & Goodwin, D. (2004). *National assessment of vocational education: Final report to Congress*. Washington, DC: U.S. Department of Education.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*, 317-329.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417-453.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effect of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571-581.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60*, 471-499.

Slavin, R. E. (1992). *Research methods in education*. Needham Heights, MA: Allyn and Bacon.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*, 15-21.

Smith, H. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology, 27*, 325-353.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics, 125*, 305-353.

Smits, N., Mellenbergh, G. J., & Vorst, H. C. M. (2002). Alternative missing data techniques to grade point average: Imputing unavailable grades. *Journal of Educational Measurement, 39*, 187-206.

Smyth, R. (2004). Exploring the usefulness of a conceptual framework as a research tool: A researcher's reflections. *Issues in Educational Research, 14*, 167-180.

Snyder, T. D., Dillow, S. A., & Hoffman, C. M. (2009). *Digest of education statistics 2008* (NCES 2009-020). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics, 39*, 312-320.

Sørensen, A. B., & Hallinan, M. T. (1978). A reconceptualization of school effects. *Sociology of Education, 50*, 273-289.

Sørensen, A. B., & Hallinan, M. T. (1986). Effects of ability grouping on growth in academic achievement. *American Educational Research Journal, 23*, 519-542.

Spring, J. (2009). *American education*. New York, NY: McGraw-Hill.

Stage, F., & Hossler, D. (1989). Differences in family influences on college attendance plans for male and female ninth graders. *Research in Higher Education, 30*, 301-315.

Stone, J. R., & Aliaga, O. A. (2005). Career and technical education and school-to-work at the end of the 20th century: Participation and outcomes. *Career and Technical Education Research, 30, 125-144.*

Stuart, E. (2007, November 6). Re: Am I ready for analysis? [Electronic mailing list message]. Retrieved from http://lists.hmdc.harvard.edu/lists/matchit/2007_11/msg00001.html

Sum, A. (1996). *Literacy in the labor force: Results from the National Adult Literacy Survey*. Washington, DC: National Center for Educational Statistics.

Swanson, C. B. (2004). Sketching a portrait of public high school graduation: Who graduates? Who doesn't? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 13-40). Cambridge, MA: Harvard Education Publishing Group.

Swanson, C. B. (2009). *Closing the graduation gap: Educational and economic conditions in America's largest cities*. Bethesda, MD: Educational Projects in Education Research Center.

Switzer, F. S., Roth, P. L., & Switzer, D. M. (1998). A Monte Carlo analysis of systematic data loss in an HRM setting. *Journal of Management*, *24*, 763-779.

Thompson, B. (1988). Misuse of chi-square contingency-table test statistics. *Educational and Psychological Research, 8*, 39-49.

Threeton, M. D. (2007). The Carl D. Perkins Career and Technical Education (CTE) Act of 2006 and the roles and responsibilities of CTE teachers and faculty members. *Journal of Industrial Teacher Education, 44*, 66-82.

Titus, M. A. (2007). Detecting selection bias, using propensity score matching, and estimating treatment effects: An application to the private returns to a Master's degree. *Research in Higher Education, 48*, 487-521.

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.

Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in education. *British Journal of Educational Studies, 49*, 316-328.

Torgerson, D. J., & Roberts, C. (1999). Understanding controlled trials: Randomization methods - concealment. *British Medical Journal, 319*, 375-376.

Trautwein, U., Lüdtke, O., Marsh, H., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*, 788-806.

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management, 24*, 53-62.

Turner, J. C., Thorpe, P. K., & Meyer, D. K. (1998). Students' reports of motivation and negative affect: A theoretical and empirical analysis. *Journal of Educational Psychology,* 90, 758-771.

Tyler, J. H. (2003). The economic benefits of the GED: Lessons from recent research. *Review of Educational Research, 73*, 369-403.

Tyler, J. H., & Lofstrom, M. (2008). *Is the GED an effective route to postsecondary education for school dropouts?* Cambridge, MA: National Bureau of Economic Research.

Tyler, J. H., & Lofstrom, L. (2009). Finishing high school: Alternative pathways and dropout recovery. *Future of Children, 19*, 77-103.

Tyler, J. H., Murnane, R. J., & Willett, J. B. (2000). Estimating the labor market signaling value of a GED. *Quarterly Journal of Economics, 115*, 431-468.

Tyler, J. H., Murnane, R. J., & Willett, J. B. (2003). Who benefits from a GED? Evidence for females from high school and beyond. *Economics of Education Review, 22*, 237-247.

United Nations. (2007). *Principles and recommendations for population and housing censuses: Revision 2*. Retrieved from http://unstats.un.org/unsd/demographic/sources/census/docs/P&R_%20Rev2.pdf

U.S. Bureau of Labor Statistics. (2005). *NLS handbook*. Retrieved from http://www.bls.gov/nls/handbook/nlshndbk.htm

U.S. Bureau of Labor Statistics. (2009a). *National longitudinal survey of youth 1997* [Data file and code book]. Retrieved from https://www.nlsinfo.org/investigator/pages/welcome.jsp

U.S. Bureau of Labor Statistics. (2009b). *NLS web investigator: Custom weights*. Retrieved from http://www.nlsinfo.org/old-web-investigator/custom_weights.php

U.S. Department of Education. (1999). *1998 revision of the secondary school taxonomy*. Washington, DC: Author.

U.S. Department of Education. (2008). *A nation accountable: Twenty-five years after a nation at risk*. Washington, DC: Author.

U.S. Department of Labor. (1991). *What work requires of schools: A SCANS report for America 2000.* Washington, DC: Author.

Useem, E. L. (1991). Student selection into course sequences in mathematics: The impact of parental involvement and school policies. *Journal of Research on Adolescence, 1*, 231-250.

Useem, E. L. (1992). Getting on the fast track in mathematics: School organizational influences on math track assignment. *American Journal of Education, 92*, 325-352.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*, 473-481.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16*, 219-242.

Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine, 18*, 681-694.

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, K., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76*, 1049-1064.

Van Buuren, S., & Oudshoorn, C. G. M. (2000). Multiple imputation by chained equations (1.0) [computer software]. Retrieved from http://www.stefvanbuuren.nl/downloads.html

Vanfossen, B., Jones, J., & Spade, J. (1987). Curriculum tracking and status maintenance. *Sociology of Education, 60*, 104-122.

Van Houtte, M., & Stevens, P. A. J. (2009). Study involvement of academic and vocational students: Does between-school tracking sharpen the difference? *American Educational Research Journal, 46*, 943-973.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., & Rahman, T. (2009). *Achievement gaps: How black and white students in public schools perform in mathematics and reading on the national assessment of educational progress*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Vauras, M., Rauhanummi, T., Kinnunen, R., & Lepola, J. (1999). Motivational vulnerability as a challenge for educational interventions. *International Journal of Educational Research*, *31*, 515-531.

Verma, R., & Goodale, J. C. (1995). Statistical power in operations management research. *Journal of Operations Management, 13*, 139-152.

Vinovskis, M. A. (1999). *The road to Charlottesville: The 1989 education summit.* Ann Arbor, MI: University of Michigan, School of Public Policy. Retrieved from http://govinfo.library.unt.edu/negp/reports/negp30.pdf

Wallace, L. (1996). Changes in demographics and motivations of distance education in students. *Journal of Distance Education, 11*, 1-31.

Washington, B. T. (1903). *Industrial education for the negro.* Retrieved from http://www.teachingamericanhistory.org/library/index.asp?document=62

Wells, A., & Serna, I. (1996). The politics of culture: Understanding local political resistance to detracking in racially mixed schools. *Harvard Educational Review, 66*, 93-118.

Weber, J. M. (1988). Dropout rates for students in different high school curricula: Another look. *Journal of Vocational Education Research, 13*, 35-49.

Werblow, J., & Duesbery, L. (2009). The impact of high school size on math achievement and dropout rate. *High School Journal, 92*, 14-23.

White, M., & Lakey, J. (1992). *The restart effect: Evaluation of a labour market programme for unemployed people*. London, England: Policy Studies Institute.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Wiersma, W. (1991). *Research methods in education*. Needham Heights, MA: Allyn and Bacon.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80-83.

Wilde, E. T., & Hollister, R. (2002). *How close is close enough? Testing nonexperimental estimates of impact against experimental estimates of impact with education test scores as outcomes* [Discussion Paper No. 1242-02]. Madison, WI: Institute for Research on Poverty.

Wirt, J., Snyder, T., Sable, J., Choy, S. P., Bae, Y., Stennett, J., Gruner, A., & Perie, M. (1998). *The condition of education 1998* (NCES 98-013). Washington, DC: U.S. Department of Education.

Witte, A., & Tauchen, H. (1994). Work and crime: An exploration using panel data. *Public Finance, 49*, 155-167.

Wolfe, B., & Zuvekas, S. (1997). Nonmarket outcomes of schooling. *International Journal of Educational Research, 27*, 491-501.

Yamada, T., Yamada, T., & Kang, J. M. (1991). *Crime rates versus labor market conditions: Theory and time-related evidence*. Cambridge, MA: National Bureau of Economic Research.

Yonezawa, S., Wells, A. S., & Serna, I. (2002). Choosing tracks: "Freedom of choice" in detracking schools. *American Educational Research Journal, 39*, 37-67.

Zhao, Z. (2003). *Data issues of using matching methods to estimate treatment effects: An illustration with NSW data set*. [Unpublished manuscript]. Beijing, China: Peking University, China Center for Economic Research. Retrieved from http://www.nd.edu/~meg/MEG2004/Zhao-Zhong.pdf

Zhou, X. H., Eckert, G. J., & Tierney, W. M. (2001). Multiple imputation of public health research. *Statistics in Medicine, 20,* 1541-1549.

APPENDIX A

COVARIATE BALANCE HYPOTHESIS TESTS FOR CTE AND GENERAL-TRACK STUDENTS

Table A1

Imputation 1 - Differences Between CTE (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	228064.87	82530.358	203063.32	87646.771	-3.764***	720		0.29
Gender	1.37	.485	1.51	.500		1	12.576***	-0.28
Race/ethnicity (dummy 1)	.22	.413	.32	.465		1	7.901**	-0.23
Race/ethnicity (dummy 2)	.16	.371	.22	.416		1	3.453	-0.15
Urbanicity	.71	.453	.72	.448		1	.086	-0.02
Household poverty ratio (square root)	15.26	5.698	14.49	6.369	-1.634	720		0.13
Grades received in eighth grade	5.14	1.651	5.25	1.618	.927	720		-0.07
PIAT math standard score	92.87	13.118	91.51	14.368	-1.255	720		0.10
Work-based learning	.18	.384	.16	.370		1	.318	0.05
Remedial English and/or math	.17	.375	.19	.390		1	.409	-0.05
ESL and/or bilingual program	.07	.260	.11	.312		1	2.527	-0.14
Educational and/or physical handicap	.09	.284	.06	.239		1	1.842	0.11
Attitudes toward school	15.88	2.686	16.45	2.880	2.632**	720		-0.20
Number of days absent from school	4.07	4.415	6.25	8.633	3.807***	720		-0.32
Ever suspended from school	.30	.458	.40	.491		1	8.184***	-0.21
School type	1.03	.182	1.06	.231		1	1.779	-0.14
Student-teacher ratio	2.23	1.065	2.40	1.077	2.007*	720		-0.16
Percent peers college-bound	3.44	.944	3.37	1.043	-.917	720		0.07

[a]$\underline{n} = 262$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_t + \sigma^2_c / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A2

Imputation 1 - Differences Between CTE (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	226475.46	83078.599	223409.02	83859.452	-.455	638		0.04
Gender	1.38	.487	1.40	.491		1	.215	-0.04
Race/ethnicity (dummy 1)	.22	.417	.22	.418		1	.005	0.00
Race/ethnicity (dummy 2)	.17	.375	.18	.389		1	.261	-0.03
Urbanicity	.71	.455	.70	.457		1	.026	0.02
Household poverty ratio (square root)	15.28	5.746	15.16	5.960	-.268	638		0.02
Grades received in eighth grade	5.12	1.659	5.17	1.690	.368	638		-0.03
PIAT math standard score	92.64	13.107	93.03	14.039	.354	638		-0.03
Work-based learning	.18	.382	.19	.394		1	.251	-0.03
Remedial English and/or math	.17	.379	.17	.379		1	.002	0.00
ESL and/or bilingual program	.07	.263	.08	.270		1	.025	-0.04
Educational and/or physical handicap	.09	.281	.08	.264		1	.251	0.04
Attitudes toward school	15.97	2.650	15.80	2.806	-.754	638		0.06
Number of days absent from school	4.14	4.449	4.05	4.249	-.267	638		0.02
Ever suspended from school	.31	.462	.28	.449		1	.479	0.07
School type	1.04	.185	1.03	.177		1	.082	0.06
Student-teacher ratio	2.25	1.065	2.24	1.044	-.139	638		0.01
Percent peers college-bound	3.43	.948	3.45	.988	.293	638		-0.02

[a]$\underline{n} = 255$ [b]$\underline{n} = 385$ [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_t + \sigma^2_c / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A3

Imputation 1 - Differences Between CTE (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	d^c
Survey weight	227330.41	82715.614	223842.47	83435.089	-.536	698		0.04
Gender	1.38	.486	1.40	.490		1	.294	-0.04
Race/ethnicity (dummy 1)	.22	.415	.22	.418		1	.018	0.00
Race/ethnicity (dummy 2)	.17	.373	.18	.382		1	.134	-0.03
Urbanicity	.71	.454	.71	.452		1	.012	0.00
Household poverty ratio (square root)	15.29	5.703	15.00	6.016	-.634	698		0.05
Grades received in eighth grade	5.14	1.658	5.05	1.750	-.638	698		0.05
PIAT math standard score	92.73	13.130	92.46	13.313	-.261	698		0.02
Work-based learning	.18	.383	.20	.399		1	.410	-0.05
Remedial English and/or math	.17	.376	.19	.389		1	.285	-0.05
ESL and/or bilingual program	.07	.261	.07	.258		1	.023	0.00
Educational and/or physical handicap	.08	.279	.06	.245		1	1.132	0.08
Attitudes toward school	15.92	2.664	15.84	2.746	-.388	698		0.03
Number of days absent from school	4.11	4.425	3.83	4.180	-.845	698		0.07
Ever suspended from school	.30	.460	.32	.466		1	.150	-0.04
School type	1.03	.183	1.03	.177		1	.046	0.00
Student-teacher ratio	2.24	1.063	2.26	1.066	.199	698		-0.02
Percent peers college-bound	3.44	.948	3.43	.993	-.111	698		0.01

[a]n = 259 [b]n = 441 [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c}} / 2$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A4

Imputation 2 - Differences Between CTE (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	d^c
Survey weight	228064.87	82530.358	203063.32	87646.771	-3.764***	720		0.29
Gender	1.37	.485	1.51	.500		1	12.576***	-0.28
Race/ethnicity (dummy 1)	.22	.413	.32	.465		1	7.901**	-0.23
Race/ethnicity (dummy 2)	.16	.371	.22	.416		1	3.453	-0.15
Urbanicity	.72	.451	.73	.444		1	.139	-0.02
Household poverty ratio (square root)	15.14	5.533	14.40	6.449	-1.569	720		0.12
Grades received in eighth grade	5.16	1.620	5.25	1.608	.784	720		-0.06
PIAT math standard score	92.65	13.384	91.21	14.634	-1.309	720		0.10
Work-based learning	.18	.384	.17	.374		1	.169	0.03
Remedial English and/or math	.17	.375	.19	.390		1	.409	-0.05
ESL and/or bilingual program	.07	.260	.11	.312		1	2.527	-0.14
Educational and/or physical handicap	.09	.284	.06	.239		1	1.842	0.11
Attitudes toward school	15.88	2.686	16.46	2.887	2.647**	720		-0.21
Number of days absent from school	4.11	4.469	6.14	8.590	3.557***	720		-0.30
Ever suspended from school	.30	.458	.40	.491		1	8.184**	-0.21
School type	1.03	.182	1.06	.231		1	1.779	-0.14
Student-teacher ratio	2.23	1.072	2.39	1.076	1.997*	720		-0.15
Percent peers college-bound	3.44	.940	3.37	1.043	-.883	720		0.07

[a]n = 262 [b]n = 460 [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c}} / 2$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A5
Imputation 2 - Differences Between CTE (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227042.75	82985.787	225882.71	82500.328	-.174	640		0.01
Gender	1.38	.487	1.40	.491		1	.225	-0.04
Race/ethnicity (dummy 1)	.22	.416	.22	.413		1	.012	0.00
Race/ethnicity (dummy 2)	.17	.374	.18	.384		1	.152	-0.03
Urbanicity	.72	.452	.73	.444		1	.149	-0.02
Household poverty ratio (square root)	15.19	5.548	15.43	6.568	.490	640		-0.04
Grades received in eighth grade	5.15	1.623	5.13	1.683	-.123	640		0.01
PIAT math standard score	92.39	13.363	92.06	14.676	-.284	640		0.02
Work-based learning	.18	.387	.18	.385		1	.001	0.00
Remedial English and/or math	.17	.377	.17	.380		1	.009	0.00
ESL and/or bilingual program	.07	.262	.07	.261		1	.003	0.00
Educational and/or physical handicap	.09	.280	.09	.292		1	.117	0.00
Attitudes toward school	15.93	2.672	15.83	2.658	-.455	640		0.04
Number of days absent from school	4.16	4.493	3.98	4.293	-.502	640		0.04
Ever suspended from school	.30	.461	.29	.455		1	.117	0.02
School type	1.04	.184	1.04	.194		1	.067	0.00
Student-teacher ratio	2.25	1.068	2.29	1.070	.506	640		-0.04
Percent peers college-bound	3.44	.930	3.47	1.004	.382	640		-0.03

[a]$\underline{n} = 257$ [b]$\underline{n} = 385$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A6
Imputation 2 - Differences Between CTE (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227877.64	82633.153	225406.13	81545.732	-.387	704		0.03
Gender	1.38	.485	1.39	.488		1	.168	-0.02
Race/ethnicity (dummy 1)	.22	.414	.23	.424		1	.219	-0.02
Race/ethnicity (dummy 2)	.16	.372	.16	.366		1	.033	0.00
Urbanicity	.72	.452	.72	.450		1	.006	0.00
Household poverty ratio (square root)	15.18	5.516	15.25	6.119	.158	704		-0.01
Grades received in eighth grade	5.16	1.621	5.13	1.750	-.248	704		0.02
PIAT math standard score	92.61	13.394	92.66	14.581	.053	704		0.00
Work-based learning	.18	.385	.18	.386		1	.004	0.00
Remedial English and/or math	.17	.375	.16	.371		1	.024	0.03
ESL and/or bilingual program	.07	.260	.07	.262		1	.004	0.00
Educational and/or physical handicap	.08	.278	.10	.305		1	.688	-0.07
Attitudes toward school	15.91	2.657	15.87	2.715	-.175	704		0.01
Number of days absent from school	4.12	4.474	3.86	4.218	-.774	704		0.06
Ever suspended from school	.30	.459	.28	.450		1	.259	0.04
School type	1.03	.183	1.05	.222		1	1.125	-0.10
Student-teacher ratio	2.23	1.071	2.27	1.053	.434	704		-0.04
Percent peers college-bound	3.44	.941	3.52	.996	1.035	704		-0.08

[a]$\underline{n} = 261$ [b]$\underline{n} = 445$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A7

Imputation 3 - Differences Between CTE (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	228064.87	82530.358	203063.32	87646.771	-3.764^{***}	720		0.29
Gender	1.37	.485	1.51	.500		1	12.576^{***}	-0.28
Race/ethnicity (dummy 1)	.22	.413	.32	.465		1	7.901^{**}	-0.23
Race/ethnicity (dummy 2)	.16	.371	.22	.416		1	3.453	-0.15
Urbanicity	.72	.449	.73	.444		1	.069	-0.02
Household poverty ratio (square root)	15.29	5.713	14.42	6.388	-1.822	720		0.14
Grades received in eighth grade	5.13	1.627	5.28	1.609	1.270	720		-0.09
PIAT math standard score	92.78	13.254	91.37	14.582	-1.293	720		0.10
Work-based learning	.18	.384	.17	.372		1	.237	0.03
Remedial English and/or math	.17	.375	.19	.390		1	.409	-0.05
ESL and/or bilingual program	.07	.260	.11	.312		1	2.527	-0.14
Educational and/or physical handicap	.09	.284	.06	.239		1	1.842	0.11
Attitudes toward school	15.88	2.686	16.45	2.884	2.599^{**}	720		-0.20
Number of days absent from school	4.10	4.443	6.24	8.650	3.733^{***}	720		-0.31
Ever suspended from school	.30	.458	.40	.491		1	8.184^{**}	-0.21
School type	1.03	.182	1.06	.231		1	1.779	-0.14
Student-teacher ratio	2.23	1.070	2.40	1.077	2.062^{*}	720		-0.16
Percent peers college-bound	3.45	.932	3.37	1.043	-1.040	720		0.08

[a]$\underline{n} = 262$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A8

Imputation 3 - Differences Between CTE (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227075.60	83021.465	221124.52	84818.985	-.869	620		0.07
Gender	1.38	.486	1.39	.489		1	.131	-0.02
Race/ethnicity (dummy 1)	.22	.416	.24	.425		1	.163	-0.05
Race/ethnicity (dummy 2)	.17	.374	.18	.383		1	.122	-0.03
Urbanicity	.72	.450	.73	.447		1	.029	-0.02
Household poverty ratio (square root)	15.30	5.728	14.95	5.679	-.749	620		0.06
Grades received in eighth grade	5.14	1.633	5.22	1.670	.573	620		-0.05
PIAT math standard score	92.75	13.241	92.62	14.878	-.115	620		0.01
Work-based learning	.18	.384	.18	.385		1	.003	0.00
Remedial English and/or math	.17	.374	.17	.379		1	.030	0.00
ESL and/or bilingual program	.07	.262	.07	.262		1	.000	0.00
Educational and/or physical handicap	.09	.280	.08	.274		1	.023	0.04
Attitudes toward school	15.93	2.662	15.87	2.798	-.271	620		0.02
Number of days absent from school	4.17	4.455	3.72	3.837	-1.340	620		0.11
Ever suspended from school	.30	.461	.28	.450		1	.424	0.04
School type	1.04	.184	1.04	.184		1	.002	0.00
Student-teacher ratio	2.25	1.069	2.28	1.057	.317	620		-0.03
Percent peers college-bound	3.44	.930	3.45	1.033	.126	620		-0.01

[a]$\underline{n} = 257$ [b]$\underline{n} = 365$ [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A9

Imputation 3 - Differences Between CTE (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227877.64	82633.153	225443.90	84233.506	-.373	701		0.03
Gender	1.38	.485	1.39	.489		1	.229	-0.02
Race/ethnicity (dummy 1)	.22	.414	.21	.410		1	.032	0.02
Race/ethnicity (dummy 2)	.16	.372	.18	.383		1	.224	-0.05
Urbanicity	.72	.450	.74	.441		1	.248	-0.04
Household poverty ratio (square root)	15.32	5.696	15.38	5.776	.131	701		-0.01
Grades received in eighth grade	5.13	1.629	5.22	1.680	.672	701		-0.05
PIAT math standard score	92.74	13.264	93.36	14.150	.571	701		-0.05
Work-based learning	.18	.385	.19	.394		1	.161	-0.03
Remedial English and/or math	.17	.375	.18	.384		1	.117	-0.03
ESL and/or bilingual program	.07	.260	.07	.259		1	.000	0.00
Educational and/or physical handicap	.08	.278	.09	.284		1	.032	-0.04
Attitudes toward school	15.91	2.657	15.88	2.729	-.146	701		0.01
Number of days absent from school	4.11	4.447	3.69	3.803	-1.316	701		0.10
Ever suspended from school	.30	.459	.28	.451		1	.206	0.04
School type	1.03	.183	1.03	.171		1	.139	0.00
Student-teacher ratio	2.24	1.069	2.28	1.042	.535	701		-0.04
Percent peers college-bound	3.45	.933	3.46	1.023	.112	701		-0.01

[a]$\underline{n} = 261$ [b]$\underline{n} = 442$ [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c}} / 2$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A10

Imputation 4 - Differences Between CTE (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	228064.87	82530.358	203063.32	87646.771	-3.764***	720		0.29
Gender	1.37	.485	1.51	.500		1	12.576***	-0.28
Race/ethnicity (dummy 1)	.22	.413	.32	.465		1	7.901**	-0.23
Race/ethnicity (dummy 2)	.16	.371	.22	.416		1	3.453	-0.15
Urbanicity	.71	.456	.72	.448		1	.261	-0.02
Household poverty ratio (square root)	15.33	5.823	14.61	6.579	-1.462	720		0.12
Grades received in eighth grade	5.15	1.606	5.26	1.610	.917	720		-0.07
PIAT math standard score	93.01	13.172	91.48	14.402	-1.417	720		0.11
Work-based learning	.18	.384	.16	.370		1	.318	0.05
Remedial English and/or math	.17	.375	.19	.390		1	.409	-0.05
ESL and/or bilingual program	.07	.260	.11	.312		1	2.527	-0.14
Educational and/or physical handicap	.09	.284	.06	.239		1	1.842	0.11
Attitudes toward school	15.88	2.686	16.45	2.889	2.586*	720		-0.20
Number of days absent from school	4.11	4.494	6.15	8.585	3.583***	720		-0.30
Ever suspended from school	.30	.458	.40	.491		1	8.184**	-0.21
School type	1.03	.182	1.06	.231		1	1.779	-0.14
Student-teacher ratio	2.21	1.068	2.41	1.086	2.380*	720		-0.19
Percent peers college-bound	3.44	.952	3.38	1.044	-.803	720		0.06

[a]$\underline{n} = 262$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c}} / 2$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A11

Imputation 4 - Differences Between CTE (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227179.11	82849.466	221248.99	85768.353	-.863	624		0.07
Gender	1.38	.486	1.39	.488		1	.049	-0.02
Race/ethnicity (dummy 1)	.22	.416	.24	.429		1	.371	-0.05
Race/ethnicity (dummy 2)	.17	.373	.18	.381		1	.056	-0.03
Urbanicity	.70	.458	.74	.441		1	.918	-0.09
Household poverty ratio (square root)	15.36	5.838	15.49	6.620	.268	624		-0.02
Grades received in eighth grade	5.16	1.617	5.00	1.691	-1.173	624		0.10
PIAT math standard score	92.88	13.233	92.78	14.509	-.094	624		0.01
Work-based learning	.18	.384	.19	.395		1	.214	-0.03
Remedial English and/or math	.17	.377	.16	.371		1	.062	0.03
ESL and/or bilingual program	.07	.262	.09	.281		1	.359	-0.07
Educational and/or physical handicap	.09	.280	.07	.255		1	.458	0.07
Attitudes toward school	15.92	2.671	15.93	2.692	.058	624		0.00
Number of days absent from school	4.15	4.513	4.07	4.227	-.219	624		0.02
Ever suspended from school	.30	.460	.32	.466		1	.172	-0.04
School type	1.03	.184	1.03	.182		1	.001	0.00
Student-teacher ratio	2.22	1.068	2.20	1.083	-.273	624		0.02
Percent peers college-bound	3.43	.953	3.39	1.012	-.494	624		0.04

[a]$\underline{n}$ = 258 [b]$\underline{n}$ = 368 [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A12

Imputation 4 - Differences Between CTE (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227616.32	82684.399	223200.80	85861.138	-.667	699		0.05
Gender	1.38	.486	1.37	.482		1	.098	0.02
Race/ethnicity (dummy 1)	.22	.415	.24	.429		1	.500	-0.05
Race/ethnicity (dummy 2)	.17	.372	.16	.365		1	.054	0.03
Urbanicity	.70	.457	.74	.437		1	1.319	-0.09
Household poverty ratio (square root)	15.36	5.817	15.91	7.171	1.055	699		-0.08
Grades received in eighth grade	5.15	1.611	5.00	1.687	-1.223	699		0.09
PIAT math standard score	92.93	13.193	92.54	14.076	-.368	699		0.03
Work-based learning	.18	.382	.20	.401		1	.652	-0.05
Remedial English and/or math	.17	.376	.15	.356		1	.594	0.05
ESL and/or bilingual program	.07	.261	.08	.275		1	.166	-0.04
Educational and/or physical handicap	.08	.279	.08	.277		1	.001	0.00
Attitudes toward school	15.91	2.662	15.85	2.779	-.276	699		0.02
Number of days absent from school	4.13	4.503	4.19	4.540	.155	699		-0.01
Ever suspended from school	.30	.459	.34	.475		1	1.201	-0.09
School type	1.03	.183	1.04	.191		1	.071	-0.05
Student-teacher ratio	2.22	1.066	2.18	1.064	-.418	699		0.04
Percent peers college-bound	3.44	.955	3.41	.964	-.389	699		0.03

[a]$\underline{n}$ = 260 [b]$\underline{n}$ = 441 [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table A13

Imputation 5 - Differences Between CTE (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	228064.87	82530.358	203063.32	87646.771	-3.764***	720		0.29
Gender	1.37	.485	1.51	.500		1	12.576***	-0.28
Race/ethnicity (dummy 1)	.22	.413	.32	.465		1	7.901**	-0.23
Race/ethnicity (dummy 2)	.16	.371	.22	.416		1	3.453	-0.15
Urbanicity	.71	.453	.72	.448		1	.086	-0.02
Household poverty ratio (square root)	15.42	5.495	14.33	6.447	-2.316	720		0.18
Grades received in eighth grade	5.16	1.606	5.27	1.603	.898	720		-0.07
PIAT math standard score	92.77	13.138	91.38	14.315	-1.296	720		0.10
Work-based learning	.18	.384	.16	.370		1	.318	0.05
Remedial English and/or math	.17	.375	.19	.390		1	.409	-0.05
ESL and/or bilingual program	.07	.260	.11	.312		1	2.527	-0.14
Educational and/or physical handicap	.09	.284	.06	.239		1	1.842	0.11
Attitudes toward school	15.88	2.686	16.45	2.896	2.602**	720		-0.20
Number of days absent from school	4.13	4.513	6.14	8.586	3.524***	720		-0.29
Ever suspended from school	.30	.458	.40	.491		1	8.184**	-0.21
School type	1.03	.182	1.06	.231		1	1.779	-0.14
Student-teacher ratio	2.21	1.076	2.39	1.066	2.119*	720		-0.17
Percent peers college-bound	3.44	.956	3.37	1.043	-.906	720		0.07

[a]$\underline{n} = 262$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

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

Table A14

Imputation 5 - Differences Between CTE (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	226677.93	83062.580	222116.93	85137.883	-.666	631		0.05
Gender	1.39	.488	1.40	.490		1	.035	-0.02
Race/ethnicity (dummy 1)	.22	.417	.23	.420		1	.033	-0.02
Race/ethnicity (dummy 2)	.17	.377	.18	.388		1	.177	-0.03
Urbanicity	.70	.458	.70	.458		1	.001	0.00
Household poverty ratio (square root)	15.43	5.553	15.26	6.206	-.353	631		0.03
Grades received in eighth grade	5.17	1.588	5.12	1.704	-.356	631		0.03
PIAT math standard score	92.08	12.468	91.88	14.582	-.181	631		0.01
Work-based learning	.17	.380	.18	.384		1	.016	-0.03
Remedial English and/or math	.17	.380	.17	.375		1	.047	0.00
ESL and/or bilingual program	.08	.265	.07	.253		1	.118	0.04
Educational and/or physical handicap	.09	.283	.10	.301		1	.273	-0.03
Attitudes toward school	15.93	2.693	15.77	2.795	-.695	631		0.06
Number of days absent from school	4.23	4.557	3.97	3.941	-.753	631		0.06
Ever suspended from school	.30	.460	.32	.467		1	.245	-0.04
School type	1.04	.186	1.03	.180		1	.011	0.05
Student-teacher ratio	2.25	1.074	2.28	1.036	.319	631		-0.03
Percent peers college-bound	3.42	.956	3.44	1.009	.293	631		-0.02

[a]$\underline{n} = 252$ [b]$\underline{n} = 381$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

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

Table A15

Imputation 5 - Differences Between CTE (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	227440.84	82795.962	223878.42	85366.442	-.539	700		0.04
Gender	1.38	.486	1.41	.493		1	.820	-0.06
Race/ethnicity (dummy 1)	.22	.415	.21	.409		1	.060	0.02
Race/ethnicity (dummy 2)	.17	.373	.21	.406		1	1.826	-0.10
Urbanicity	.71	.454	.71	.453		1	.000	0.00
Household poverty ratio (square root)	15.45	5.50	15.36	6.583	-.178	700		0.01
Grades received in eighth grade	5.15	1.604	5.18	1.677	.228	700		-0.02
PIAT math standard score	92.57	13.018	92.43	15.025	-.125	700		0.01
Work-based learning	.17	.380	.17	.375		1	.023	0.00
Remedial English and/or math	.17	.376	.18	.386		1	.188	-0.03
ESL and/or bilingual program	.07	.261	.06	.243		1	.270	0.04
Educational and/or physical handicap	.08	.279	.12	.321		1	1.598	-0.13
Attitudes toward school	15.91	2.667	15.85	2.884	-.284	700		0.02
Number of days absent from school	4.17	4.523	3.91	4.320	-.757	700		0.06
Ever suspended from school	.30	.460	.28	.449		1	.442	0.04
School type	1.03	.183	1.03	.180		1	.004	0.00
Student-teacher ratio	2.23	1.074	2.30	1.019	.865	700		-0.07
Percent peers college-bound	3.44	.960	3.39	.982	-.631	700		0.05

[a]n = 259 [b]n = 443 [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

APPENDIX B

COVARIATE BALANCE IMPROVEMENT

FOR

CTE AND GENERAL-TRACK STUDENTS

Table B1

Imputation 1 - Standardized Mean Difference[a] Improvement for the Sample of CTE (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.3029	0.0372	87.7350	0.3029	0.0423	86.0491
Gender	-0.2822	-0.0365	87.0546	-0.2822	-0.0433	84.6506
Race/ethnicity (dummy 1)	-0.2363	-0.0028	98.7953	-0.2363	-0.0117	95.0390
Race/ethnicity (dummy 2)	-0.1553	-0.0432	72.2077	-0.1553	-0.0294	81.0554
Urbanicity	-0.0225	0.0141	37.0344	-0.0225	-0.0082	63.3027
Household poverty ratio (square root)	0.1361	0.0223	83.6114	0.1361	0.0514	62.2619
Grades received in eighth grade	-0.0709	-0.0302	57.3560	-0.0709	0.0519	26.7066
PIAT math standard score	0.1032	-0.0298	71.1154	0.1032	0.0206	80.0333
Work-based learning	0.0425	-0.0400	6.0343	0.0425	-0.0537	-26.1895
Remedial English and/or math	-0.0508	-0.0014	97.2506	-0.0508	-0.0414	18.4918
ESL and/or bilingual program	-0.1392	-0.0158	88.6179	-0.1392	0.0075	94.6487
Educational and/or physical handicap	0.0949	0.0390	58.9632	0.0949	0.0731	23.0261
Attitudes toward school	-0.2132	0.0622	70.8052	-0.2132	0.0307	85.5783
Number of days absent from school	-0.4931	0.0211	95.7157	-0.4931	0.0640	87.0126
Ever suspended from school	-0.2328	0.0575	75.2997	-0.2328	-0.0331	85.7972
School type	-0.1215	0.0172	85.8495	-0.1215	0.0136	88.8368
Student-teacher ratio	-0.1564	0.0111	92.9010	-0.1564	-0.0155	90.0601
Percent peers college-bound	0.0758	-0.0244	67.8410	0.0758	0.0090	88.1295

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table B2

Imputation 2 - Standardized Mean Difference[a] Improvement for the Sample of CTE (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.3029	0.0141	95.3601	0.3029	0.0299	90.1145
Gender	-0.2822	-0.0392	86.1126	-0.2822	-0.0315	88.8527
Race/ethnicity (dummy 1)	-0.2363	0.0089	96.2149	-0.2363	-0.0366	84.5138
Race/ethnicity (dummy 2)	-0.1553	-0.0315	79.7401	-0.1553	0.0160	89.7172
Urbanicity	-0.0286	-0.0326	-14.3165	-0.0286	-0.0073	74.5276
Household poverty ratio (square root)	0.1346	-0.0441	67.2462	0.1346	-0.0131	90.2577
Grades received in eighth grade	-0.0604	0.0101	83.2338	-0.0604	0.0203	66.3895
PIAT math standard score	0.1075	0.0242	77.4681	0.1075	-0.0044	95.9324
Work-based learning	0.0312	0.0051	83.7846	0.0312	-0.0027	91.4648
Remedial English and/or math	-0.0508	-0.0099	80.5627	-0.0508	0.0121	76.1893
ESL and/or bilingual program	-0.1392	0.0022	98.3866	-0.1392	-0.0055	96.0549
Educational and/or physical handicap	0.0949	-0.0293	69.1607	0.0949	-0.0677	28.6362
Attitudes toward school	-0.2148	0.0364	83.0741	-0.2148	0.0137	93.6301
Number of days absent from school	-0.4538	0.0396	91.2752	-0.4538	0.0583	87.1639
Ever suspended from school	-0.2328	0.0265	88.6278	-0.2328	0.0370	84.1083
School type	-0.1215	-0.0224	81.5720	-0.1215	-0.0958	21.1596
Student-teacher ratio	-0.1550	-0.0407	73.7651	-0.1550	-0.0335	78.4175
Percent peers college-bound	0.0732	-0.0319	56.4741	0.0732	-0.0838	-14.4906

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table B3

Imputation 3 - Standardized Mean Difference[a] Improvement for the Sample of CTE (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.3029	0.0721	76.1971	0.3029	0.0295	90.2656
Gender	-0.2822	-0.0272	90.3783	-0.2822	-0.0374	86.7426
Race/ethnicity (dummy 1)	-0.2363	-0.0326	86.1878	-0.2363	0.0135	94.2765
Race/ethnicity (dummy 2)	-0.1553	-0.0302	80.5280	-0.1553	-0.0361	76.7363
Urbanicity	-0.0202	-0.0128	36.2996	-0.0202	-0.0379	-88.0362
Household poverty ratio (square root)	0.1519	0.0609	59.9195	0.1519	-0.0103	93.2072
Grades received in eighth grade	-0.0976	-0.0474	51.4114	-0.0976	-0.0535	45.1627
PIAT math standard score	0.1066	0.0101	90.5355	0.1066	-0.0465	56.3382
Work-based learning	0.0369	-0.0046	87.6448	0.0369	-0.0298	19.1647
Remedial English and/or math	-0.0508	-0.0159	68.6276	-0.0508	-0.0295	41.8160
ESL and/or bilingual program	-0.1392	0.0010	99.2830	-0.1392	0.0033	97.6126
Educational and/or physical handicap	0.0949	0.0151	84.0985	0.0949	-0.0136	85.6301
Attitudes toward school	-0.2108	0.0225	89.3012	-0.2108	0.0115	94.5441
Number of days absent from school	-0.4817	0.1008	79.0800	-0.4817	0.0937	80.5383
Ever suspended from school	-0.2328	0.0515	77.8637	-0.2328	0.0350	84.9730
School type	-0.1215	-0.0007	99.4150	-0.1215	0.0249	79.5456
Student-teacher ratio	-0.1603	-0.0256	84.0069	-0.1603	-0.0411	74.3860
Percent peers college-bound	0.0867	-0.0109	87.4016	0.0867	-0.0093	89.2870

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table B4

Imputation 4 - Standardized Mean Difference[a] Improvement for the Sample of CTE (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.3029	0.0719	76.2810	0.3029	0.0535	82.3390
Gender	-0.2822	-0.0179	93.6734	-0.2822	0.0224	92.0609
Race/ethnicity (dummy 1)	-0.2363	-0.0525	77.7745	-0.2363	-0.0571	75.8271
Race/ethnicity (dummy 2)	-0.1553	-0.0230	85.2003	-0.1553	0.0205	86.7837
Urbanicity	-0.0390	-0.0783	-100.6243	-0.0390	-0.0893	-128.7700
Household poverty ratio (square root)	0.1227	-0.0236	80.7741	0.1227	-0.0949	22.6730
Grades received in eighth grade	-0.0711	0.0985	-38.5464	-0.0711	0.0988	-38.9703
PIAT math standard score	0.1163	0.0081	93.0067	0.1163	0.0300	74.1768
Work-based learning	0.0425	-0.0378	11.0790	0.0425	-0.0635	-49.3306
Remedial English and/or math	-0.0508	0.0174	65.6920	-0.0508	0.0555	-9.2297
ESL and/or bilingual program	-0.1392	-0.0487	65.0008	-0.1392	-0.0347	75.1092
Educational and/or physical handicap	0.0949	0.0551	41.9205	0.0949	0.0046	95.1516
Attitudes toward school	-0.2099	-0.0047	97.7549	-0.2099	0.0219	89.5460
Number of days absent from school	-0.4547	0.0172	96.2163	-0.4547	-0.0122	97.3109
Ever suspended from school	-0.2328	-0.0321	86.1881	-0.2328	-0.0897	61.4583
School type	-0.1215	0.0042	96.5035	-0.1215	-0.0185	84.7824
Student-teacher ratio	-0.1862	0.0224	87.9753	-0.1862	0.0326	82.4897
Percent peers college-bound	0.0660	0.0416	36.9880	0.0660	0.0307	53.5073

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table B5

Imputation 5 - Standardized Mean Difference[a] Improvement for the Sample of CTE (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.3029	0.0553	81.7571	0.3029	0.0432	85.7512
Gender	-0.2822	-0.0180	93.6194	-0.2822	-0.0696	75.3321
Race/ethnicity (dummy 1)	-0.2363	-0.0150	93.6341	-0.2363	0.0190	91.9439
Race/ethnicity (dummy 2)	-0.1553	-0.0346	77.7311	-0.1553	-0.1101	29.0582
Urbanicity	-0.0225	-0.0006	97.3994	-0.0225	-0.0038	83.0558
Household poverty ratio (square root)	0.1996	0.0311	84.4343	0.1996	0.0157	92.1382
Grades received in eighth grade	-0.0694	0.0299	56.9108	-0.0694	-0.0183	73.6163
PIAT math standard score	0.1061	0.0154	85.4622	0.1061	0.0106	89.9732
Work-based learning	0.0425	-0.0110	74.1047	0.0425	0.0136	68.0183
Remedial English and/or math	-0.0508	0.0152	70.0917	-0.0508	-0.0325	36.0801
ESL and/or bilingual program	-0.1392	0.0265	80.9868	-0.1392	0.0398	71.3829
Educational and/or physical handicap	0.0949	-0.0469	50.6119	0.0949	-0.1104	-16.2683
Attitudes toward school	-0.2116	0.0579	72.6405	-0.2116	0.0232	89.0311
Number of days absent from school	-0.4457	0.0569	87.2429	-0.4457	0.0577	87.0604
Ever suspended from school	-0.2328	-0.0409	82.4481	-0.2328	0.0503	78.4010
School type	-0.1215	0.0120	90.1557	-0.1215	0.0062	94.9137
Student-teacher ratio	-0.1630	-0.0253	84.4639	-0.1630	-0.0654	59.8894
Percent peers college-bound	0.0743	-0.0246	66.9354	0.0743	0.0503	32.3072

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

APPENDIX C

PROPENSITY SCORE DISTRIBUTION JITTER PLOTS

FOR

CTE AND GENERAL-TRACK STUDENTS

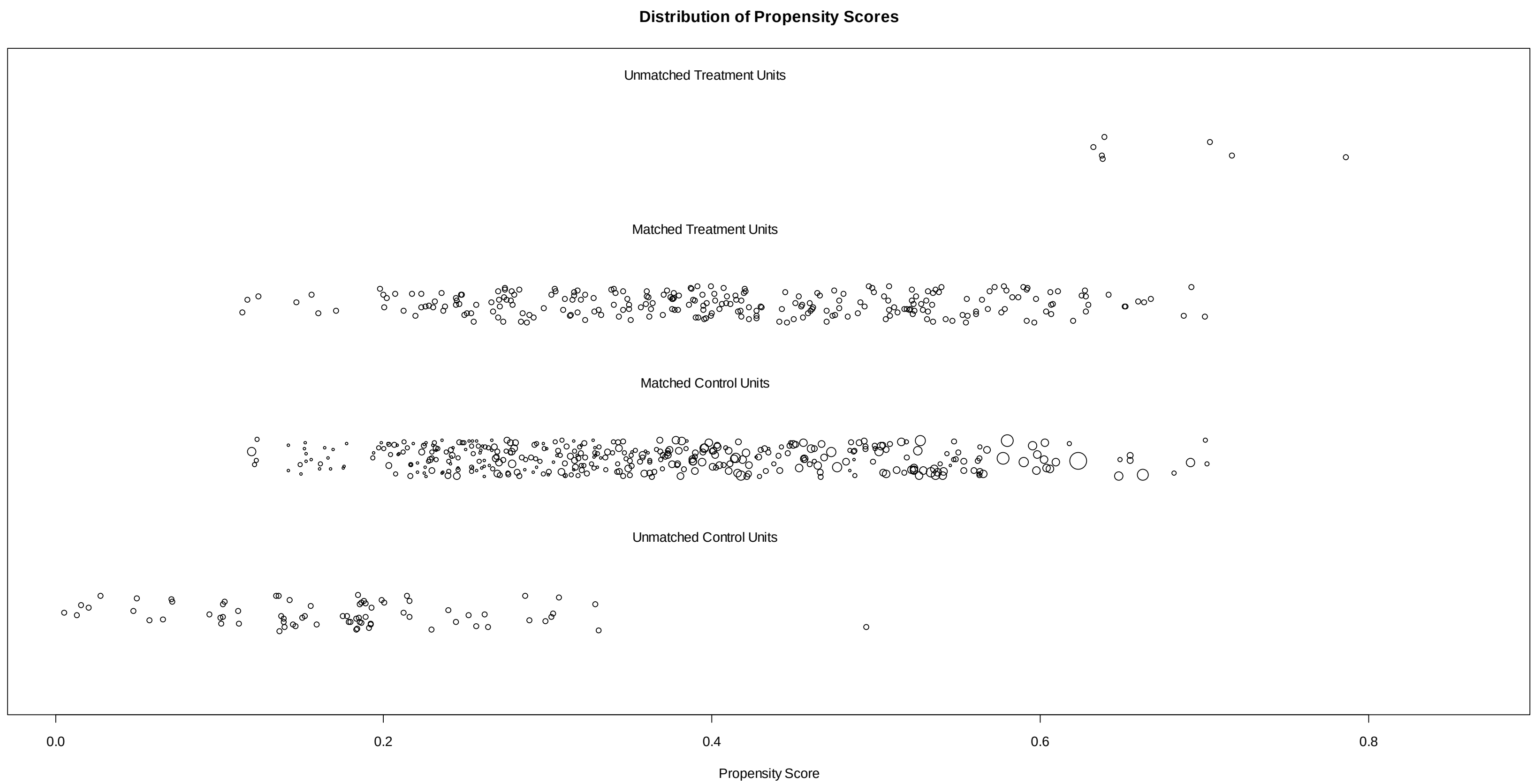

Figure C1. Imputation 1 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Seven treatment cases and 75 control cases remain unmatched due to common support and caliper size restrictions.

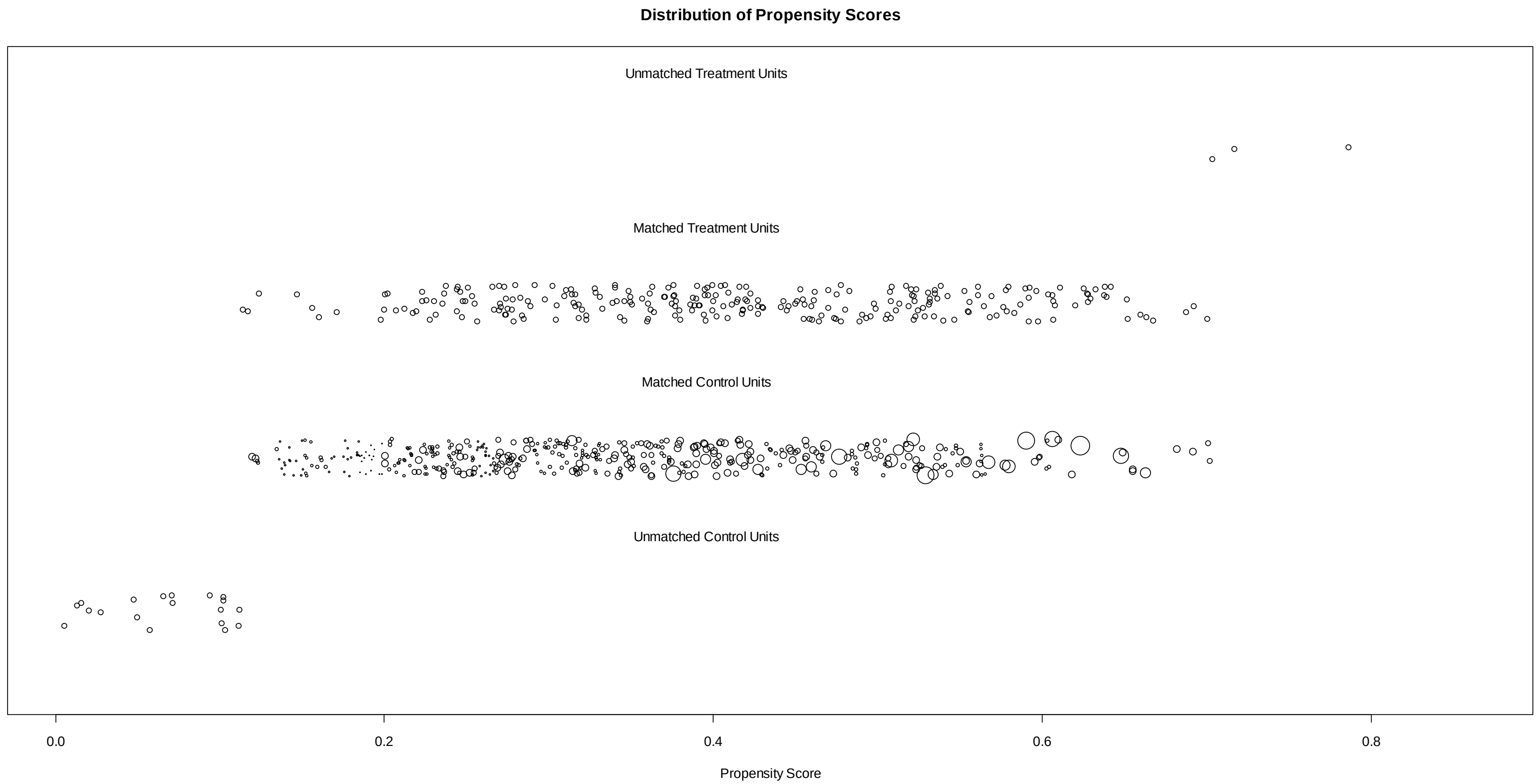

Figure C2. Imputation 1 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using full matching. Three treatment cases and 19 control cases remain unmatched due to common support restrictions.

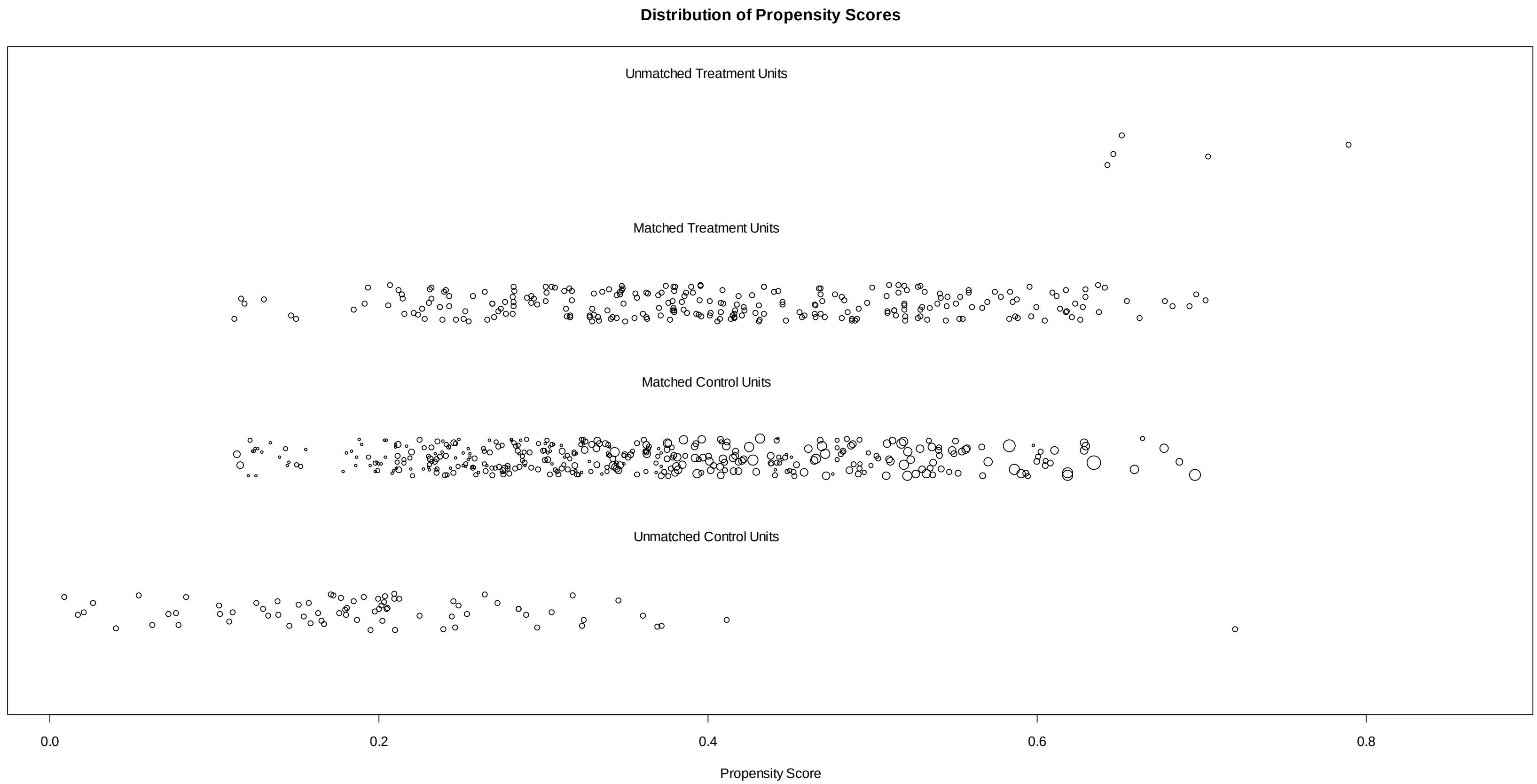

Figure C3. Imputation 2 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Five treatment cases and 75 control cases remain unmatched due to common support and caliper size restrictions.

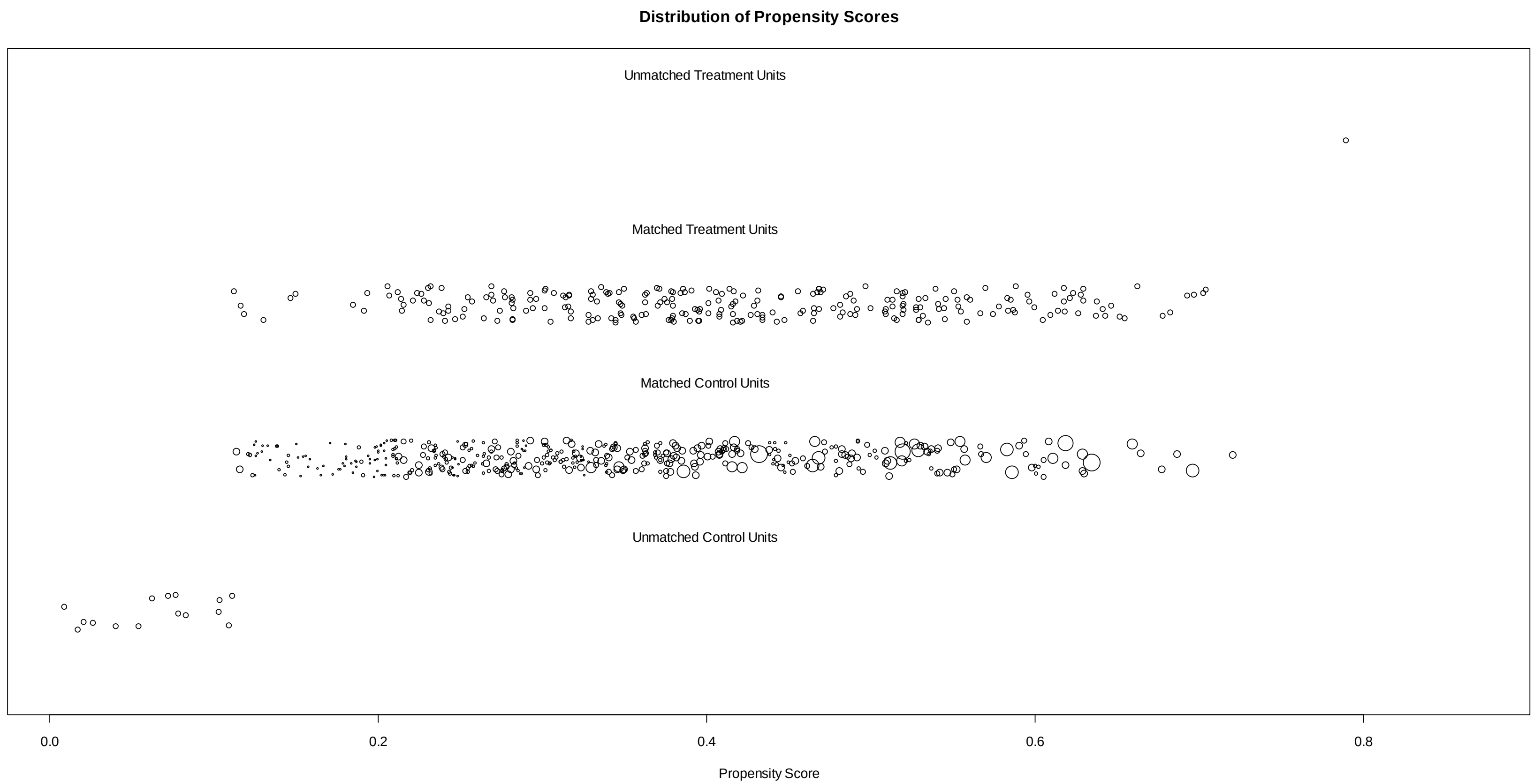

Figure C4. Imputation 2 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using full matching. One treatment case and 15 control cases remain unmatched due to common support restrictions.

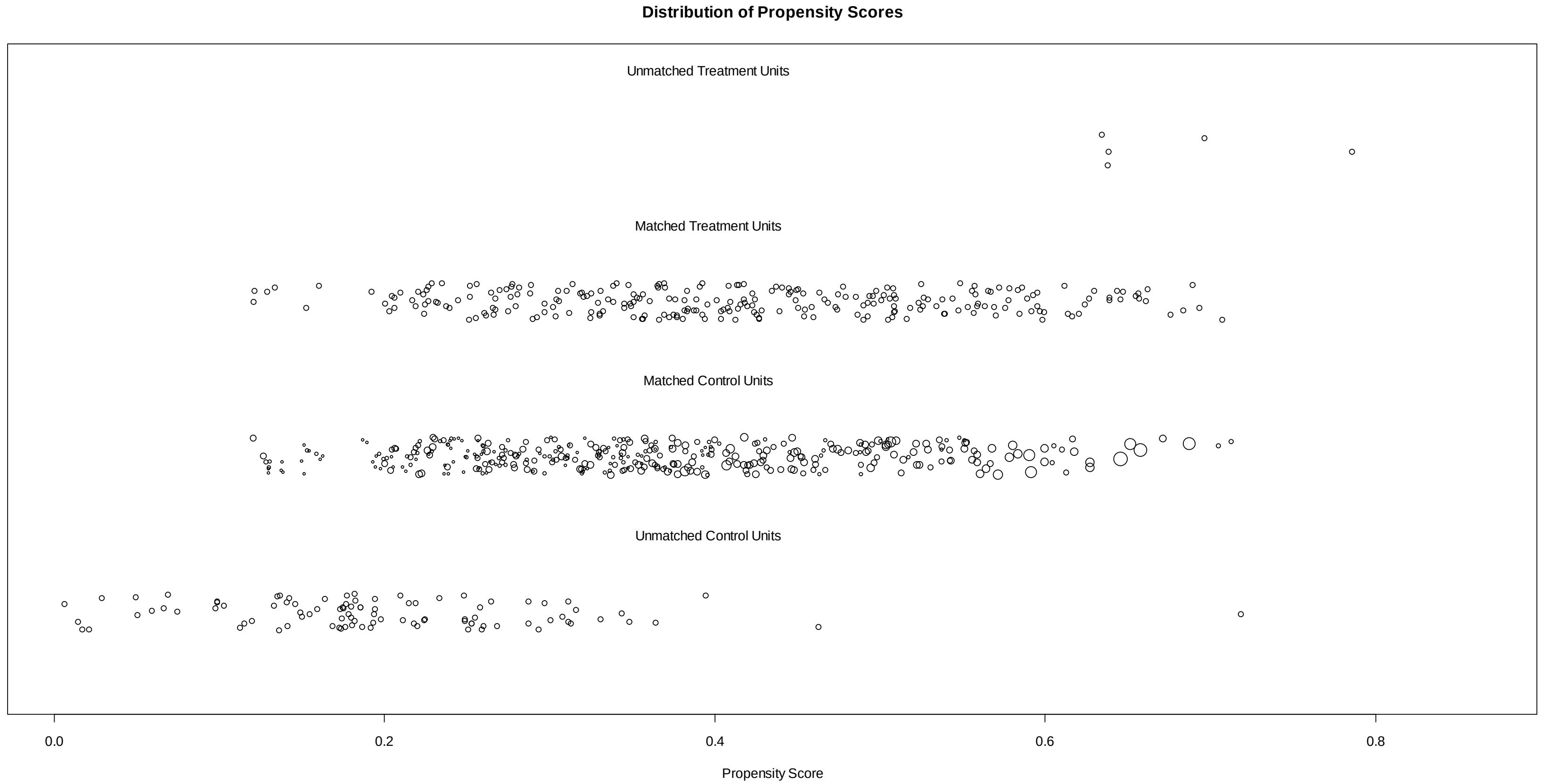

Figure C5. Imputation 3 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Five treatment cases and 95 control cases remain unmatched due to common support and caliper size restrictions.

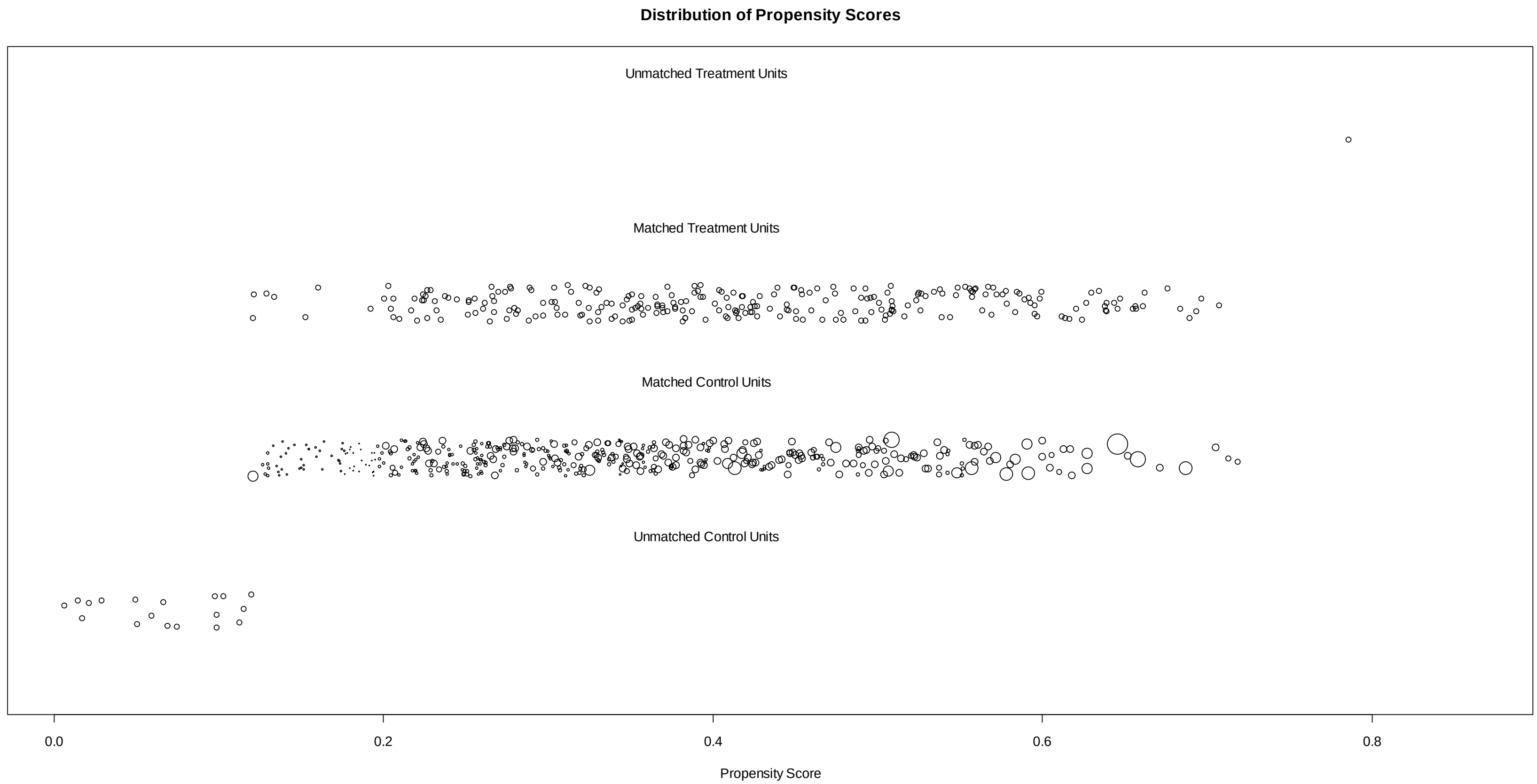

Figure C6. Imputation 3 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using full matching. One treatment case and 18 control cases remain unmatched due to common support restrictions.

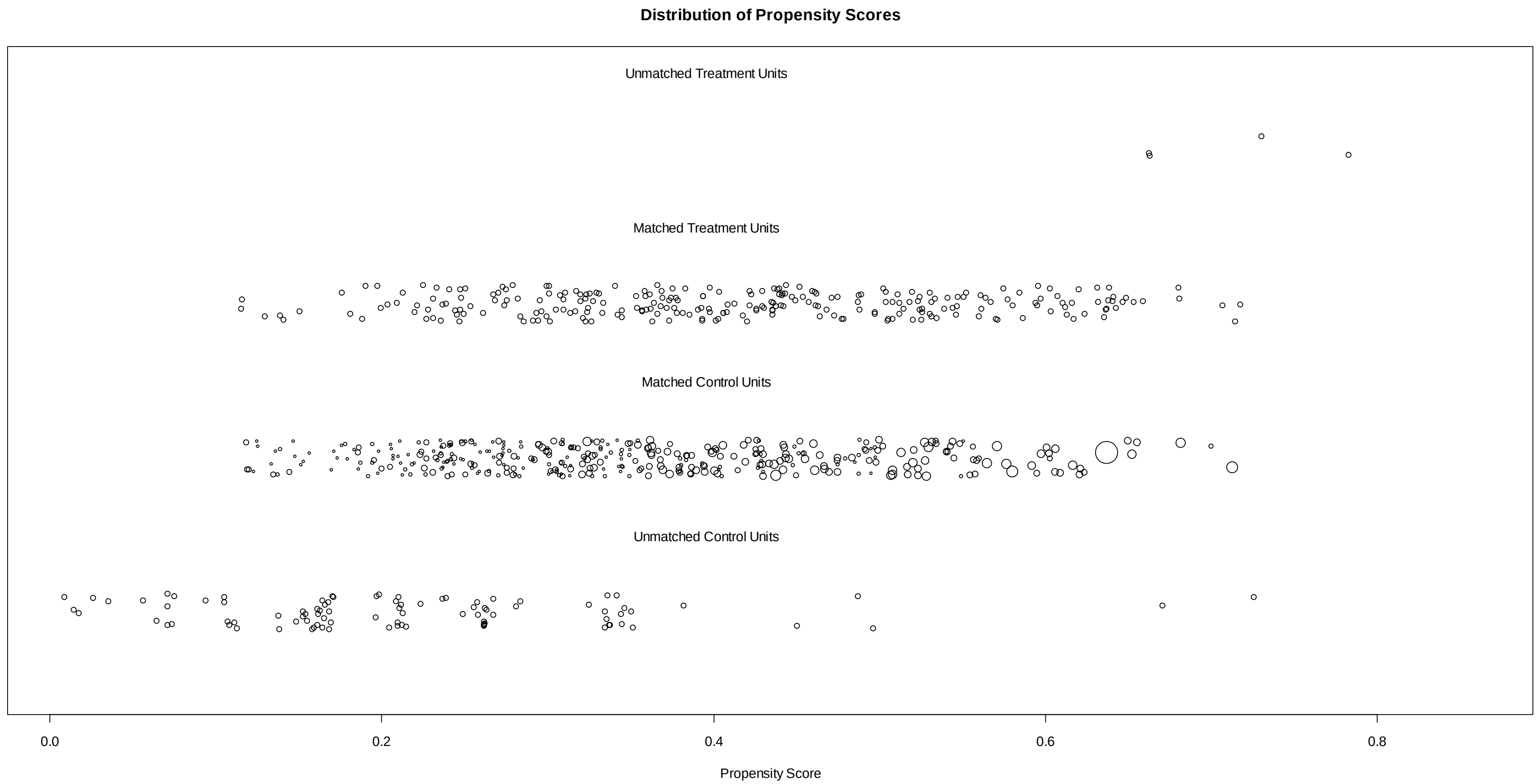

Figure C7. Imputation 4 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Four treatment cases and 92 control cases remain unmatched due to common support and caliper size restrictions.

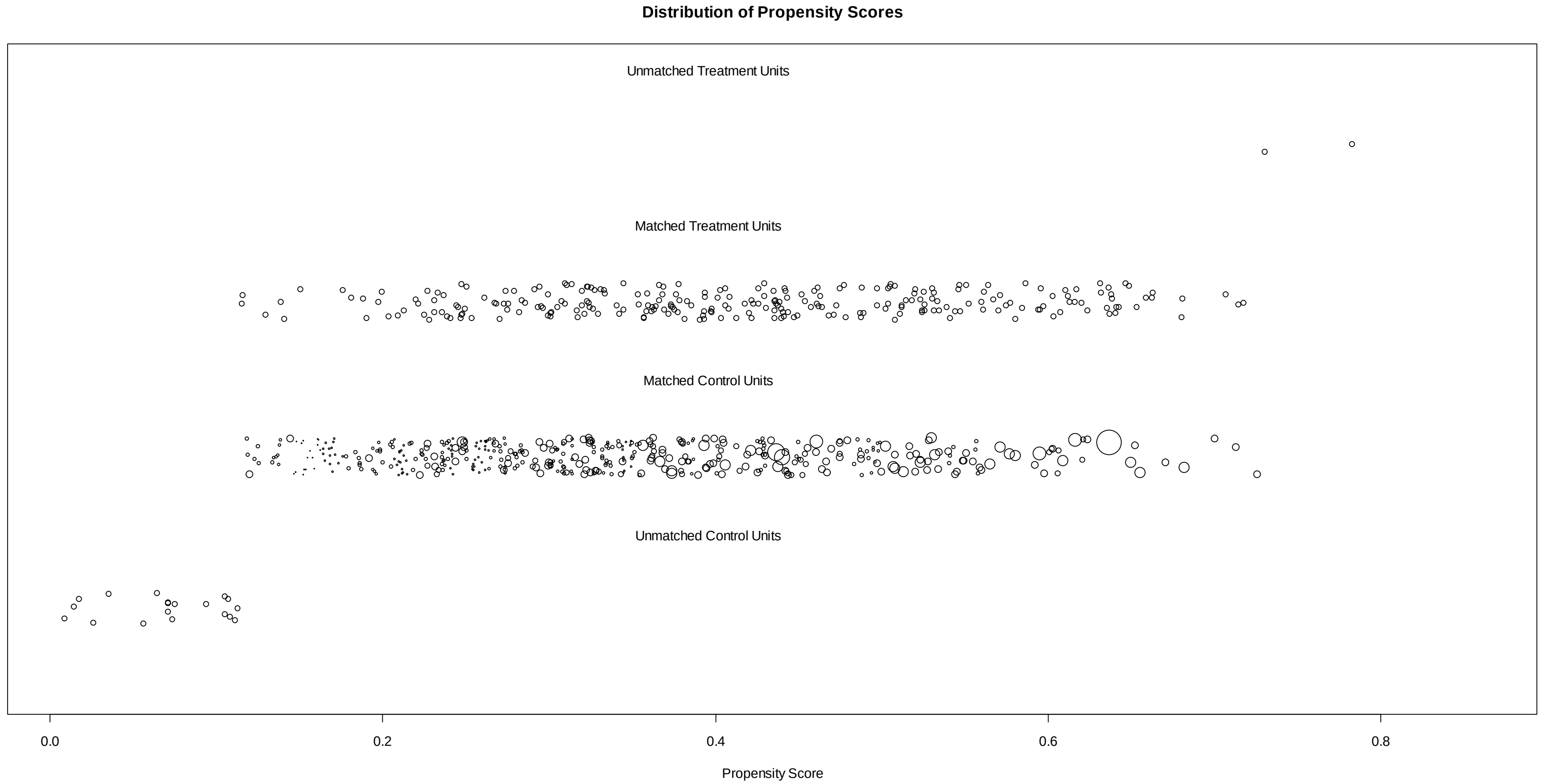

Figure C8. Imputation 4 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using full matching. Two treatment cases and 19 control cases remain unmatched due to common support restrictions.

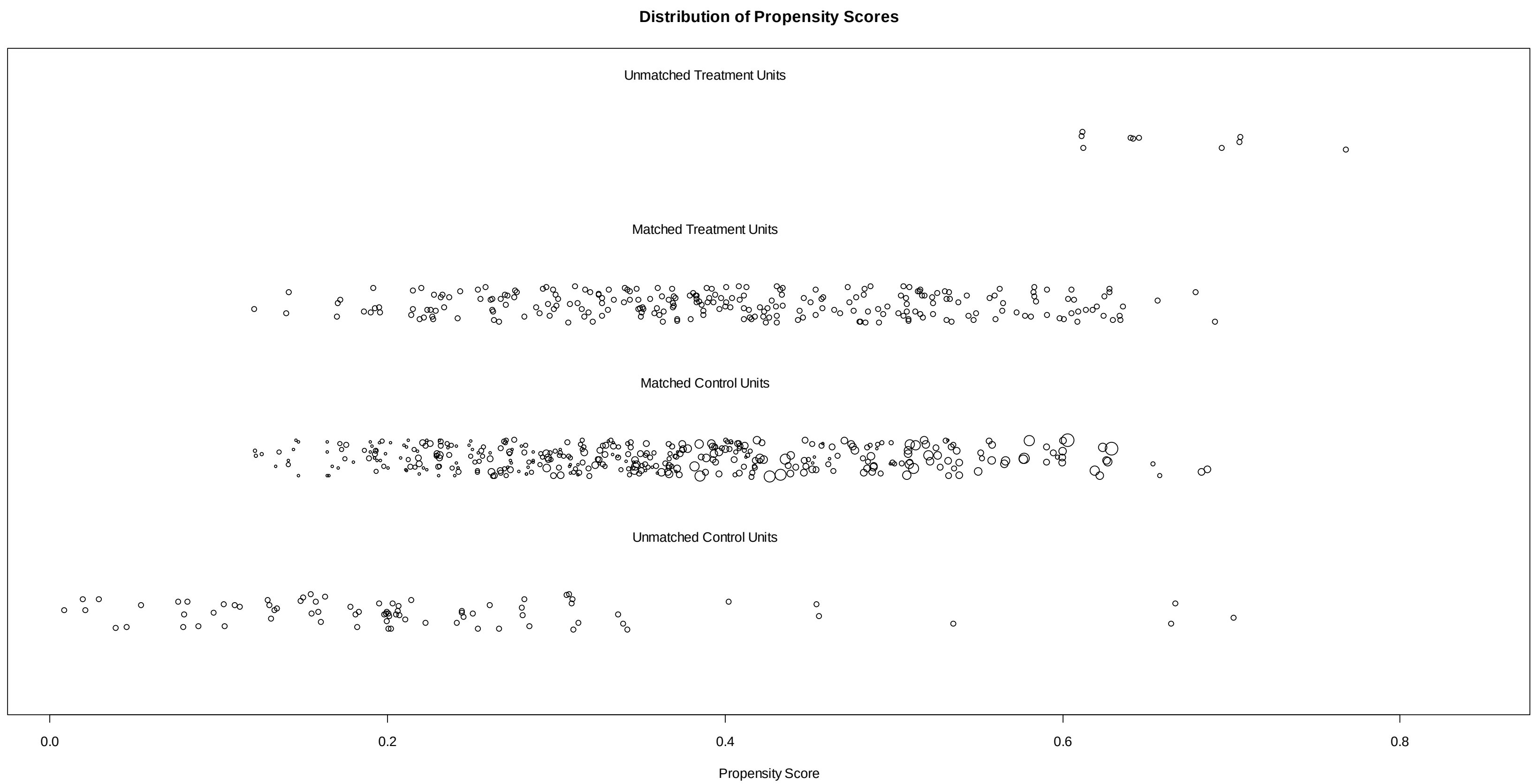

Figure C9. Imputation 5 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Ten treatment cases and 79 control cases remain unmatched due to common support and caliper size restrictions.

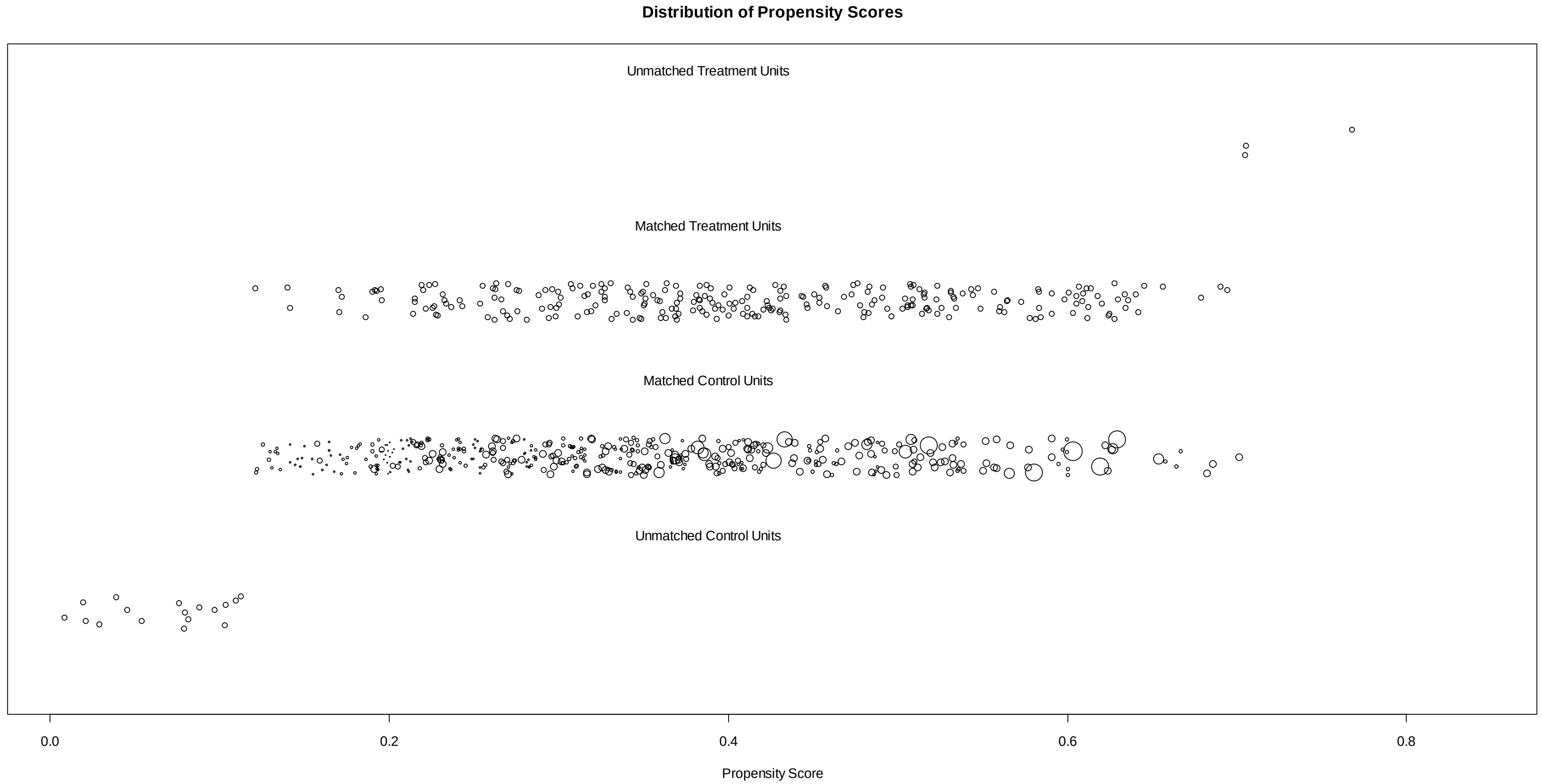

Figure C10. Imputation 5 jitter plot of the overall propensity score distribution for CTE (treatment) and general-track students (control) using full matching. Three treatment cases and 17 control cases remain unmatched due to common support restrictions.

APPENDIX D

COVARIATE BALANCE QQ PLOTS FOR IMPUTATION CYCLE 1

FOR

CTE AND GENERAL-TRACK STUDENTS

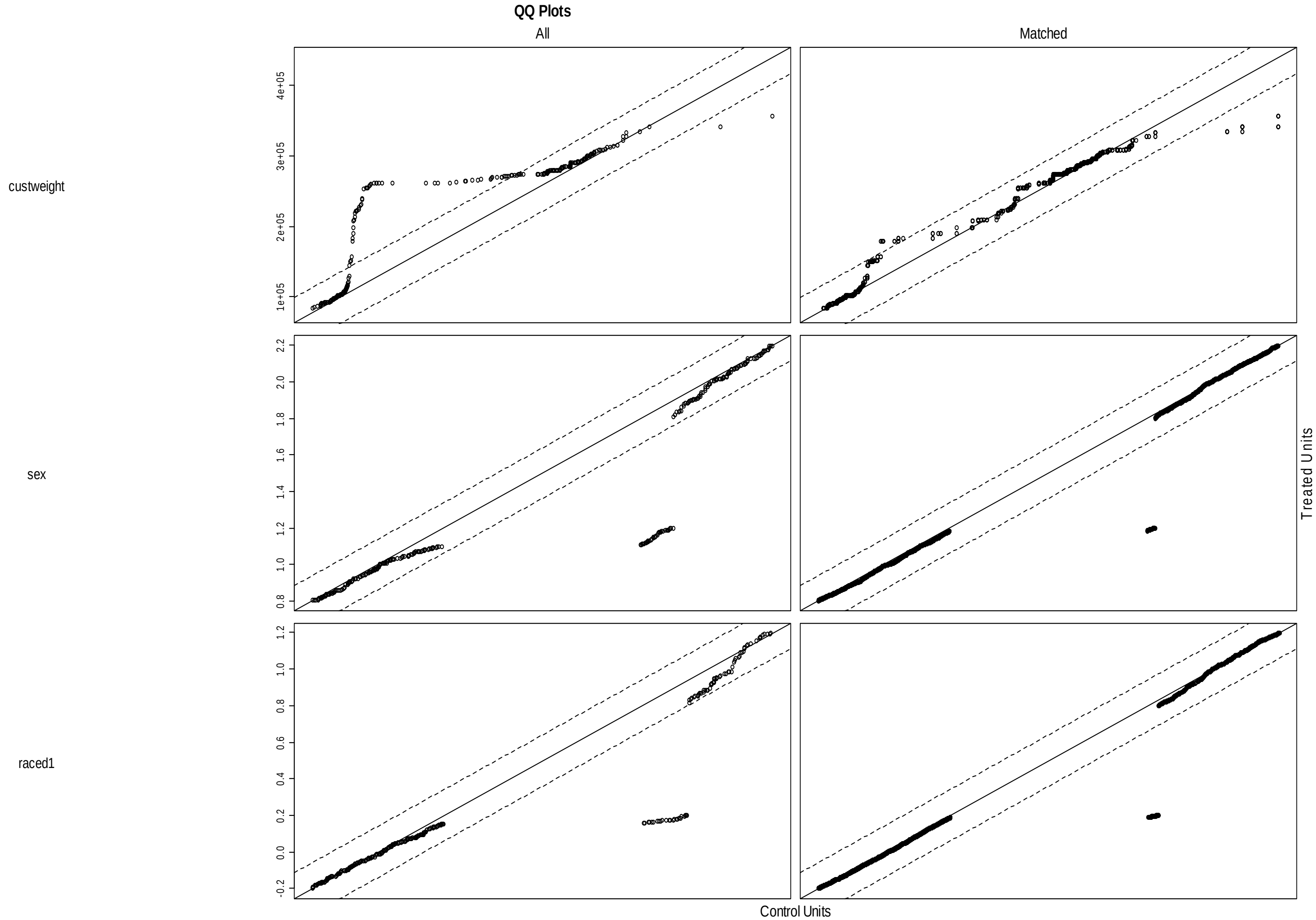

Figure D1. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for survey weight (custweight), gender (sex), and race/ethnicity dummy 1 (raced1). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

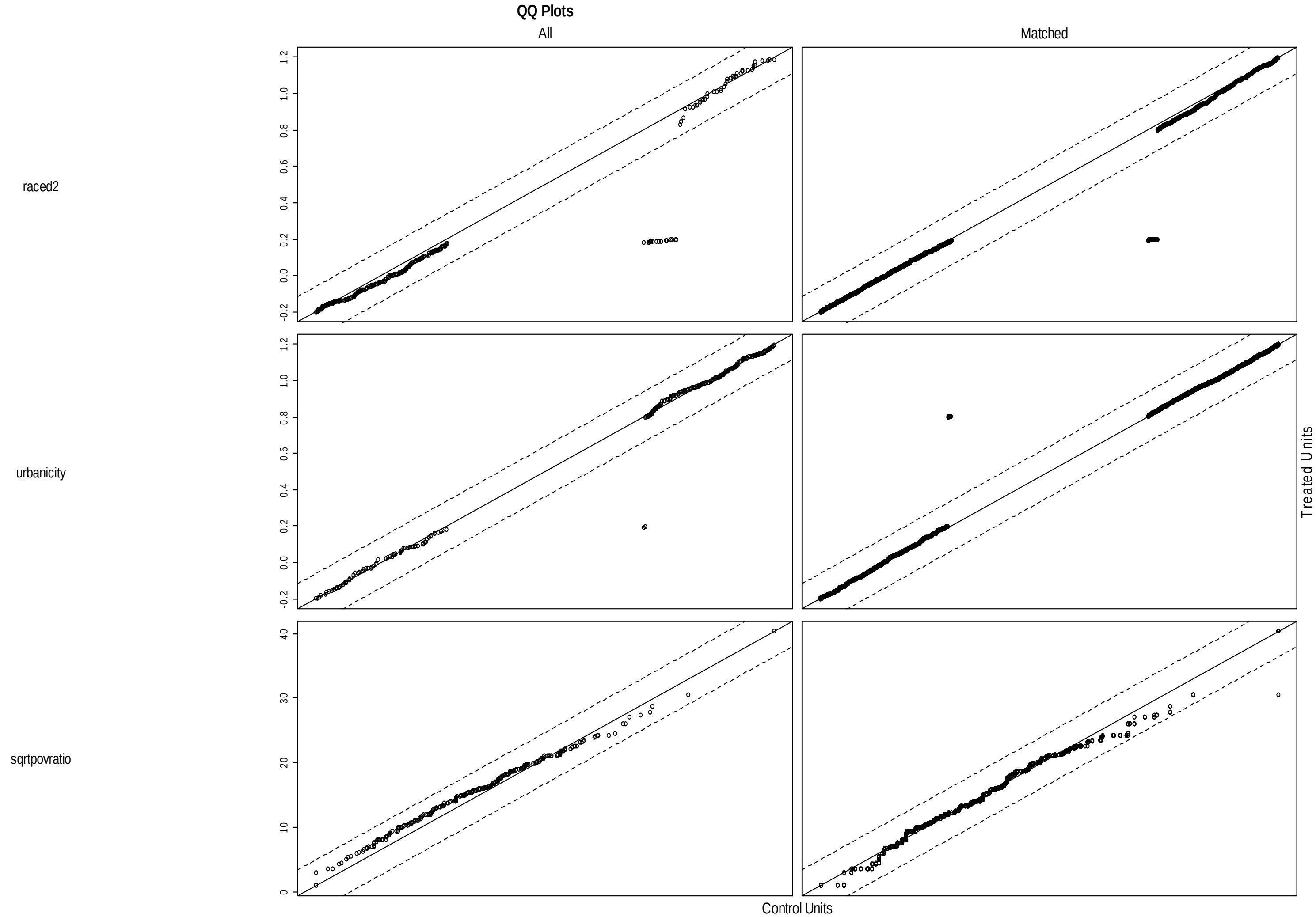

Figure D2. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for race/ethnicity dummy 2 (raced2), urbanicity (urbanicity), and household poverty square root (sqrtpovratio). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

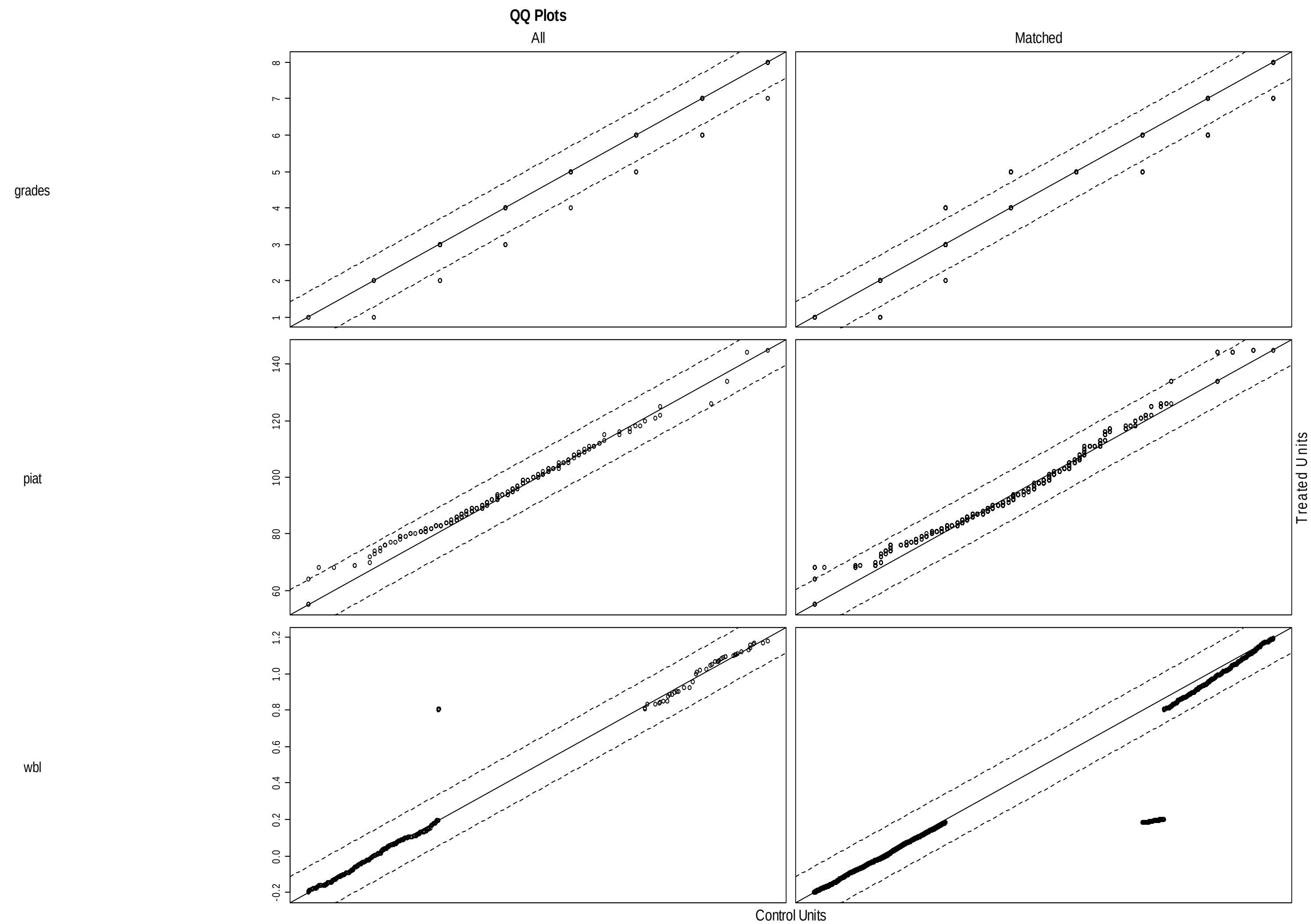

Figure D3. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for grades received in eighth grade (grades), PIAT math standard score (piat), and work-based learning (wbl). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

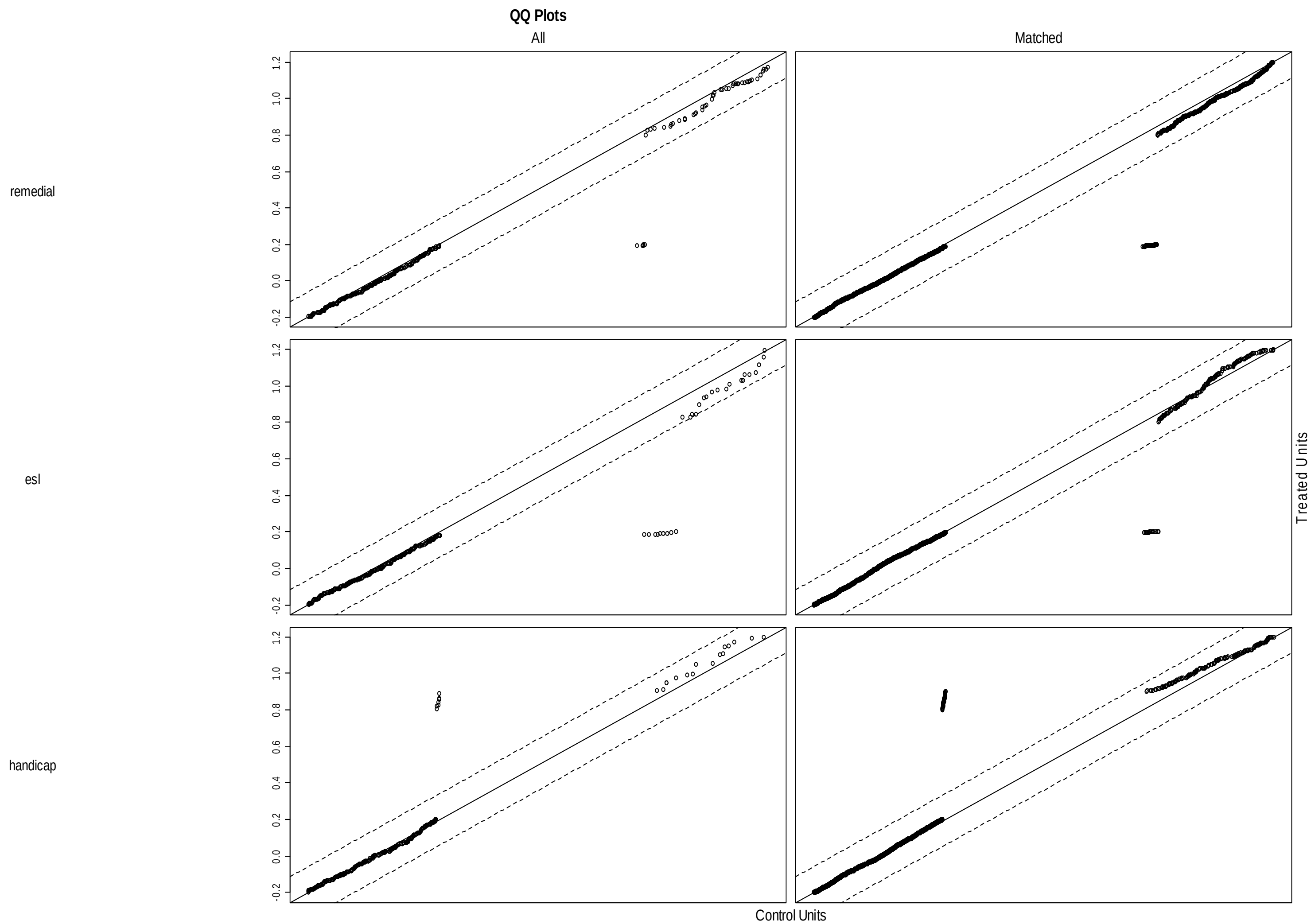

Figure D4. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for remedial English and/or math (remedial), ESL and/or bilingual program (esl), and educational and/or physical handicap (handicap). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

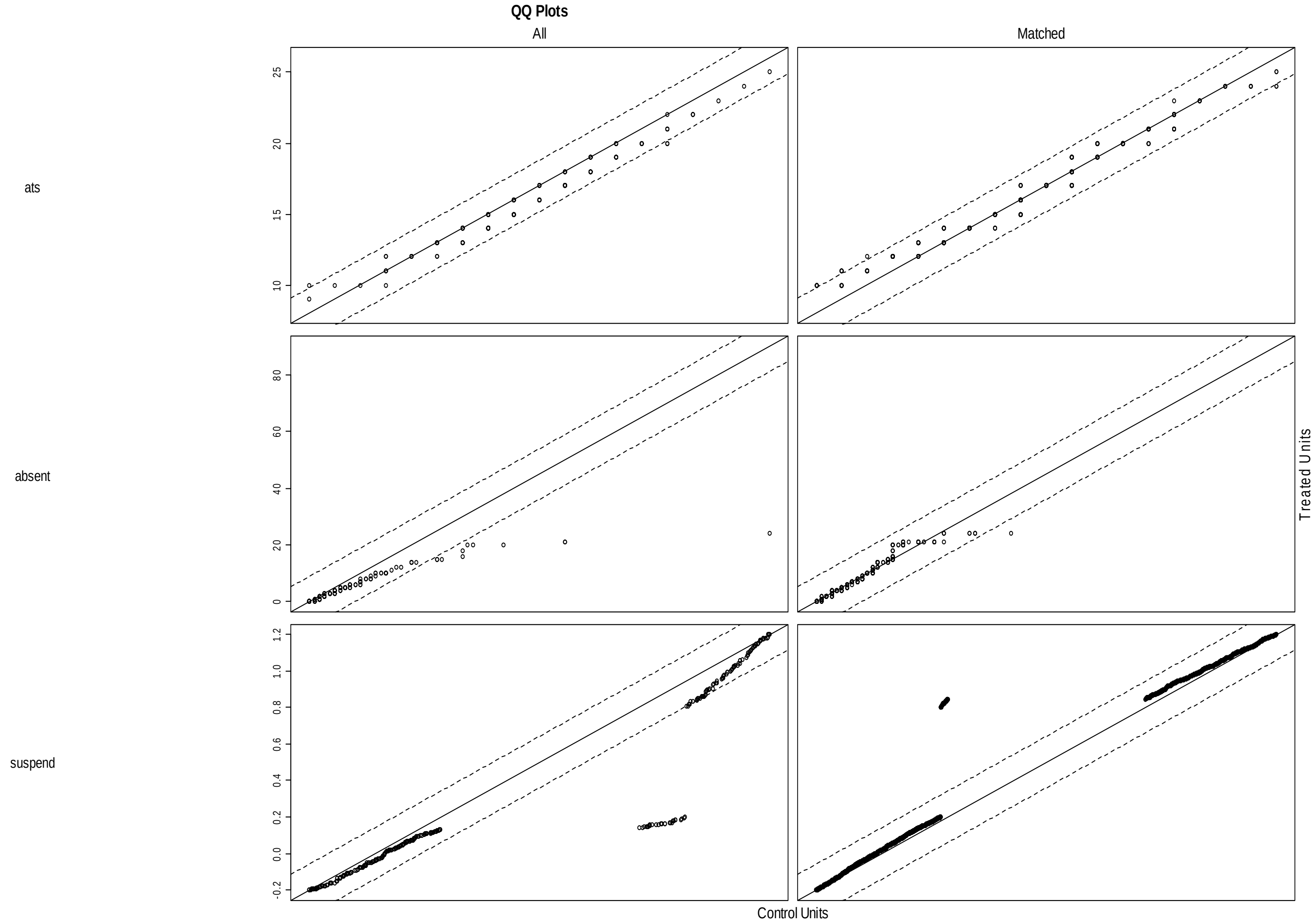

Figure D5. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for attitudes toward science (ats), number of days absent from school (absent), and ever suspended from school (suspend). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

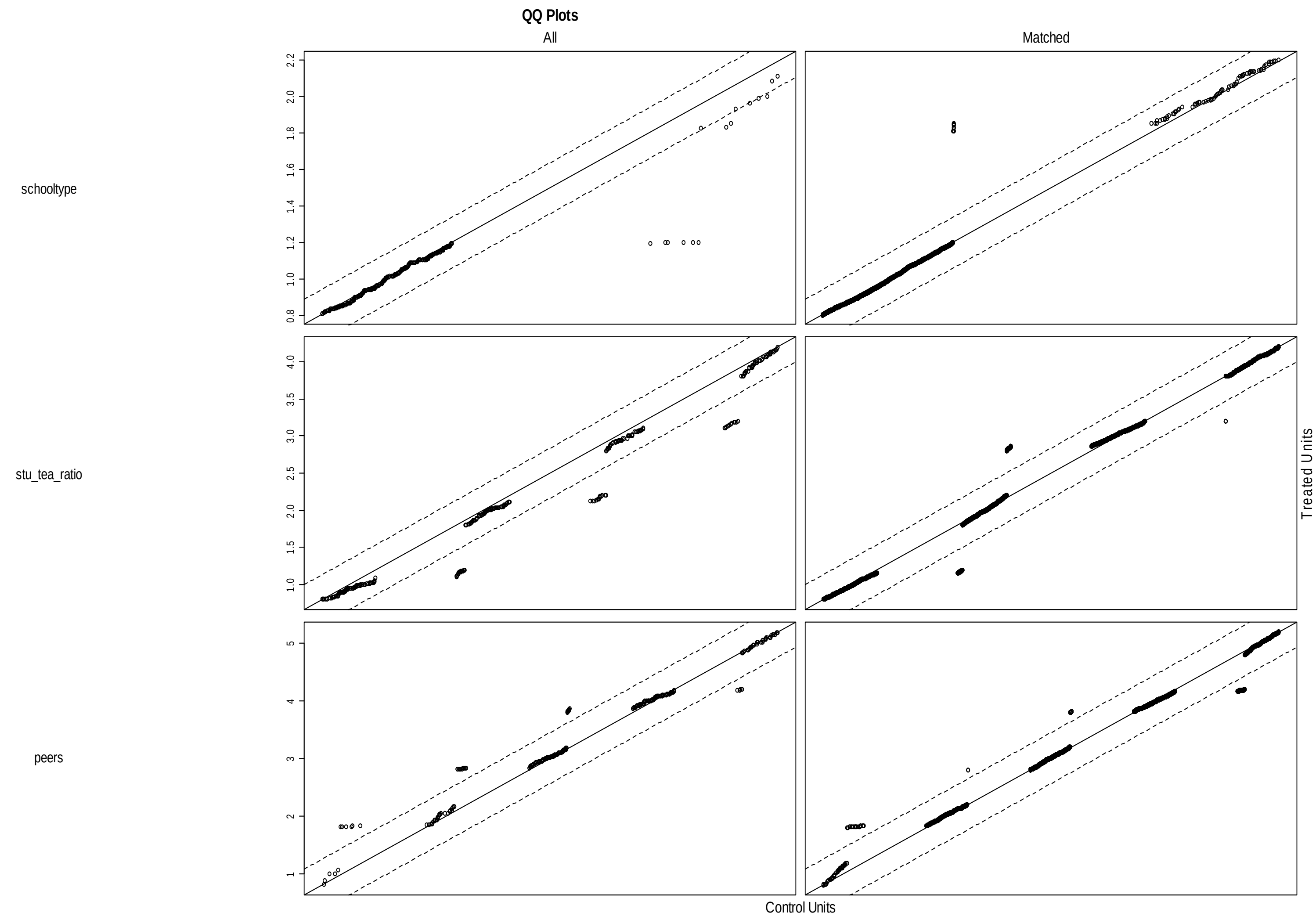

Figure D6. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for school type (schooltype), student-teacher ratio (stu_tea_ratio), and percent peers college-bound (peers). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

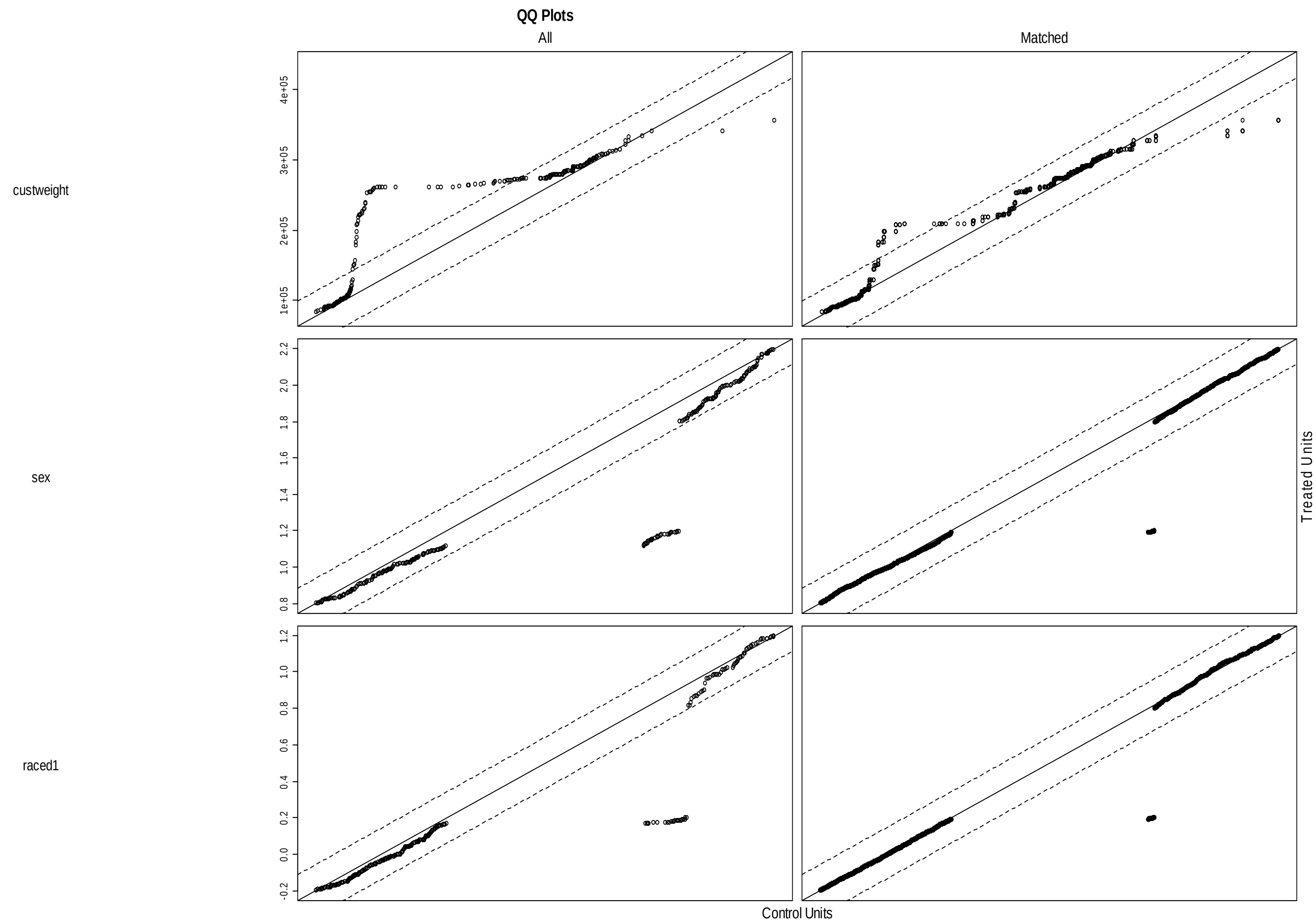

Figure D7. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for survey weight (custweight), gender (sex), and race/ethnicity dummy 1 (raced1). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

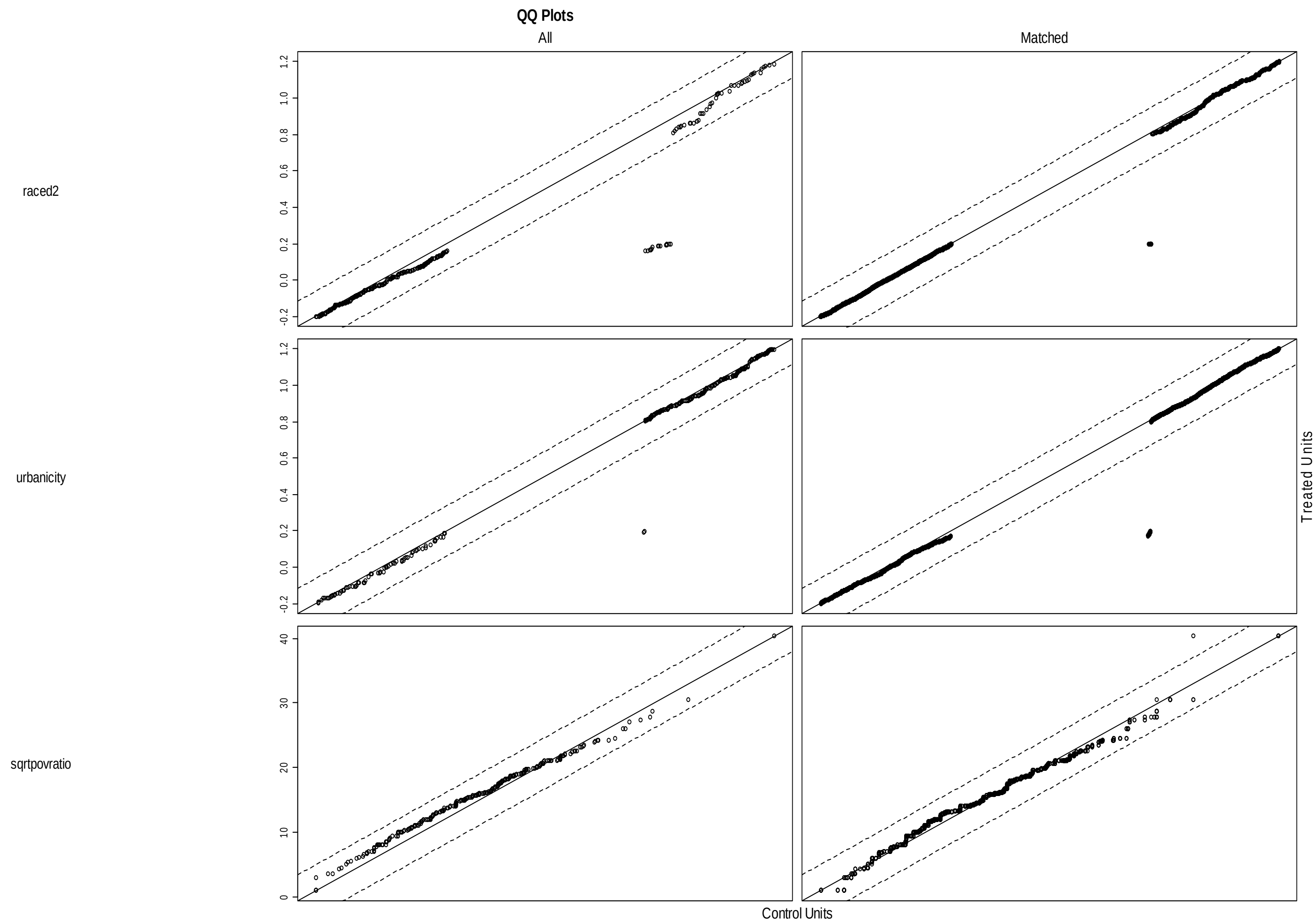

Figure D8. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for race/ethnicity dummy 2 (raced2), urbanicity (urbanicity), and household poverty ratio (sqrtpovratio). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

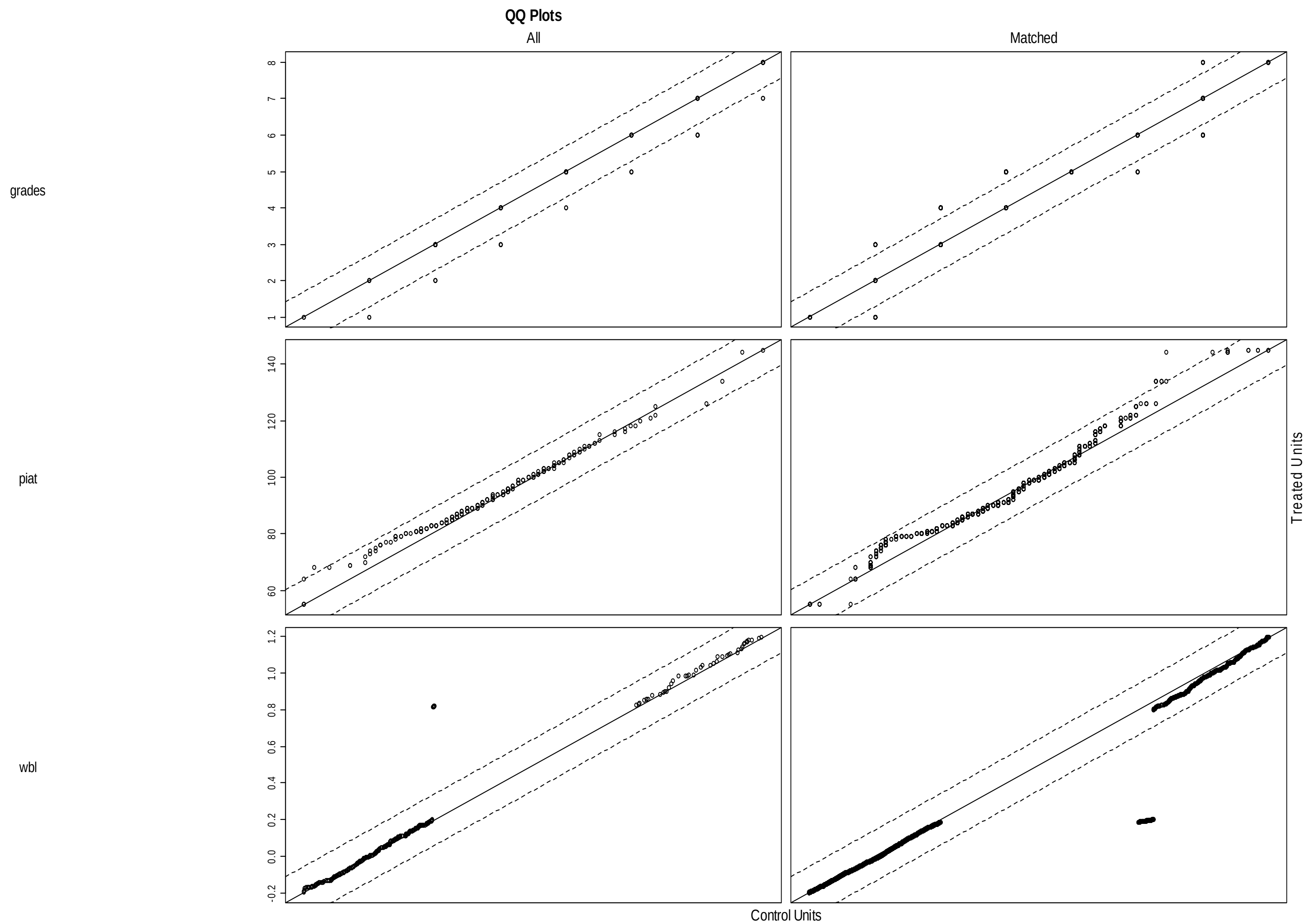

Figure D9. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for grades received in eighth grade (grades), PIAT math standard score (piat), work-based learning (wbl). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

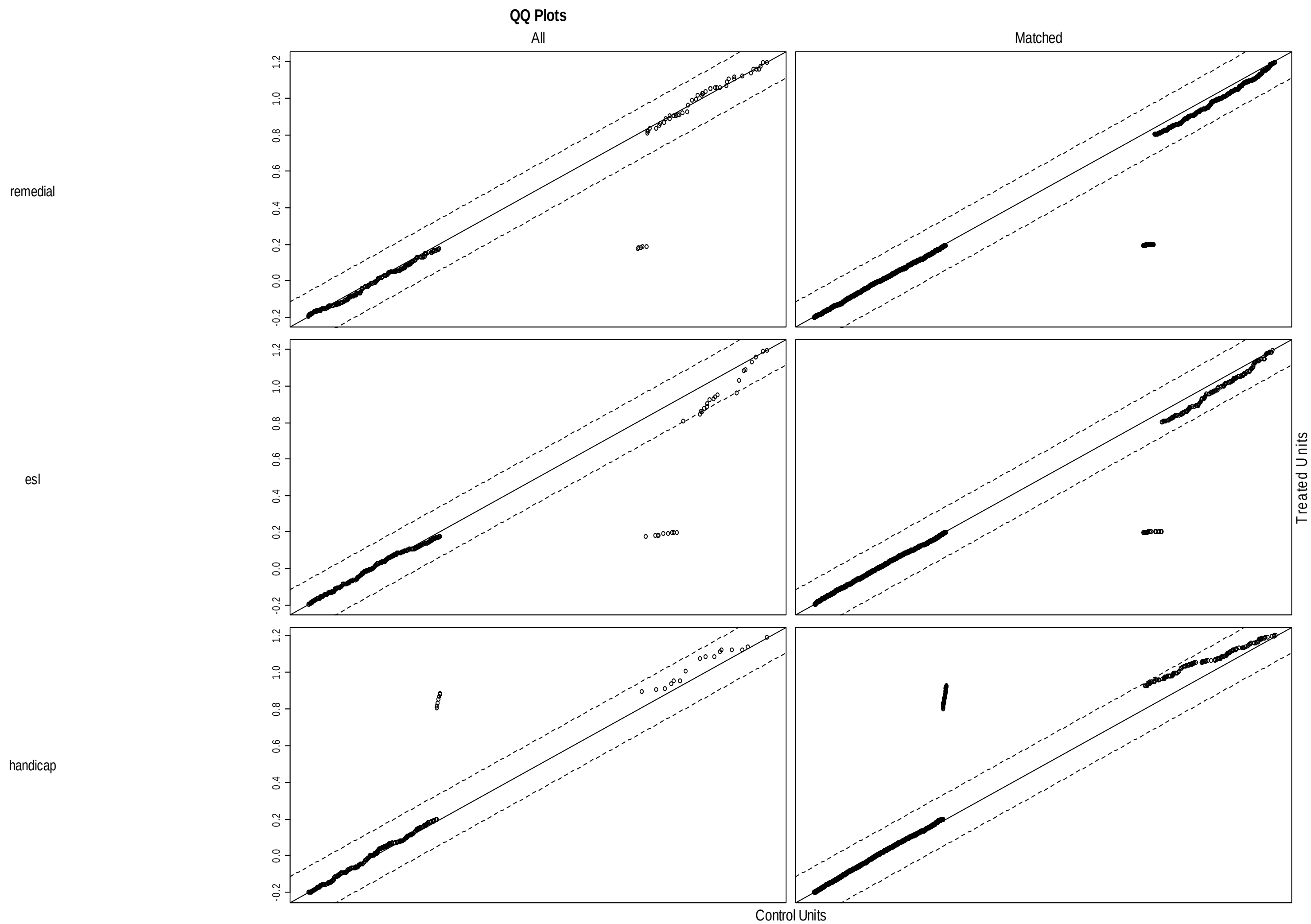

Figure D10. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for remedial and/or math (remedial), ESL and/or bilingual program (esl), and educational and/or physical handicap (handicap). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

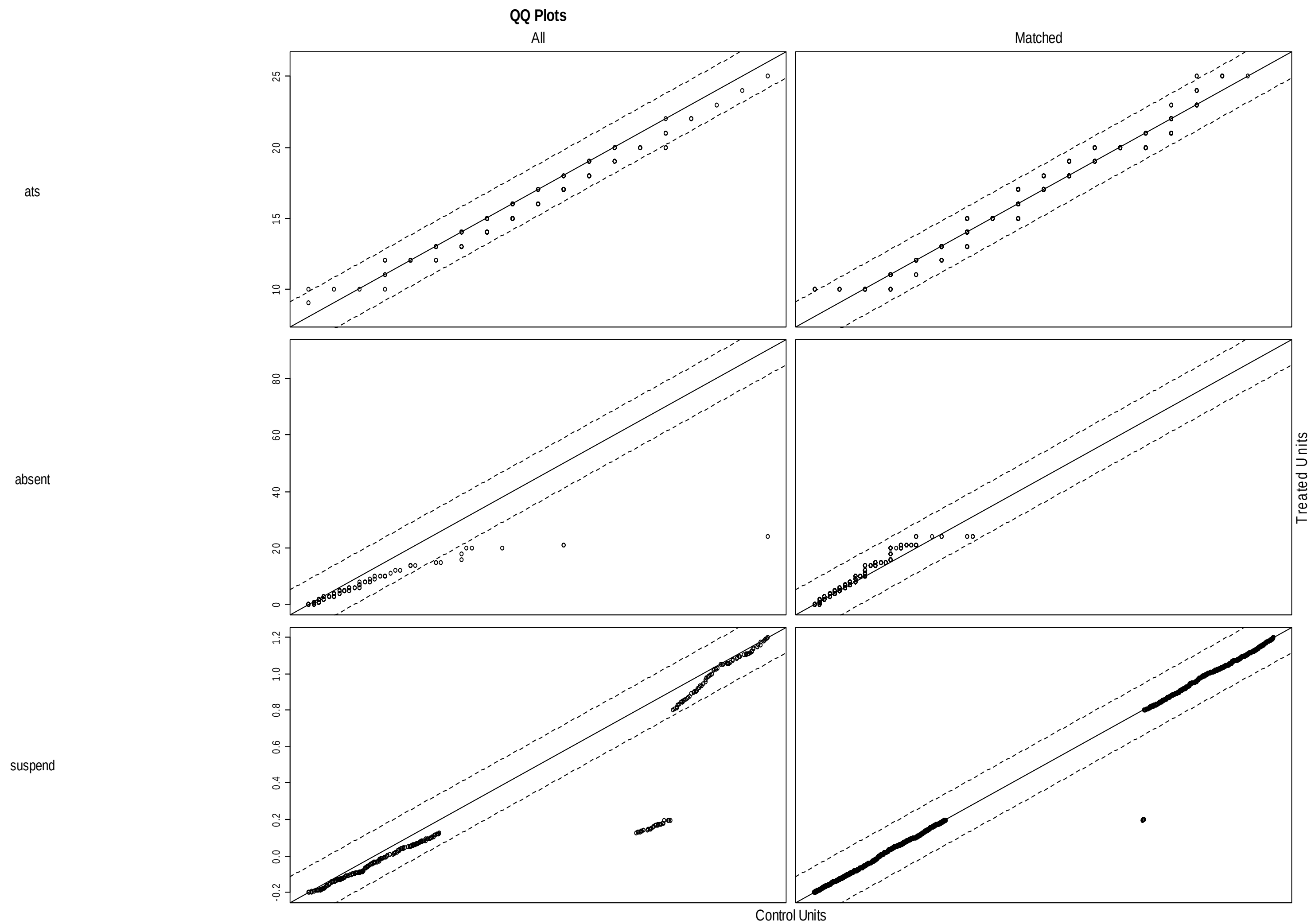

Figure D11. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for attitudes toward school (ats), number of days absent from school (absent), and ever suspended from school (suspend). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

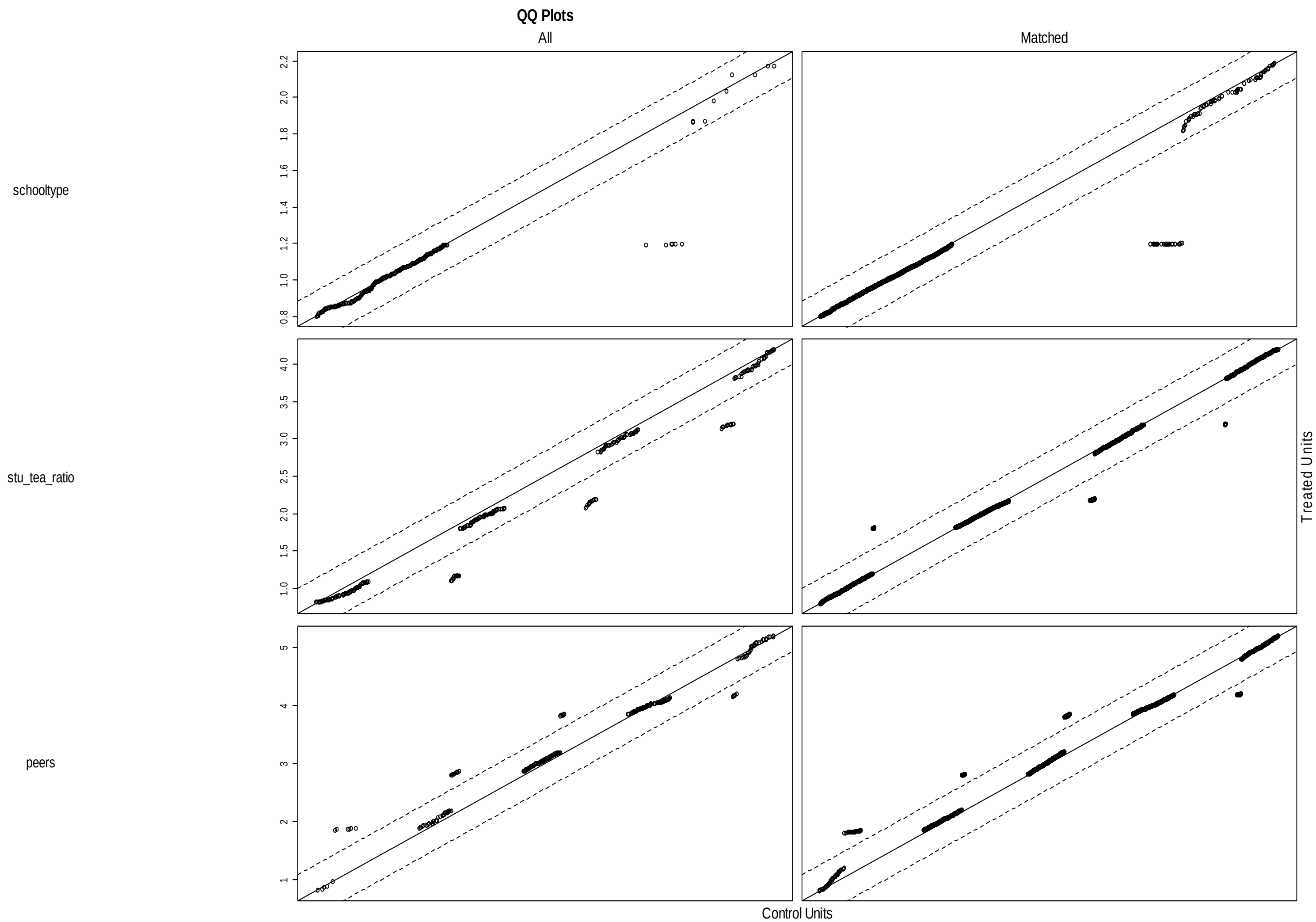

Figure D12. Imputation 1 QQ-plots for the sample of CTE (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for school type (schooltype), student-teacher ratio (stu_tea_ratio), and percent peers college-bound (peers). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

APPENDIX E

COVARIATE BALANCE HYPOTHESIS TESTS

FOR

COLLEGE-PREPARATORY AND GENERAL-TRACK STUDENTS

Table E1
Imputation 1 - Differences Between College-preparatory (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238530.25	80080.018	203063.32	87646.771	-4.937***	662		0.42
Gender	1.64	.481	1.51	.500		1	9.847**	0.26
Race/ethnicity (dummy 1)	.17	.378	.32	.465		1	14.758***	-0.35
Race/ethnicity (dummy 2)	.15	.355	.22	.416		1	4.949*	-0.18
Urbanicity	.77	.419	.72	.448		1	1.878	0.12
Household poverty ratio (square root)	18.75	6.537	14.49	6.369	-7.898***	662		0.66
Grades received in eighth grade	6.92	1.146	5.25	1.618	-13.310***	662		1.19
PIAT math standard score	105.22	14.381	91.51	14.368	-11.335***	662		0.95
Work-based learning	.10	.298	.16	.370		1	4.871*	-0.18
Remedial English and/or math	.06	.245	.19	.390		1	16.917***	-0.40
ESL and/or bilingual program	.08	.270	.11	.312		1	1.446	-0.10
Educational and/or physical handicap	.01	.121	.06	.239		1	6.767**	-0.26
Attitudes toward school	15.00	2.910	16.45	2.880	5.964***	662		-0.50
Number of days absent from school	2.87	3.565	6.25	8.633	5.387***	662		-0.51
Ever suspended from school	.07	.262	.40	.491		1	73.275***	-0.84
School type	1.16	.365	1.06	.231		1	17.849***	0.33
Student-teacher ratio	2.41	1.044	2.40	1.077	-.125	662		0.01
Percent peers college-bound	3.78	.896	3.37	1.043	-4.898***	662		0.42

[a]$\underline{n}$ = 204 [b]$\underline{n}$ = 460 [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2{}_{t} + \sigma^2{}_{c} / 2}$
*$p < .05$ **$p < .01$ ***$p < .001$

Table E2
Imputation 1 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*T*	*df*	χ^2	*d*[c]
Survey weight	238354.39	80288.037	233512.55	83001.108	-.605	417		0.06
Gender	1.65	.477	1.66	.476		1	.001	-0.02
Race/ethnicity (dummy 1)	.17	.380	.18	.381		1	.001	-0.03
Race/ethnicity (dummy 2)	.15	.356	.15	.353		1	.017	0.00
Urbanicity	.77	.422	.74	.438		1	.523	0.07
Household poverty ratio (square root)	18.61	6.415	18.21	6.120	-.655	417		0.06
Grades received in eighth grade	6.90	1.155	6.99	1.240	.785	417		-0.08
PIAT math standard score	104.49	13.658	103.59	13.578	-.671	417		0.07
Work-based learning	.10	.303	.11	.315		1	.110	-0.03
Remedial English and/or math	.07	.249	.06	.234		1	.116	0.04
ESL and/or bilingual program	.08	.275	.09	.283		1	.017	-0.04
Educational and/or physical handicap	.02	.123	.01	.119		1	.025	0.08
Attitudes toward school	15.09	2.897	15.02	2.739	-.251	417		0.02
Number of days absent from school	2.90	3.609	2.67	3.466	-.661	417		0.07
Ever suspended from school	.08	.267	.07	.260		1	.035	0.04
School type	1.15	.356	1.17	.376		1	.391	-0.05
Student-teacher ratio	2.41	1.046	2.43	1.036	.148	417		-0.02
Percent peers college-bound	3.76	.899	3.78	.913	.185	417		-0.02

[a]$\underline{n}$ = 196 [b]$\underline{n}$ = 223 [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2{}_{t} + \sigma^2{}_{c} / 2}$
*$p < .05$ **$p < .01$ ***$p < .001$

Table E3

Imputation 1 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238372.63	80318.288	234035.04	78697.207	-.637	617		0.05
Gender	1.65	.477	1.64	.480		1	.100	0.02
Race/ethnicity (dummy 1)	.17	.381	.15	.357		1	.617	0.05
Race/ethnicity (dummy 2)	.15	.353	.17	.373		1	.491	-0.06
Urbanicity	.78	.419	.74	.439		1	.895	0.09
Household poverty ratio (square root)	18.73	6.560	18.19	6.073	-1.016	617		0.09
Grades received in eighth grade	6.91	1.150	6.98	1.223	.688	617		-0.06
PIAT math standard score	104.50	13.598	104.03	13.897	-.401	617		0.03
Work-based learning	.10	.301	.08	.270		1	.780	0.07
Remedial English and/or math	.07	.247	.05	.222		1	.396	0.09
ESL and/or bilingual program	.08	.272	.09	.286		1	.194	-0.04
Educational and/or physical handicap	.02	.122	.01	.111		1	.100	0.09
Attitudes toward school	15.02	2.933	14.59	2.739	-1.783	617		0.15
Number of days absent from school	2.88	3.587	2.62	3.053	-.922	617		0.08
Ever suspended from school	.08	.264	.07	.251		1	.140	0.04
School type	1.16	.368	1.20	.397		1	1.149	-0.10
Student-teacher ratio	2.40	1.047	2.43	.993	.370	617		-0.03
Percent peers college-bound	3.78	.899	3.90	.914	1.591	617		-0.13

[a]$\underline{n} = 200$ [b]$\underline{n} = 419$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c}} / 2$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E4

Imputation 2 - Differences Between College-preparatory (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238530.25	80080.018	203063.32	87646.771	-4.937***	662		0.42
Gender	1.64	.481	1.51	.500		1	9.847**	0.26
Race/ethnicity (dummy 1)	.17	.378	.32	.465		1	14.758***	-0.35
Race/ethnicity (dummy 2)	.15	.355	.22	.416		1	4.949*	-0.18
Urbanicity	.77	.419	.73	.444		1	1.441	0.09
Household poverty ratio (square root)	18.35	6.471	14.40	6.449	-7.275***	662		0.61
Grades received in eighth grade	6.92	1.161	5.25	1.608	-13.303***	662		1.19
PIAT math standard score	105.09	14.326	91.21	14.634	-11.350***	662		0.96
Work-based learning	.10	.298	.17	.374		1	5.449*	-0.21
Remedial English and/or math	.06	.245	.19	.390		1	16.917***	-0.40
ESL and/or bilingual program	.08	.270	.11	.312		1	1.446	-0.10
Educational and/or physical handicap	.01	.121	.06	.239		1	6.767**	-0.26
Attitudes toward school	15.00	2.886	16.46	2.887	5.987***	662		-0.51
Number of days absent from school	2.80	3.122	6.14	8.590	5.396***	662		-0.52
Ever suspended from school	.07	.262	.40	.491		1	73.275***	-0.84
School type	1.16	.365	1.06	.231		1	17.849***	0.33
Student-teacher ratio	2.38	1.041	2.39	1.076	.155	662		-0.01
Percent peers college-bound	3.78	.896	3.37	1.043	-4.818***	662		0.42

[a]$\underline{n} = 204$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c}} / 2$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E5

Imputation 2 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	d^c
Survey weight	238205.96	80408.093	237811.55	78758.133	-.051	423		0.00
Gender	1.67	.472	1.65	.477		1	.084	0.04
Race/ethnicity (dummy 1)	.18	.384	.17	.373		1	.118	0.03
Race/ethnicity (dummy 2)	.14	.351	.14	.348		1	.008	0.00
Urbanicity	.77	.422	.75	.432		1	.216	0.05
Household poverty ratio (square root)	18.25	6.495	17.92	6.333	-.519	423		0.05
Grades received in eighth grade	6.88	1.169	6.96	1.263	.637	423		-0.07
PIAT math standard score	103.96	13.240	103.39	13.212	-.446	423		0.04
Work-based learning	.10	.303	.10	.295		1	.042	0.00
Remedial English and/or math	.07	.249	.05	.219		1	.370	0.09
ESL and/or bilingual program	.08	.267	.08	.279		1	.059	0.00
Educational and/or physical handicap	.02	.123	.01	.110		1	.037	0.09
Attitudes toward school	15.02	2.916	14.92	2.557	-.387	423		0.04
Number of days absent from school	2.78	3.144	2.55	2.581	-.831	423		0.08
Ever suspended from school	.08	.267	.06	.244		1	.394	0.08
School type	1.15	.361	1.18	.385		1	.512	-0.08
Student-teacher ratio	2.38	1.048	2.38	1.064	.041	423		0.00
Percent peers college-bound	3.77	.891	3.83	.920	.621	423		-0.07

[a]$\underline{n} = 196$ [b]$\underline{n} = 229$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^*p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E6

Imputation 2 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	d^c
Survey weight	238372.63	80318.288	247752.46	74841.868	1.420	610		-0.12
Gender	1.65	.477	1.67	.471		1	.094	-0.04
Race/ethnicity (dummy 1)	.18	.381	.14	.352		1	1.047	0.11
Race/ethnicity (dummy 2)	.15	.353	.11	.313		1	1.621	0.12
Urbanicity	.77	.419	.72	.449		1	2.043	0.12
Household poverty ratio (square root)	18.32	6.491	18.54	6.529	.394	610		-0.03
Grades received in eighth grade	6.90	1.165	6.88	1.228	-.153	610		0.02
PIAT math standard score	104.38	13.531	103.76	14.015	-.514	610		0.05
Work-based learning	.10	.301	.09	.280		1	.373	0.03
Remedial English and/or math	.06	.247	.07	.254		1	.019	-0.04
ESL and/or bilingual program	.08	.272	.07	.262		1	.100	0.04
Educational and/or physical handicap	.02	.122	.01	.104		1	.086	0.09
Attitudes toward school	15.02	2.909	14.88	2.514	-.633	610		0.05
Number of days absent from school	2.81	3.138	2.49	2.694	-1.264	610		0.11
Ever suspended from school	.08	.264	.07	.248		1	.189	0.04
School type	1.16	.368	1.19	.396		1	1.052	-0.08
Student-teacher ratio	2.37	1.043	2.46	1.024	.971	610		-0.09
Percent peers college-bound	3.77	.899	3.87	.896	1.288	610		-0.11

[a]$\underline{n} = 200$ [b]$\underline{n} = 412$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^*p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E7

Imputation 3 - Differences Between College-preparatory (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	d^c
Survey weight	238530.25	80080.018	203063.32	87646.771	-4.937^{***}	662		0.42
Gender	1.64	.481	1.51	.500		1	9.847^{**}	0.26
Race/ethnicity (dummy 1)	.17	.378	.32	.465		1	14.758^{***}	-0.35
Race/ethnicity (dummy 2)	.15	.355	.22	.416		1	4.949^{*}	-0.18
Urbanicity	.78	.412	.73	.444		1	2.171	0.12
Household poverty ratio (square root)	18.61	6.391	14.42	6.388	-7.795^{***}	662		0.66
Grades received in eighth grade	6.91	1.154	5.28	1.609	-13.032^{***}	662		1.16
PIAT math standard score	104.99	14.329	91.37	14.582	-11.163^{***}	662		0.94
Work-based learning	.10	.298	.17	.372		1	5.157^{*}	-0.21
Remedial English and/or math	.06	.245	.19	.390		1	16.917^{***}	-0.40
ESL and/or bilingual program	.08	.270	.11	.312		1	1.446	-0.10
Educational and/or physical handicap	.01	.121	.06	.239		1	6.767^{**}	-0.26
Attitudes toward school	15.00	2.890	16.45	2.884	5.944^{***}	662		-0.50
Number of days absent from school	2.75	3.077	6.24	8.650	5.597^{***}	662		-0.54
Ever suspended from school	.07	.262	.40	.491		1	73.275^{***}	-0.84
School type	1.16	.365	1.06	.231		1	17.849^{***}	0.33
Student-teacher ratio	2.42	1.040	2.40	1.077	-.137	662		0.02
Percent peers college-bound	3.78	.896	3.37	1.043	-4.871^{***}	662		0.42

[a]$\underline{n} = 204$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E8

Imputation 3 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	d^c
Survey weight	238506.63	80317.233	237020.39	80346.950	-.190	421		0.02
Gender	1.66	.475	1.67	.472		1	.032	-0.02
Race/ethnicity (dummy 1)	.18	.383	.19	.389		1	.047	-0.03
Race/ethnicity (dummy 2)	.14	.350	.13	.342		1	.078	0.03
Urbanicity	.79	.411	.75	.435		1	.894	0.09
Household poverty ratio (square root)	18.56	6.456	18.57	6.125	.010	421		0.00
Grades received in eighth grade	6.88	1.161	6.99	1.189	.946	421		-0.09
PIAT math standard score	104.14	13.235	103.67	13.193	-.368	421		0.04
Work-based learning	.10	.303	.13	.337		1	.738	-0.09
Remedial English and/or math	.07	.249	.05	.226		1	.315	0.08
ESL and/or bilingual program	.08	.274	.08	.276		1	.011	0.00
Educational and/or physical handicap	.02	.123	.01	.095		1	.367	0.09
Attitudes toward school	15.01	2.924	15.07	2.641	.229	421		-0.02
Number of days absent from school	2.75	3.098	2.49	2.569	-.956	421		0.09
Ever suspended from school	.08	.266	.08	.269		1	.018	0.00
School type	1.16	.370	1.14	.350		1	.356	0.06
Student-teacher ratio	2.41	1.049	2.44	1.042	.261	421		-0.03
Percent peers college-bound	3.79	.893	3.84	.936	.533	421		-0.05

[a]$\underline{n} = 197$ [b]$\underline{n} = 226$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E9

Imputation 3 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238983.66	80053.490	235266.33	80407.065	-.536	608		0.05
Gender	1.66	.475	1.67	.472		1	.042	-0.02
Race/ethnicity (dummy 1)	.18	.382	.18	.388		1	.040	0.00
Race/ethnicity (dummy 2)	.14	.349	.15	.356		1	.064	-0.03
Urbanicity	.78	.413	.77	.422		1	.174	0.02
Household poverty ratio (square root)	18.58	6.427	19.03	6.165	.825	608		-0.07
Grades received in eighth grade	6.89	1.161	6.91	1.233	.173	608		-0.02
PIAT math standard score	104.08	13.265	103.90	13.554	-.149	608		0.01
Work-based learning	.10	.301	.11	.314		1	.114	-0.03
Remedial English and/or math	.07	.248	.07	.249		1	.000	0.00
ESL and/or bilingual program	.08	.273	.06	.233		1	1.060	0.08
Educational and/or physical handicap	.02	.122	.01	.113		1	.088	0.09
Attitudes toward school	15.01	2.916	15.25	2.539	1.058	608		-0.09
Number of days absent from school	2.74	3.085	2.52	2.691	-.877	608		0.08
Ever suspended from school	.08	.265	.08	.268		1	.012	0.00
School type	1.16	.368	1.16	.370		1	.005	0.00
Student-teacher ratio	2.41	1.045	2.44	1.063	.261	608		-0.03
Percent peers college-bound	3.78	.892	3.79	.924	.115	608		-0.01

[a]$\underline{n}$ = 199 [b]$\underline{n}$ = 411 [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E10

Imputation 4 - Differences Between College-preparatory (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238530.25	80080.018	203063.32	87646.771	-4.937***	662		0.42
Gender	1.64	.481	1.51	.500		1	9.847**	0.26
Race/ethnicity (dummy 1)	.17	.378	.32	.465		1	14.758***	-0.35
Race/ethnicity (dummy 2)	.15	.355	.22	.416		1	4.949*	-0.18
Urbanicity	.77	.419	.72	.448		1	1.878	0.12
Household poverty ratio (square root)	18.48	6.565	14.61	6.579	-6.998***	662		0.59
Grades received in eighth grade	6.91	1.177	5.26	1.610	-13.112***	662		1.17
PIAT math standard score	105.10	14.387	91.48	14.402	-11.252***	662		0.95
Work-based learning	.10	.298	.16	.370		1	4.871*	-0.18
Remedial English and/or math	.06	.245	.19	.390		1	16.917***	-0.40
ESL and/or bilingual program	.08	.270	.11	.312		1	1.446	-0.10
Educational and/or physical handicap	.01	.121	.06	.239		1	6.767**	-0.26
Attitudes toward school	15.01	2.911	16.45	2.889	5.894***	662		-0.50
Number of days absent from school	2.75	3.093	6.15	8.585	5.505***	662		-0.53
Ever suspended from school	.07	.262	.40	.491		1	73.275***	-0.84
School type	1.16	.365	1.06	.231		1	17.849***	0.33
Student-teacher ratio	2.41	1.044	2.41	1.086	-.034	662		0.00
Percent peers college-bound	3.78	.896	3.38	1.044	-4.791***	662		0.41

[a]$\underline{n}$ = 204 [b]$\underline{n}$ = 460 [c]$d = M_t - M_c / \sigma_{pooled}$; *where* $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$
$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table E11

Imputation 4 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238055.36	80395.113	242869.70	78616.688	.623	423		-0.06
Gender	1.66	.475	1.61	.489		1	1.035	0.10
Race/ethnicity (dummy 1)	.18	.382	.15	.362		1	.340	0.08
Race/ethnicity (dummy 2)	.15	.354	.12	.329		1	.434	0.09
Urbanicity	.77	.419	.78	.416		1	.015	-0.02
Household poverty ratio (square root)	18.44	6.596	18.97	6.166	.861	423		-0.08
Grades received in eighth grade	6.88	1.181	6.95	1.240	.588	423		-0.06
PIAT math standard score	104.31	13.587	103.68	13.279	-.485	423		0.05
Work-based learning	.10	.301	.09	.294		1	.070	0.03
Remedial English and/or math	.07	.248	.05	.210		1	.919	0.09
ESL and/or bilingual program	.08	.273	.09	.283		1	.089	-0.04
Educational and/or physical handicap	.02	.122	.01	.105		1	.353	0.09
Attitudes toward school	15.04	2.933	15.19	2.603	.550	423		-0.05
Number of days absent from school	2.74	3.112	2.48	2.598	-.936	423		0.09
Ever suspended from school	.08	.265	.07	.251		1	.131	0.04
School type	1.16	.364	1.13	.339		1	.457	0.09
Student-teacher ratio	2.41	1.049	2.50	1.074	.866	423		-0.08
Percent peers college-bound	3.77	.897	3.70	.946	-.732	423		0.08

[a]$n = 199$ [b]$n = 226$ [c]$d = M_t – M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_t + \sigma^2_c / 2}$

[*]$p < .05$ [**]$p < .01$ [***]$p < .001$

Table E12

Imputation 4 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238372.63	80318.288	241321.86	77104.774	.438	612		-0.04
Gender	1.65	.477	1.61	.489		1	1.362	0.08
Race/ethnicity (dummy 1)	.17	.381	.15	.354		1	.782	0.05
Race/ethnicity (dummy 2)	.15	.353	.13	.332		1	.443	0.06
Urbanicity	.78	.419	.77	.424		1	.066	0.02
Household poverty ratio (square root)	18.46	6.587	19.06	6.283	1.098	612		-0.09
Grades received in eighth grade	6.89	1.181	7.01	1.239	1.100	612		-0.10
PIAT math standard score	104.39	13.598	104.54	12.656	.138	612		-0.01
Work-based learning	.10	.301	.10	.294		1	.052	0.00
Remedial English and/or math	.07	.247	.04	.192		1	2.081	0.14
ESL and/or bilingual program	.08	.272	.09	.279		1	.037	-0.04
Educational and/or physical handicap	.02	.122	.01	.113		1	.090	0.09
Attitudes toward school	15.03	2.934	15.29	2.695	1.121	612		-0.09
Number of days absent from school	2.75	3.109	2.63	2.786	-.506	612		0.04
Ever suspended from school	.08	.264	.08	.268		1	.010	0.00
School type	1.16	.368	1.14	.344		1	.542	0.06
Student-teacher ratio	2.41	1.047	2.41	1.067	.018	612		0.00
Percent peers college-bound	3.77	.899	3.70	.959	-.934	612		0.08

[a]$n = 200$ [b]$n = 414$ [c]$d = M_t – M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_t + \sigma^2_c / 2}$

[*]$p < .05$ [**]$p < .01$ [***]$p < .001$

Table E13

Imputation 5 - Differences Between College-preparatory (Treatment) and General-track Students (Control) Before Matching (Nonweighted)

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238530.25	80080.018	203063.32	87646.771	-4.937***	662		0.42
Gender	1.64	.481	1.51	.500		1	9.847**	0.26
Race/ethnicity (dummy 1)	.17	.378	.32	.465		1	14.758***	-0.35
Race/ethnicity (dummy 2)	.15	.355	.22	.416		1	4.949*	-0.18
Urbanicity	.78	.416	.72	.448		1	2.268	0.14
Household poverty ratio (square root)	18.36	6.602	14.33	6.447	-7.376***	662		0.62
Grades received in eighth grade	6.91	1.169	5.27	1.603	-13.106***	662		1.17
PIAT math standard score	105.13	14.343	91.38	14.315	-11.414***	662		0.96
Work-based learning	.10	.305	.16	.370		1	4.128*	-0.18
Remedial English and/or math	.06	.245	.19	.390		1	16.917***	-0.40
ESL and/or bilingual program	.08	.270	.11	.312		1	1.446	-0.10
Educational and/or physical handicap	.01	.121	.06	.239		1	6.767**	-0.26
Attitudes toward school	15.01	2.889	16.45	2.896	5.916***	662		-0.50
Number of days absent from school	2.72	3.042	6.14	8.586	5.536***	662		-0.53
Ever suspended from school	.07	.262	.40	.491		1	73.275***	-0.84
School type	1.16	.365	1.06	.231		1	17.849***	0.33
Student-teacher ratio	2.40	1.029	2.39	1.066	-.089	662		0.01
Percent peers college-bound	3.78	.896	3.37	1.043	-4.845***	662		0.42

[a]$\underline{n} = 204$ [b]$\underline{n} = 460$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

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

Table E14

Imputation 5 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After 5:1 Nearest-neighbor Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238372.64	80318.288	238380.98	80504.926	.001	420		0.00
Gender	1.65	.477	1.67	.471		1	.123	-0.04
Race/ethnicity (dummy 1)	.17	.381	.17	.376		1	.011	0.00
Race/ethnicity (dummy 2)	.15	.353	.13	.333		1	.321	0.06
Urbanicity	.78	.415	.76	.428		1	.209	0.05
Household poverty ratio (square root)	18.33	6.625	17.55	6.344	-1.237	420		0.12
Grades received in eighth grade	6.89	1.173	6.96	1.259	.550	420		-0.06
PIAT math standard score	104.42	13.552	105.02	14.003	.444	420		-0.04
Work-based learning	.11	.307	.10	.303		1	.002	0.03
Remedial English and/or math	.06	.247	.04	.201		1	1.274	0.09
ESL and/or bilingual program	.08	.272	.08	.277		1	.002	0.00
Educational and/or physical handicap	.02	.122	.01	.095		1	.323	0.09
Attitudes toward school	15.03	2.911	15.13	2.533	.394	420		-0.04
Number of days absent from school	2.73	3.057	2.60	2.920	-.431	420		0.04
Ever suspended from school	.08	.264	.07	.257		1	.013	0.04
School type	1.16	.368	1.14	.352		1	.206	0.06
Student-teacher ratio	2.39	1.031	2.51	1.013	1.167	420		-0.12
Percent peers college-bound	3.77	.899	3.77	.938	-.078	420		0.00

[a]$\underline{n} = 200$ [b]$\underline{n} = 222$ [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

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

Table E15

Imputation 5 - Differences Between College-preparatory (Treatment) and General-track Students (Control) After Full Matching

Variable	Treatment[a]		Control[b]					
	M	*SD*	*M*	*SD*	*t*	*df*	χ^2	*d*[c]
Survey weight	238372.64	80318.288	232962.96	82002.567	-.772	616		0.07
Gender	1.65	.477	1.65	.476		1	.000	0.00
Race/ethnicity (dummy 1)	.17	.381	.20	.401		1	.586	-0.08
Race/ethnicity (dummy 2)	.15	.353	.12	.325		1	.782	0.09
Urbanicity	.78	.415	.71	.455		1	3.556	0.16
Household poverty ratio (square root)	18.33	6.625	17.39	6.134	-1.738	616		0.15
Grades received in eighth grade	6.89	1.173	7.00	1.271	1.031	616		-0.09
PIAT math standard score	104.42	13.552	104.96	13.932	.455	616		-0.04
Work-based learning	.11	.307	.12	.321		1	.201	-0.03
Remedial English and/or math	.06	.247	.05	.216		1	.567	0.04
ESL and/or bilingual program	.08	.272	.07	.258		1	.133	0.04
Educational and/or physical handicap	.02	.122	.02	.126		1	.026	0.00
Attitudes toward school	15.03	2.911	15.17	2.458	.656	616		-0.05
Number of days absent from school	2.73	3.057	2.47	2.856	-1.032	616		0.09
Ever suspended from school	.08	.264	.08	.268		1	.005	0.00
School type	1.16	.368	1.12	.323		1	2.173	0.12
Student-teacher ratio	2.39	1.031	2.40	.998	.125	616		-0.01
Percent peers college-bound	3.77	.899	3.71	.949	-.828	616		0.06

[a]n = 200 [b]n = 418 [c]$d = M_t - M_c / \sigma_{pooled}$; where $\sigma_{pooled} = \sqrt{\sigma^2_{t} + \sigma^2_{c} / 2}$

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

APPENDIX F

COVARIATE BALANCE IMPROVEMENT

FOR

COLLEGE-PREPARATORY AND GENERAL-TRACK STUDENTS

Table F1

Imputation 1 - Standardized Mean Difference[a] Improvement for the Sample of College-preparatory (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.4429	0.0605	86.3483	0.4429	0.0542	87.7700
Gender	0.2732	-0.0067	97.5388	0.2732	0.0277	89.8717
Race/ethnicity (dummy 1)	-0.3801	-0.0061	98.4017	-0.3801	0.0663	82.5442
Race/ethnicity (dummy 2)	-0.2103	0.0072	96.5841	-0.2103	-0.0603	71.3396
Urbanicity	0.1208	0.0676	44.0353	0.1208	0.0831	31.1617
Household poverty ratio (square root)	0.6526	0.0614	90.5957	0.6526	0.0833	87.2414
Grades received in eighth grade	1.4544	-0.0805	94.4661	1.4544	-0.0619	95.7434
PIAT math standard score	0.9528	0.0622	93.4685	0.9528	0.0331	96.5253
Work-based learning	-0.2181	-0.0294	86.5263	-0.2181	0.0713	67.3179
Remedial English and/or math	-0.5033	0.0347	93.0996	-0.5033	0.0544	89.1917
ESL and/or bilingual program	-0.1123	-0.0215	80.8939	-0.1123	-0.0361	67.8519
Educational and/or physical handicap	-0.3826	0.0085	97.7896	-0.3826	0.0203	94.7047
Attitudes toward school	-0.4981	0.0238	95.2245	-0.4981	0.1477	70.3541
Number of days absent from school	-0.9476	0.0642	93.2285	-0.9476	0.0719	92.4151
Ever suspended from school	-1.2644	0.0156	98.7662	-1.2644	0.0293	97.6858
School type	0.2752	-0.0590	78.5595	0.2752	-0.0986	64.1873
Student-teacher ratio	0.0107	-0.0144	-34.2574	0.0107	-0.0308	-186.4030
Percent peers college-bound	0.4597	-0.0183	96.0168	0.4597	-0.1387	69.8304

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table F2

Imputation 2 - Standardized Mean Difference[a] Improvement for the Sample of College-preparatory (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.4429	0.0049	98.8879	0.4429	-0.1171	73.5533
Gender	0.2732	0.0315	88.4710	0.2732	-0.0276	89.8901
Race/ethnicity (dummy 1)	-0.3801	0.0328	91.3574	-0.3801	0.0810	78.6998
Race/ethnicity (dummy 2)	-0.2103	0.0074	96.4702	-0.2103	0.0995	52.7169
Urbanicity	0.1052	0.0416	60.4493	0.1052	0.1265	-20.2680
Household poverty ratio (square root)	0.6105	0.0501	91.8007	0.6105	-0.0342	94.4034
Grades received in eighth grade	1.4319	-0.0652	95.4473	1.4319	0.0137	99.0424
PIAT math standard score	0.9690	0.0400	95.8683	0.9690	0.0429	95.5768
Work-based learning	-0.2326	0.0217	90.6815	-0.2326	0.0492	78.8311
Remedial English and/or math	-0.5033	0.0646	87.1653	-0.5033	-0.0159	96.8343
ESL and/or bilingual program	-0.1123	-0.0293	73.8696	-0.1123	0.0222	80.2251
Educational and/or physical handicap	-0.3826	0.0254	93.3688	-0.3826	0.0332	91.3148
Attitudes toward school	-0.5037	0.0356	92.9401	-0.5037	0.0500	90.0662
Number of days absent from school	-1.0698	0.0740	93.0855	-1.0698	0.0993	90.7178
Ever suspended from school	-1.2644	0.0507	95.9902	-1.2644	0.0356	97.1849
School type	0.2752	-0.0737	73.2206	0.2752	-0.0938	65.9353
Student-teacher ratio	-0.0133	-0.0041	69.3093	-0.0133	-0.0828	-522.2727
Percent peers college-bound	0.4524	-0.0611	86.4952	0.4524	-0.1111	75.4431

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table F3

Imputation 3 - Standardized Mean Difference[a] Improvement for the Sample of College-preparatory (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.4429	0.0186	95.8095	0.4429	0.0464	89.5189
Gender	0.2732	-0.0160	94.1359	0.2732	-0.0185	93.2374
Race/ethnicity (dummy 1)	-0.3801	-0.0195	94.8761	-0.3801	-0.0201	94.7055
Race/ethnicity (dummy 2)	-0.2103	0.0224	89.3511	-0.2103	-0.0209	90.0466
Urbanicity	0.1307	0.0930	28.8686	0.1307	0.0350	73.2449
Household poverty ratio (square root)	0.6554	-0.0010	99.8480	0.6554	-0.0697	89.3722
Grades received in eighth grade	1.4097	-0.0940	93.3337	1.4097	-0.0157	98.8871
PIAT math standard score	0.9505	0.0331	96.5192	0.9505	0.0121	98.7266
Work-based learning	-0.2254	-0.0962	57.3073	-0.2254	-0.0338	85.0219
Remedial English and/or math	-0.5033	0.0498	90.1139	-0.5033	-0.0048	99.0506
ESL and/or bilingual program	-0.1123	-0.0057	94.9682	-0.1123	0.0854	23.9344
Educational and/or physical handicap	-0.3826	0.0505	86.8048	-0.3826	0.0173	95.4748
Attitudes toward school	-0.4994	-0.0214	95.7081	-0.4994	-0.0844	83.1054
Number of days absent from school	-1.1324	0.0857	92.4363	-1.1324	0.0696	93.8554
Ever suspended from school	-1.2644	-0.0078	99.3862	-1.2644	-0.0082	99.3535
School type	0.2752	0.0543	80.2703	0.2752	-0.0068	97.5247
Student-teacher ratio	0.0118	-0.0255	-115.6465	0.0118	-0.0229	-93.6960
Percent peers college-bound	0.4573	-0.0531	88.3783	0.4573	-0.0101	97.7873

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table F4

Imputation 4 - Standardized Mean Difference[a] Improvement for the Sample of College-preparatory (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.4429	-0.0601	86.4258	0.4429	-0.0368	91.6846
Gender	0.2732	0.1025	62.4897	0.2732	0.1023	62.5526
Race/ethnicity (dummy 1)	-0.3801	0.0572	84.9577	-0.3801	0.0755	80.1380
Race/ethnicity (dummy 2)	-0.2103	0.0637	69.7202	-0.2103	0.0530	74.8015
Urbanicity	0.1208	-0.0128	89.4062	0.1208	0.0220	81.7528
Household poverty ratio (square root)	0.5895	-0.0812	86.2290	0.5895	-0.0920	84.3988
Grades received in eighth grade	1.3966	-0.0589	95.7814	1.3966	-0.0982	92.9678
PIAT math standard score	0.9472	0.0440	95.3554	0.9472	-0.0107	98.8712
Work-based learning	-0.2181	0.0191	91.2388	-0.2181	0.0158	92.7732
Remedial English and/or math	-0.5033	0.0780	84.5043	-0.5033	0.1101	78.1237
ESL and/or bilingual program	-0.1123	-0.0264	76.4775	-0.1123	-0.0190	83.1213
Educational and/or physical handicap	-0.3826	0.0333	91.2916	-0.3826	0.0170	95.5561
Attitudes toward school	-0.4933	-0.0508	89.7107	-0.4933	-0.0921	81.3354
Number of days absent from school	-1.1006	0.0838	92.3834	-1.1006	0.0408	96.2923
Ever suspended from school	-1.2644	0.0307	97.5696	-1.2644	-0.0091	99.2815
School type	0.2752	0.0657	76.1283	0.2752	0.0627	77.2017
Student-teacher ratio	0.0029	-0.0856	-2814.4891	0.0029	-0.0016	47.2316
Percent peers college-bound	0.4500	0.0733	83.7199	0.4500	0.0843	81.2585

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

Table F5

Imputation 5 - Standardized Mean Difference[a] Improvement for the Sample of College-preparatory (Treatment) and General-track Students (Control)

Variable	Standardized mean difference (5:1 nearest-neighbor matching)			Standardized mean difference (full matching)		
	Pre-matching imbalance	Post-matching imbalance	% Balance improvement	Pre-matching imbalance	Post-matching imbalance	% Balance improvement
Survey weight	0.4429	-0.0001	99.9765	0.4429	0.0676	84.7473
Gender	0.2732	-0.0307	88.7651	0.2732	0.0011	99.6130
Race/ethnicity (dummy 1)	-0.3801	0.0150	96.0552	-0.3801	-0.0681	82.0783
Race/ethnicity (dummy 2)	-0.2103	0.0514	75.5625	-0.2103	0.0719	65.8217
Urbanicity	0.1335	0.0475	64.4136	0.1335	0.1731	-29.6101
Household poverty ratio (square root)	0.6104	0.1183	80.6129	0.6104	0.1425	76.6452
Grades received in eighth grade	1.3992	-0.0559	96.0044	1.3992	-0.0941	93.2759
PIAT math standard score	0.9588	-0.0416	95.6576	0.9588	-0.0377	96.0711
Work-based learning	-0.1973	0.0098	95.0085	-0.1973	-0.0382	80.6207
Remedial English and/or math	-0.5033	0.0939	81.3359	-0.5033	0.0649	87.1049
ESL and/or bilingual program	-0.1123	-0.0121	89.2613	-0.1123	0.0307	72.6447
Educational and/or physical handicap	-0.3826	0.0497	87.0028	-0.3826	-0.0087	97.7130
Attitudes toward school	-0.4986	-0.0361	92.7556	-0.4986	-0.0510	89.7671
Number of days absent from school	-1.1243	0.0412	96.3336	-1.1243	0.0853	92.4162
Ever suspended from school	-1.2644	0.0162	98.7153	-1.2644	-0.0104	99.1804
School type	0.2752	0.0434	84.2205	0.2752	0.1163	57.7476
Student-teacher ratio	0.0077	-0.1130	-1366.2500	0.0077	-0.0105	-36.2465
Percent peers college-bound	0.4548	0.0078	98.2829	0.4548	0.0741	83.6981

[a] Standardized mean difference = $100(\bar{x}_1 - \bar{x}_0)/[(s_1^2 + s_0^2)/2]^{1/2}$, where for each covariate, $\bar{x}_1$ and $\bar{x}_0$ are the sample means in the treatment and control groups, and s_1^2 and s_0^2 are the corresponding sample variances (Rosenbaum & Rubin, 1985).

APPENDIX G

PROPENSITY SCORE DISTRIBUTION JITTER PLOTS

FOR

COLLEGE-PREPARATORY AND GENERAL-TRACK STUDENTS

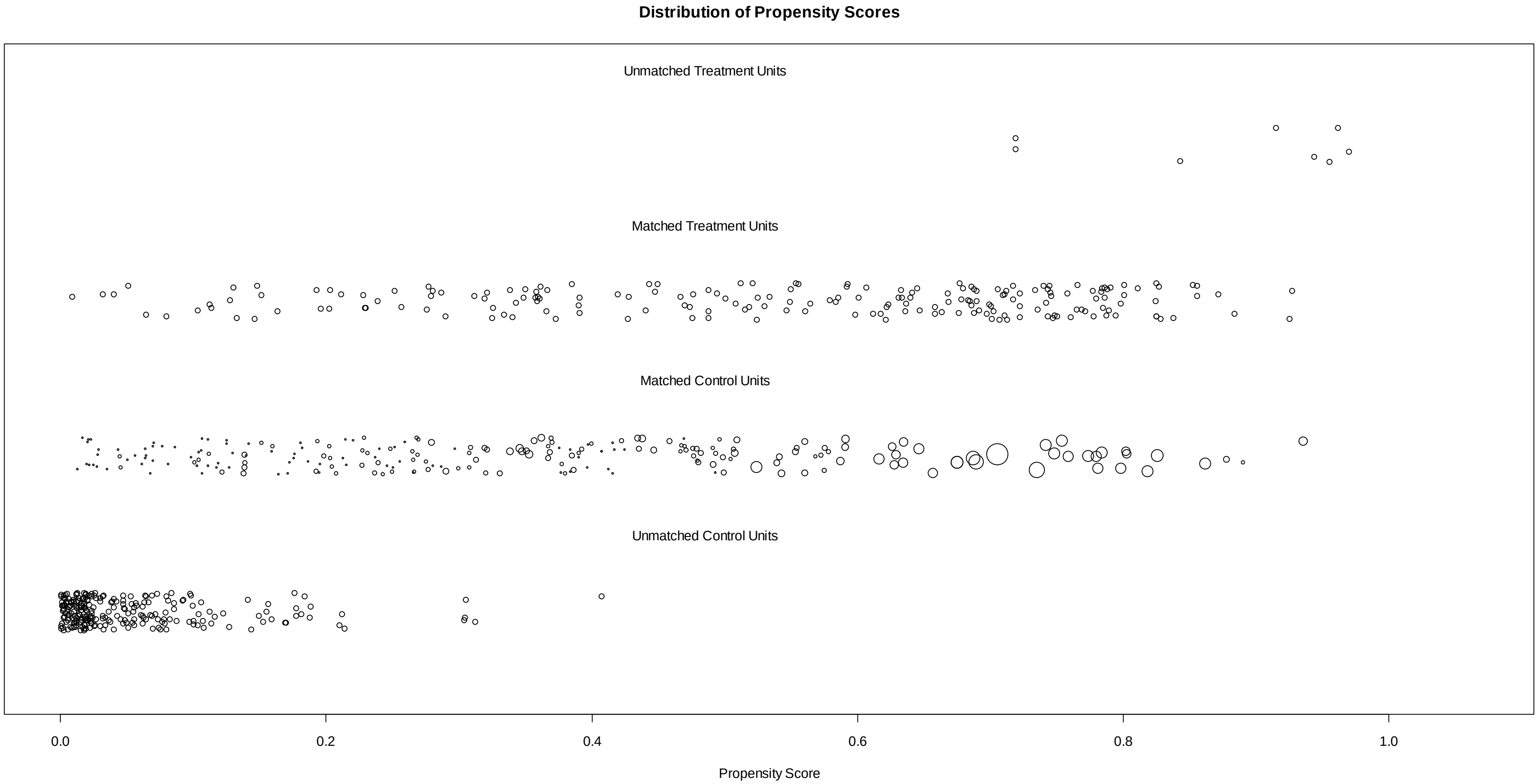

Figure G1. Imputation 1 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Eight treatment cases and 237 control cases remain unmatched due to common support and caliper size restrictions.

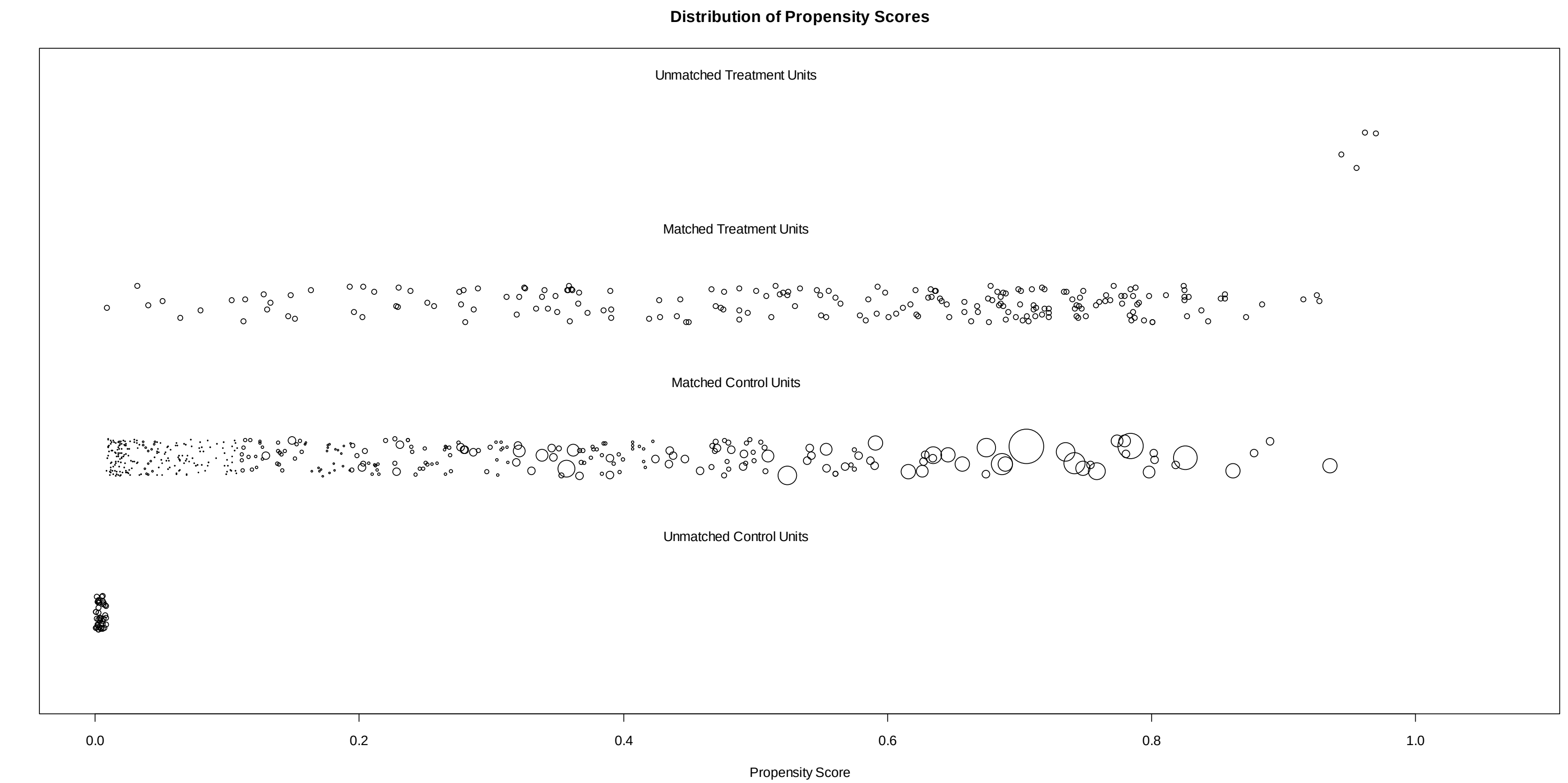

Figure G2. Imputation 1 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using full matching. Four treatment cases and 41 control cases remain unmatched due to common support restrictions.

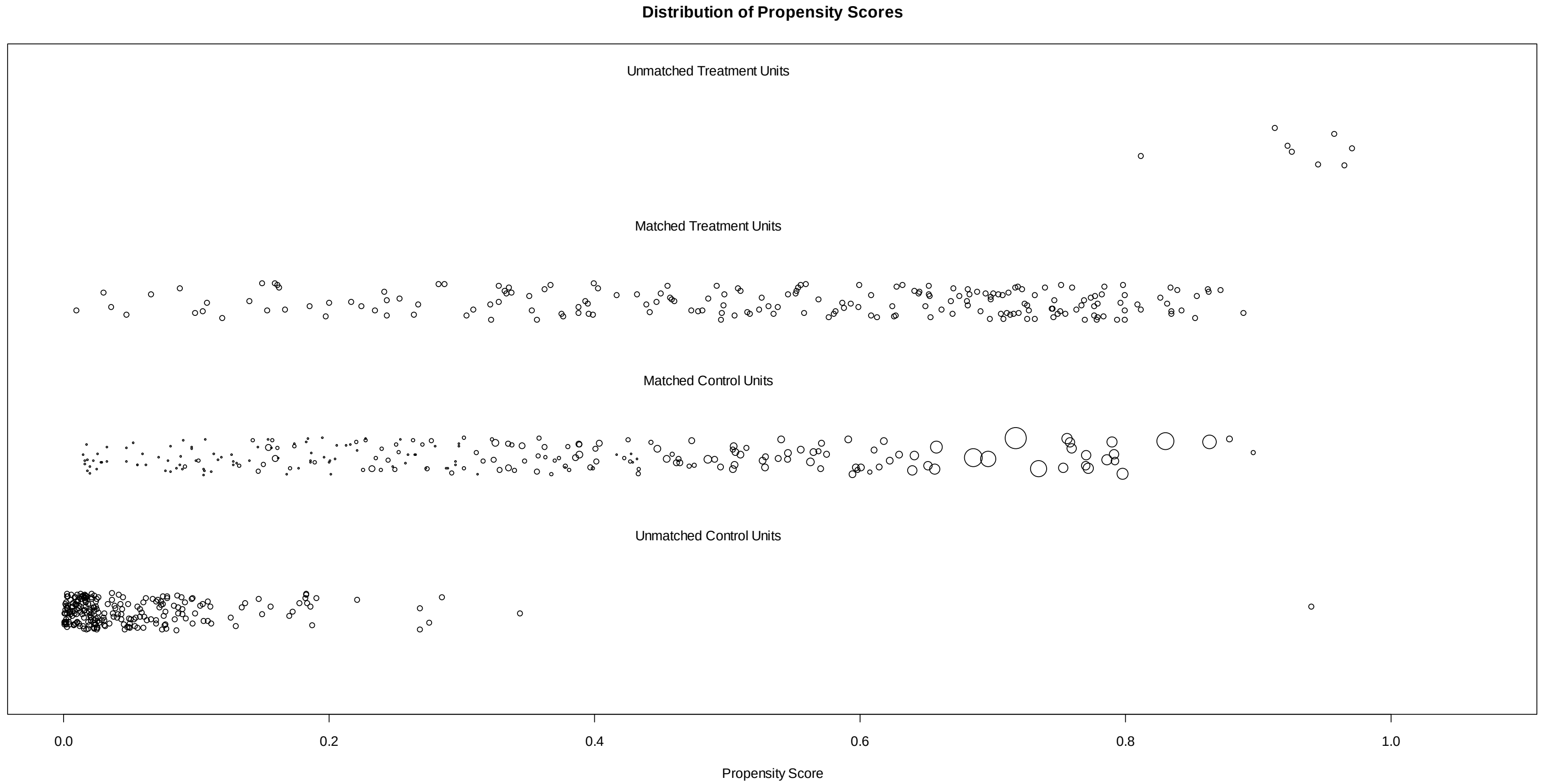

Figure G3. Imputation 2 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. Eight treatment cases and 231 control cases remain unmatched due to common support and caliper size restrictions.

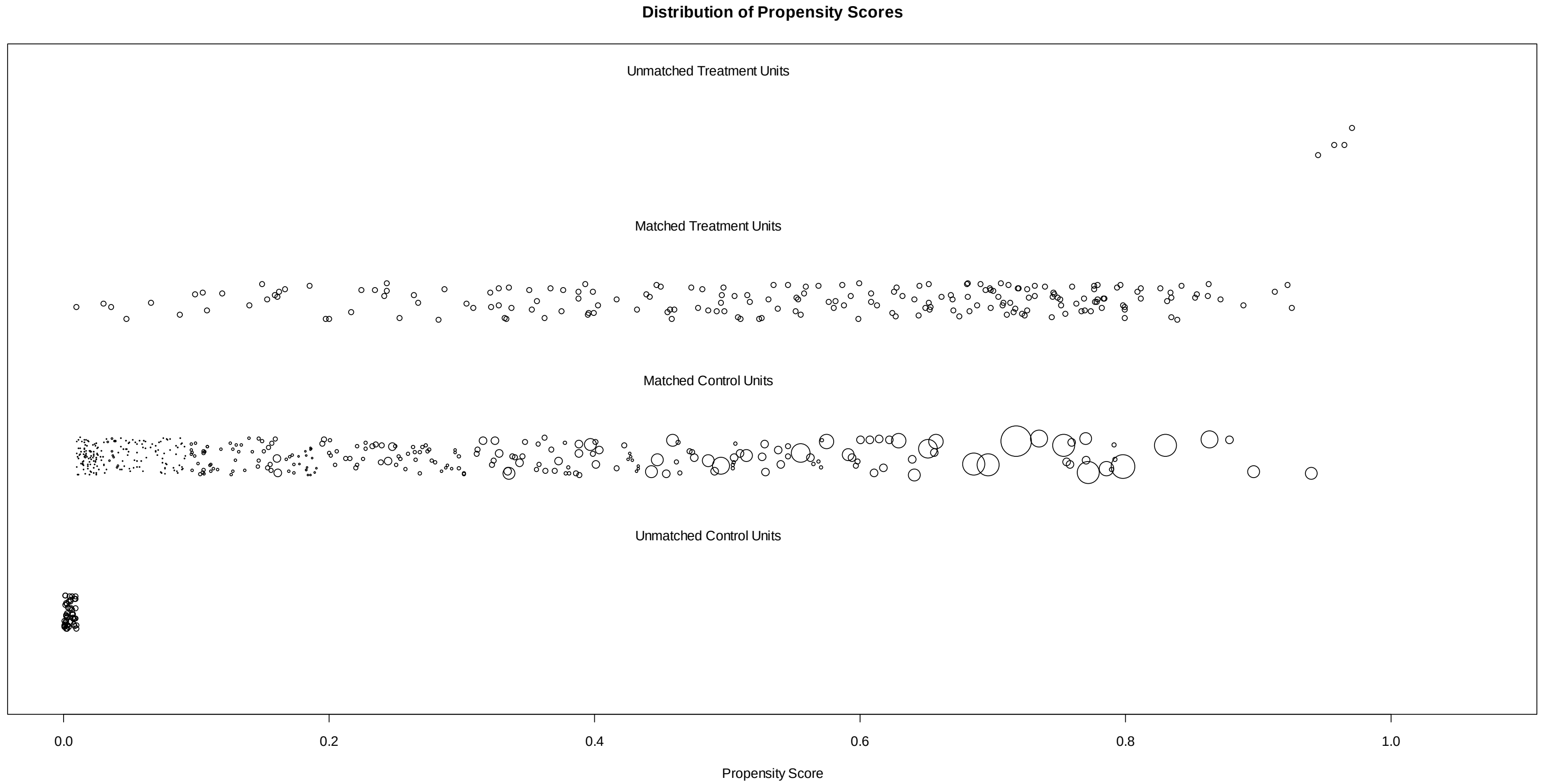

Figure G4. Imputation 2 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using full matching. Four treatment cases and 48 control cases remain unmatched due to common support restrictions.

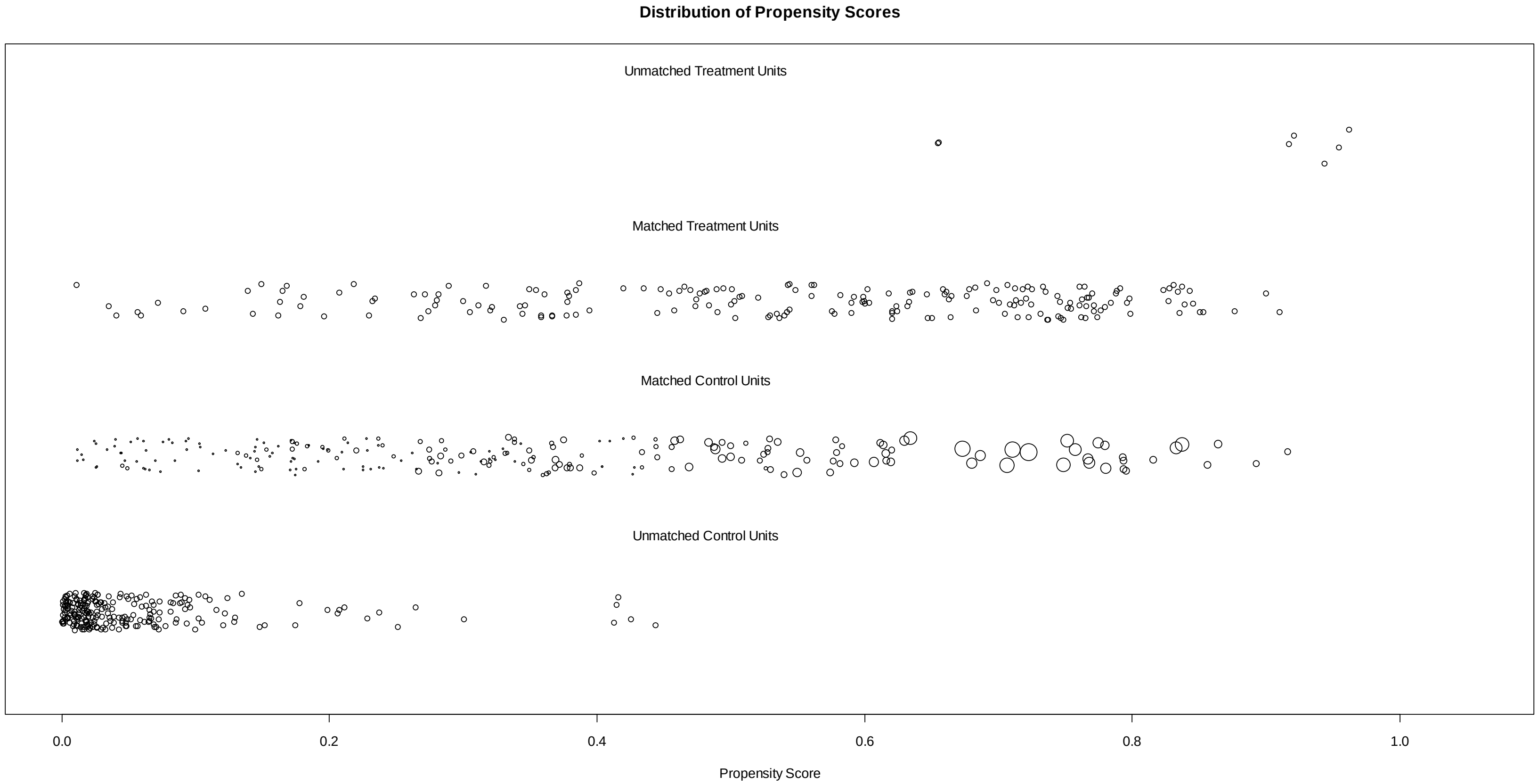

Figure G5. Imputation 3 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .06. Seven treatment cases and 234 control cases remain unmatched due to common support and caliper size restrictions.

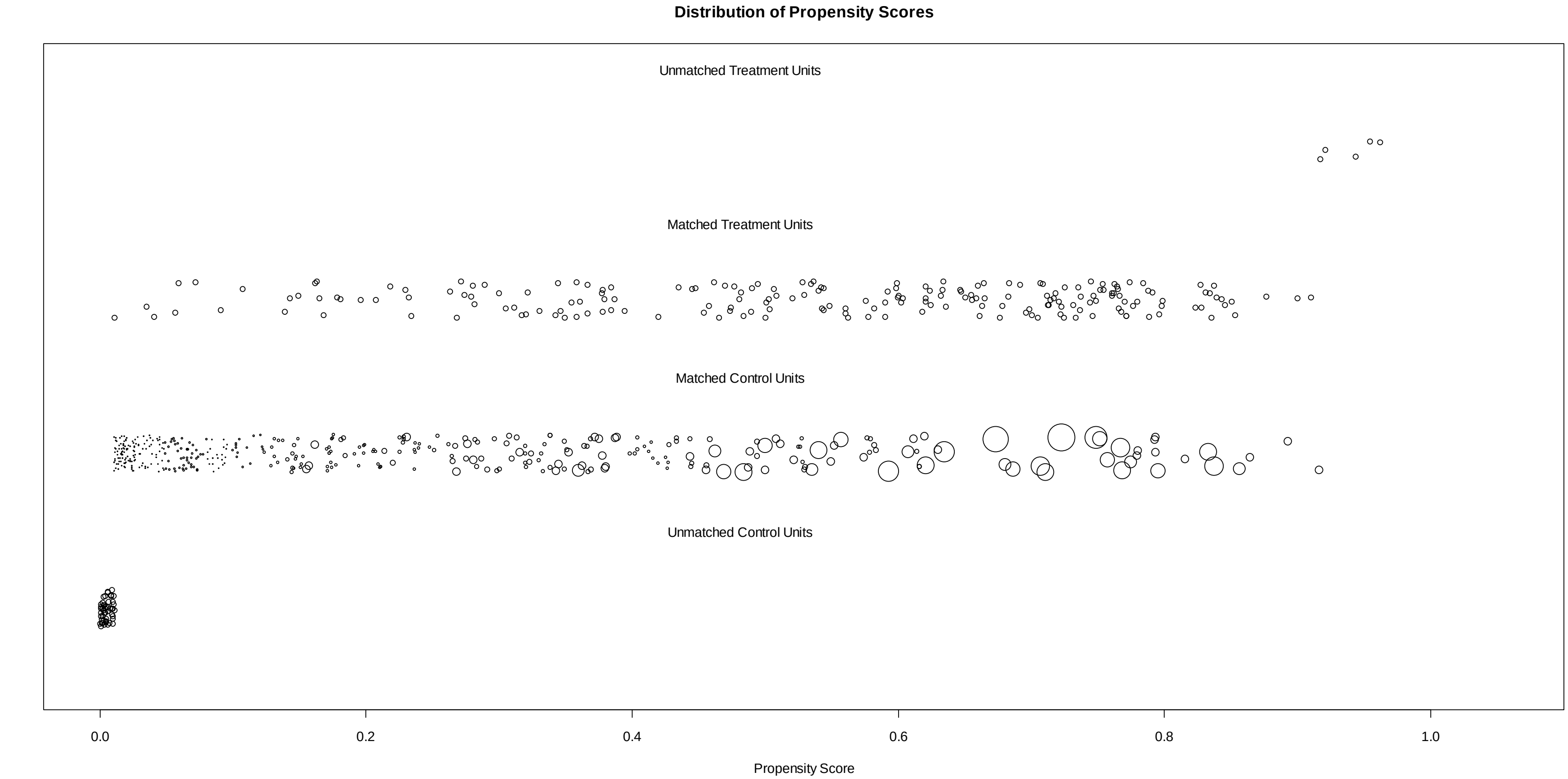

Figure G6. Imputation 3 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using full matching. Five treatment cases and 49 control cases remain unmatched due to common support restrictions.

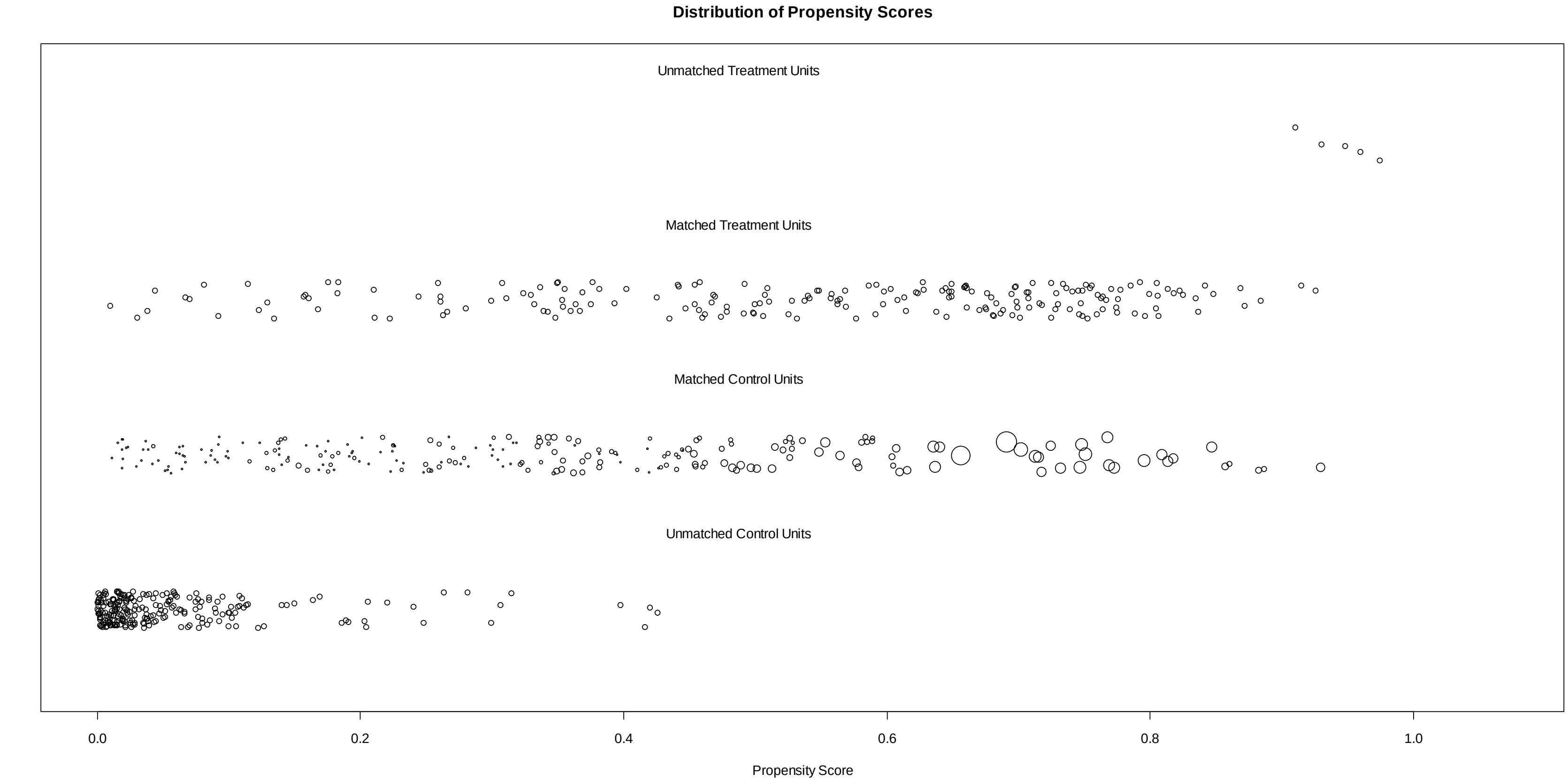

Figure G7. Imputation 4 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .06. Five treatment cases and 234 control cases remain unmatched due to common support and caliper size restrictions.

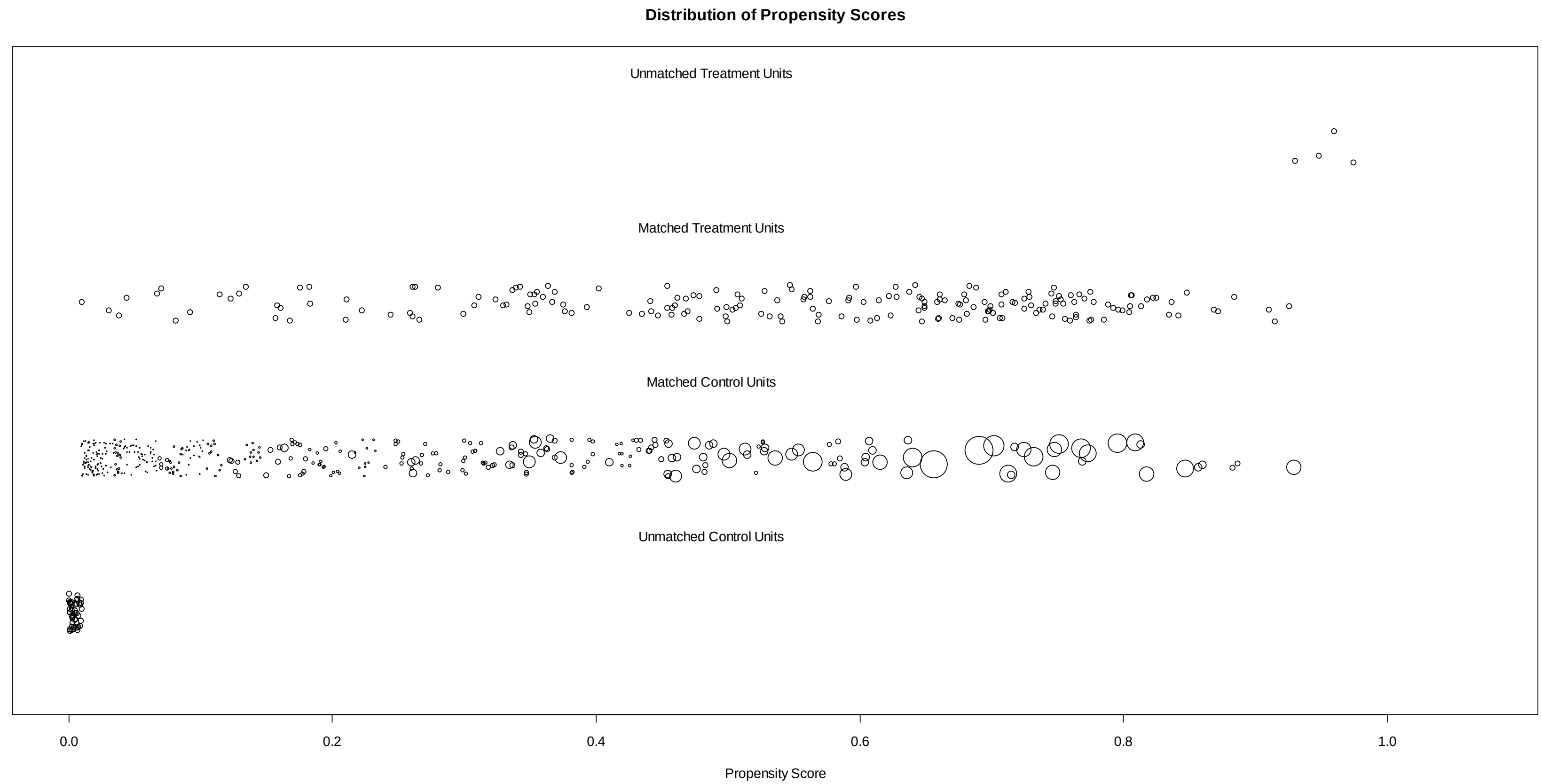

Figure G8. Imputation 4 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using full matching. Four treatment cases and 46 control cases remain unmatched due to common support restrictions.

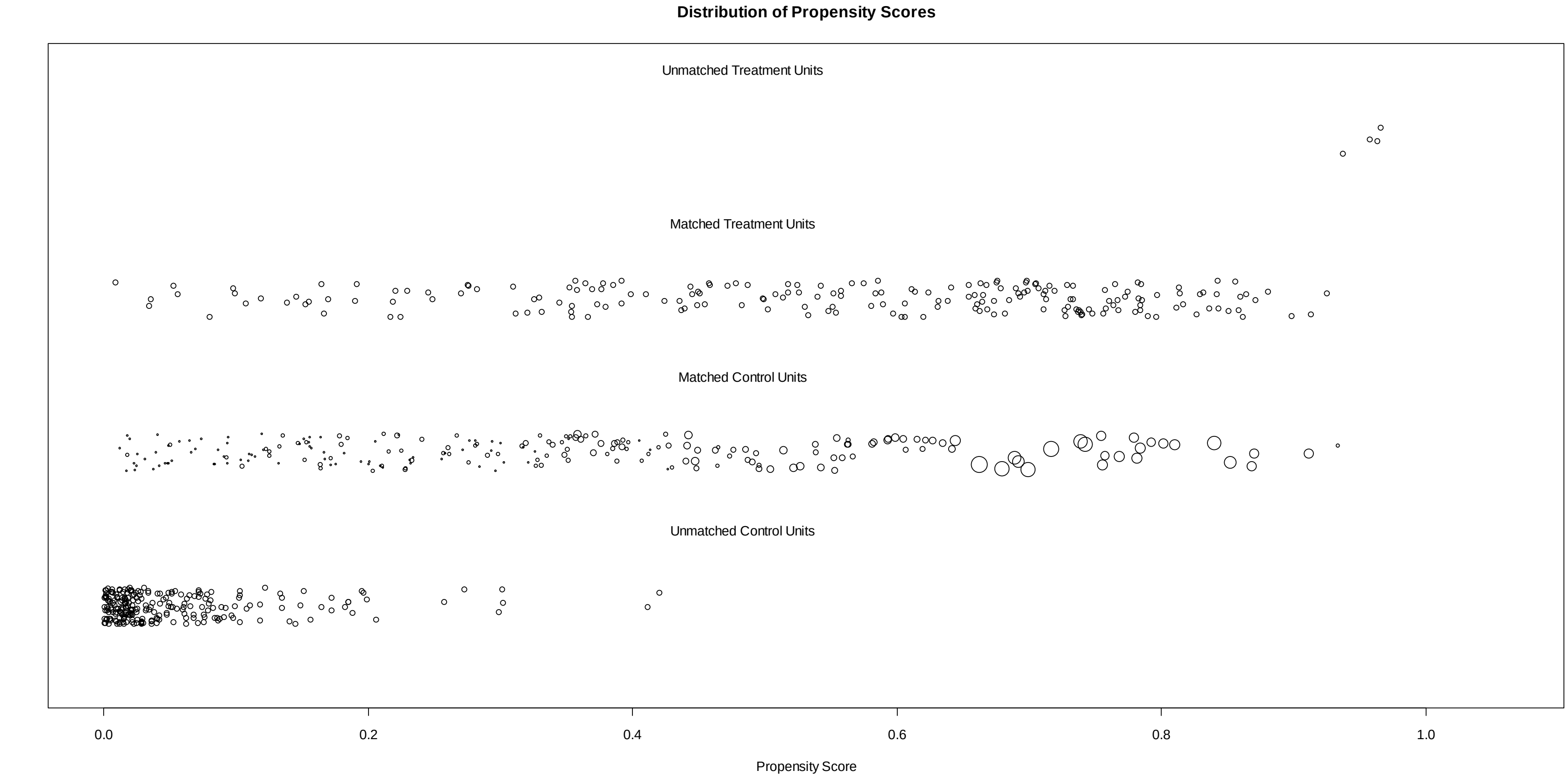

Figure G9. Imputation 5 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .06. Four treatment cases and 238 control cases remain unmatched due to common support and caliper size restrictions.

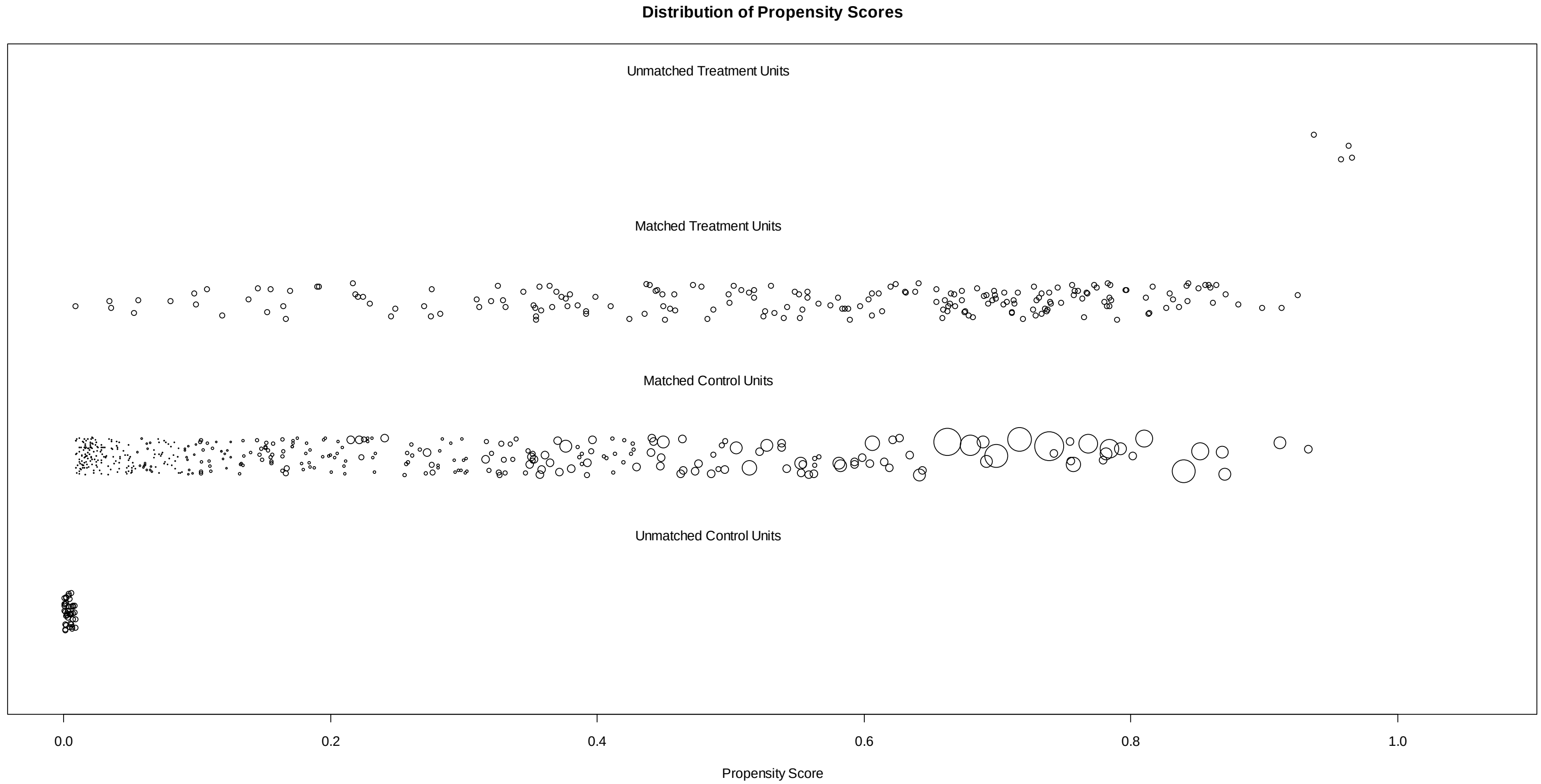

Figure G10. Imputation 5 jitter plot of the overall propensity score distribution for college-preparatory (treatment) and general-track students (control) using full matching. Four treatment cases and 42 control cases remain unmatched due to common support restrictions.

APPENDIX H

COVARIATE BALANCE QQ PLOTS FOR IMPUTATION CYCLE 1

FOR

COLLEGE-PREPARATORY AND GENERAL-TRACK STUDENTS

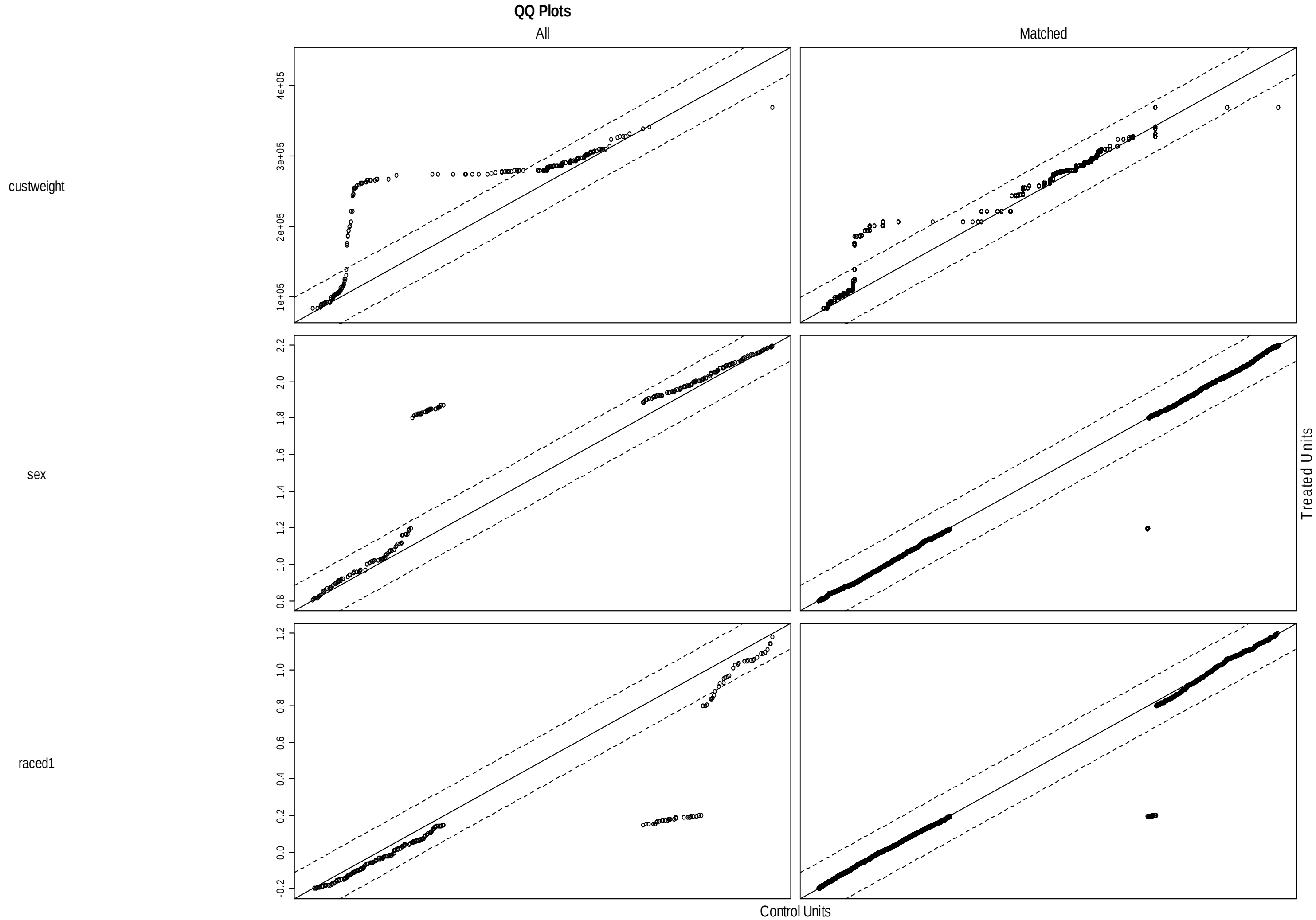

Figure H1. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for survey weight (custweight), gender (sex), and race/ethnicity dummy 1 (raced1). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

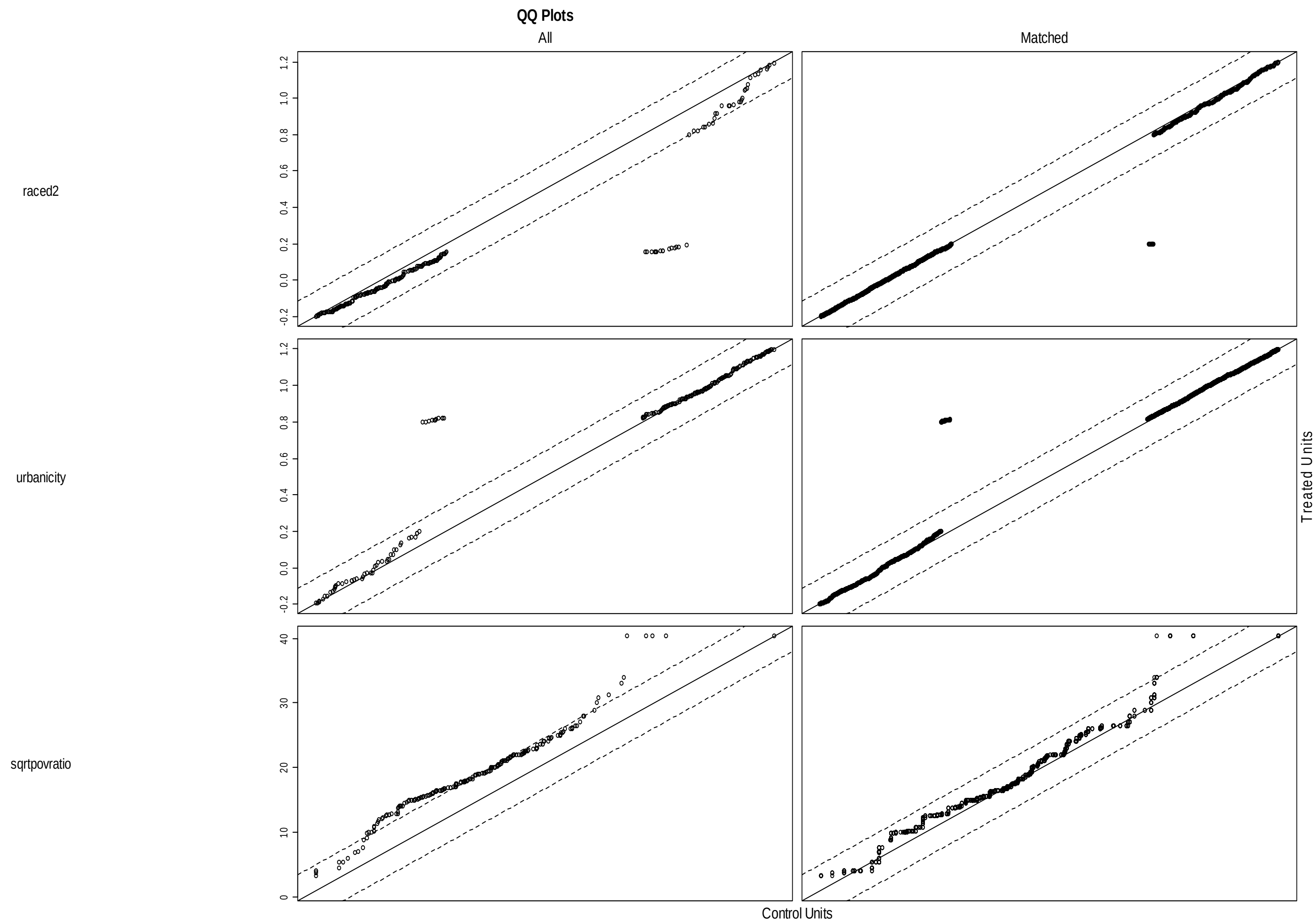

Figure H2. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for race/ethnicity dummy 2 (raced2), urbanicity (urbanicity), and household poverty ratio (sqrtpovratio). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

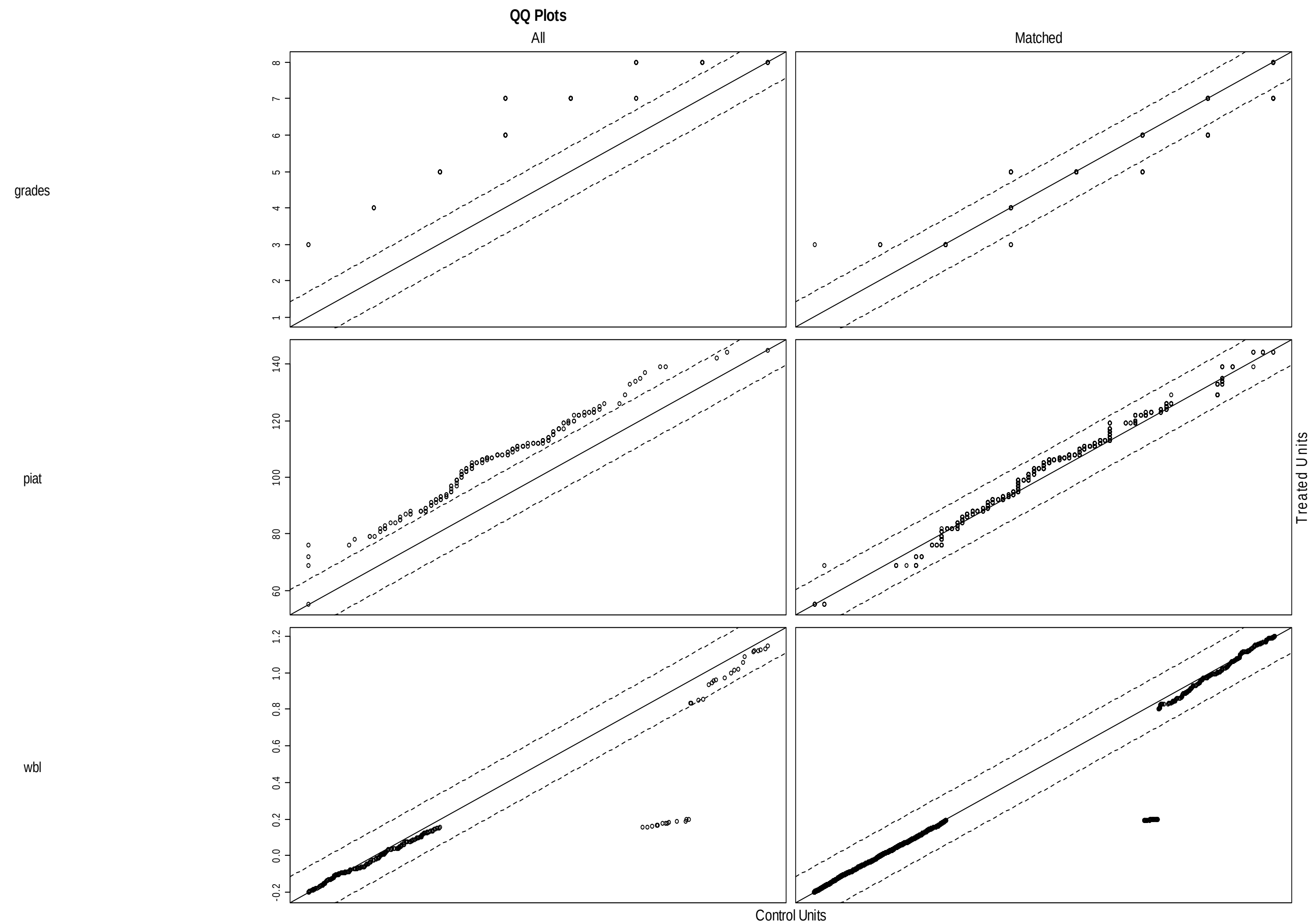

Figure H3. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for grades received in eighth grade (grades), PIAT math standard score (piat), and work-based learning (wbl). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

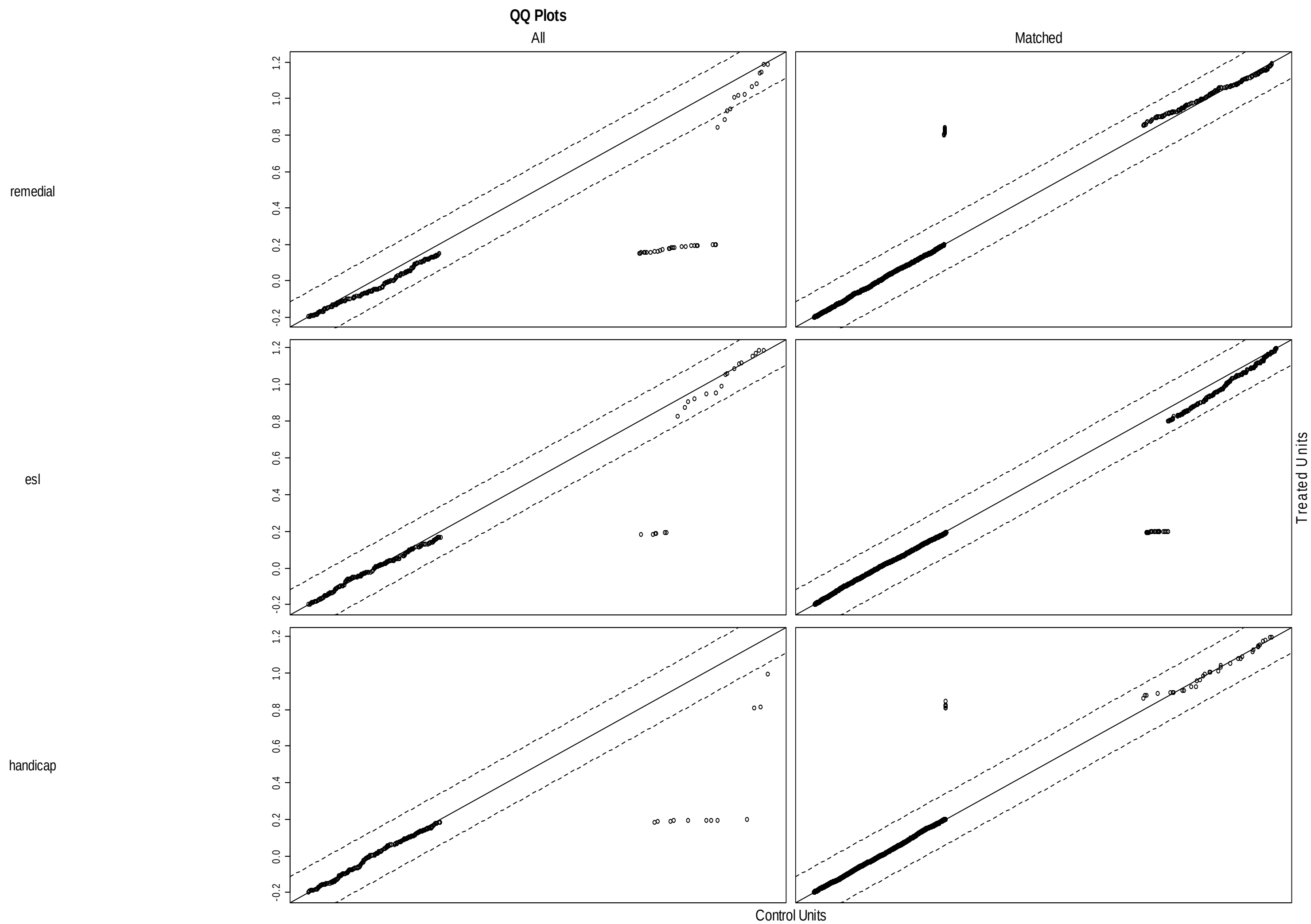

Figure H4. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for remedial English and/or math (remedial), ESL and/or bilingual program (esl), and educational and/or physical handicap (handicap). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

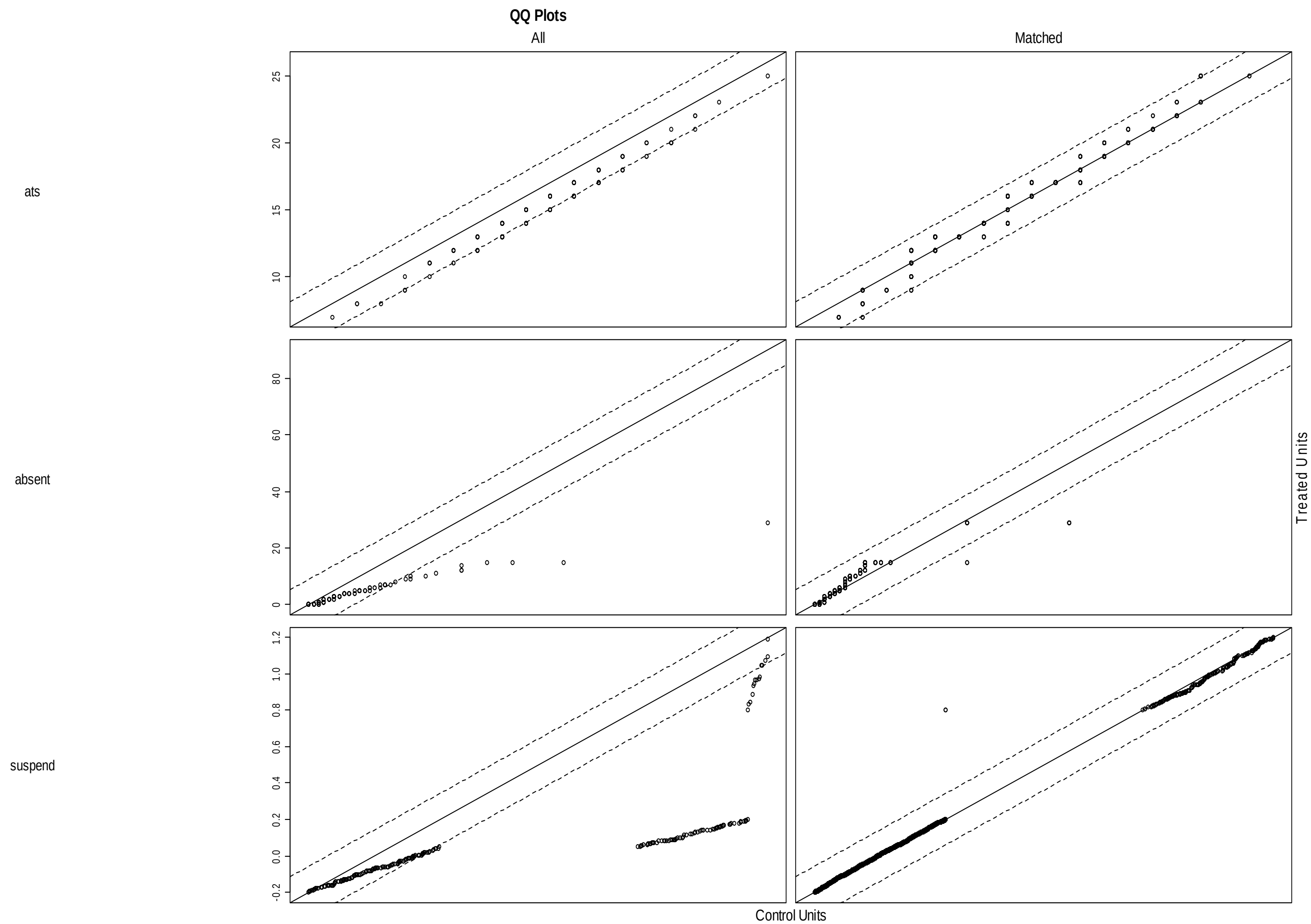

Figure H5. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for attitudes toward school (ats), number of days absent from school (absent), and ever suspended from school (suspend). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

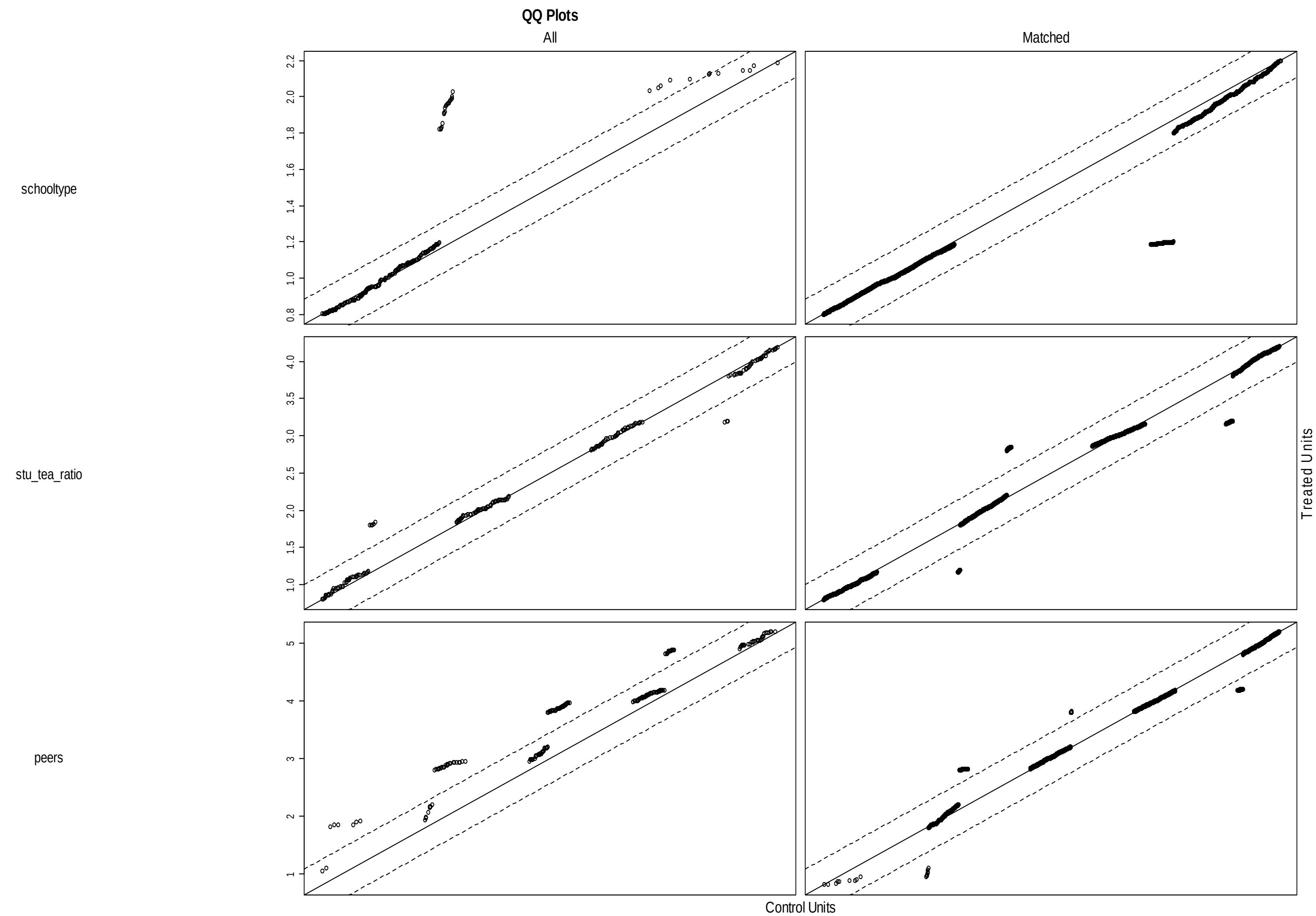

Figure H6. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using 5:1 nearest neighbor matching with replacement and a caliper size of .05. The plots illustrate the pre and post-matching covariate balance for school type (schooltype), student-teacher ratio (stu_tea_ratio), and percent peers college-bound (peers). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

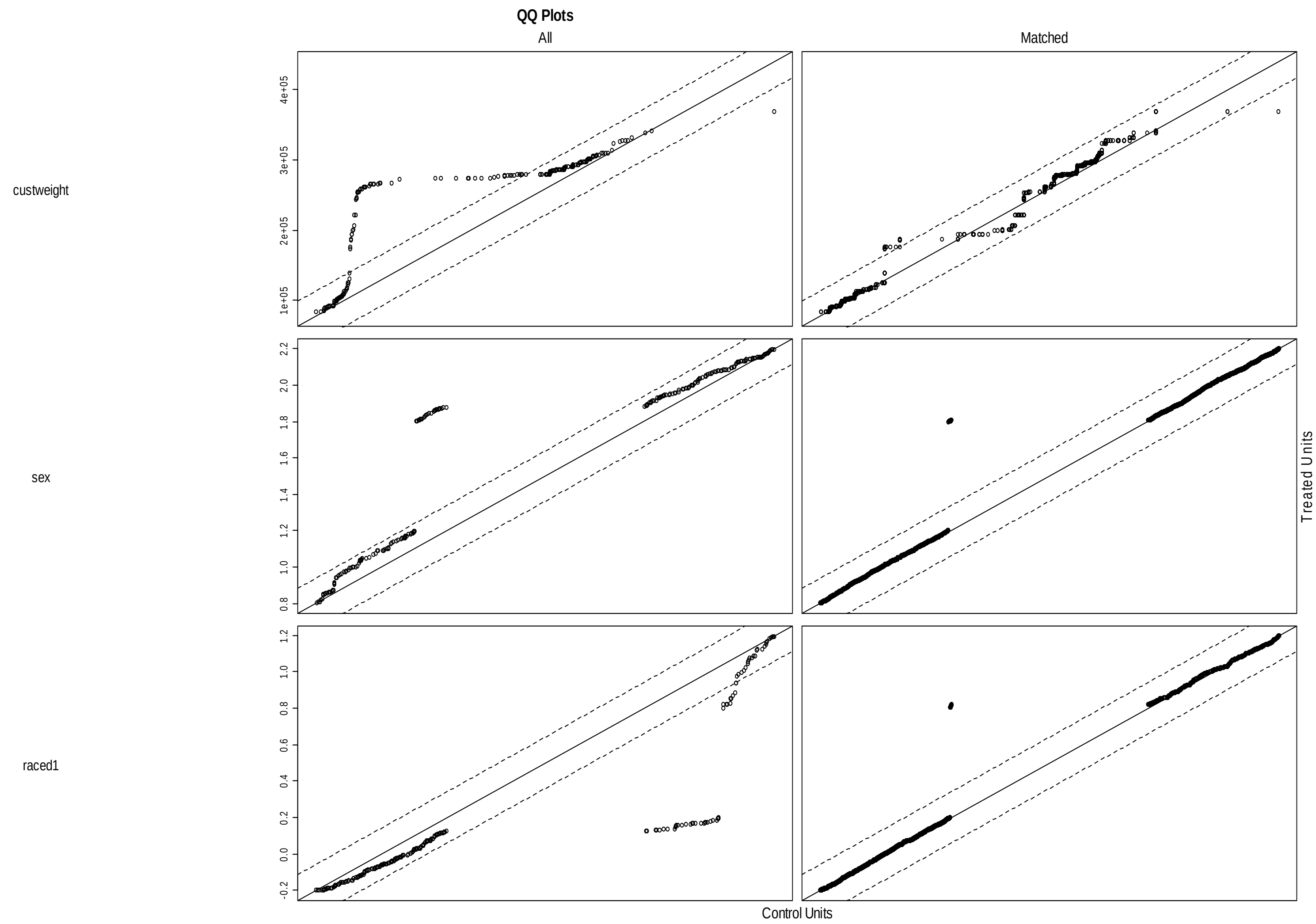

Figure H7. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for survey weight (custweight), gender (sex), and race/ethnicity dummy 1 (raced1). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

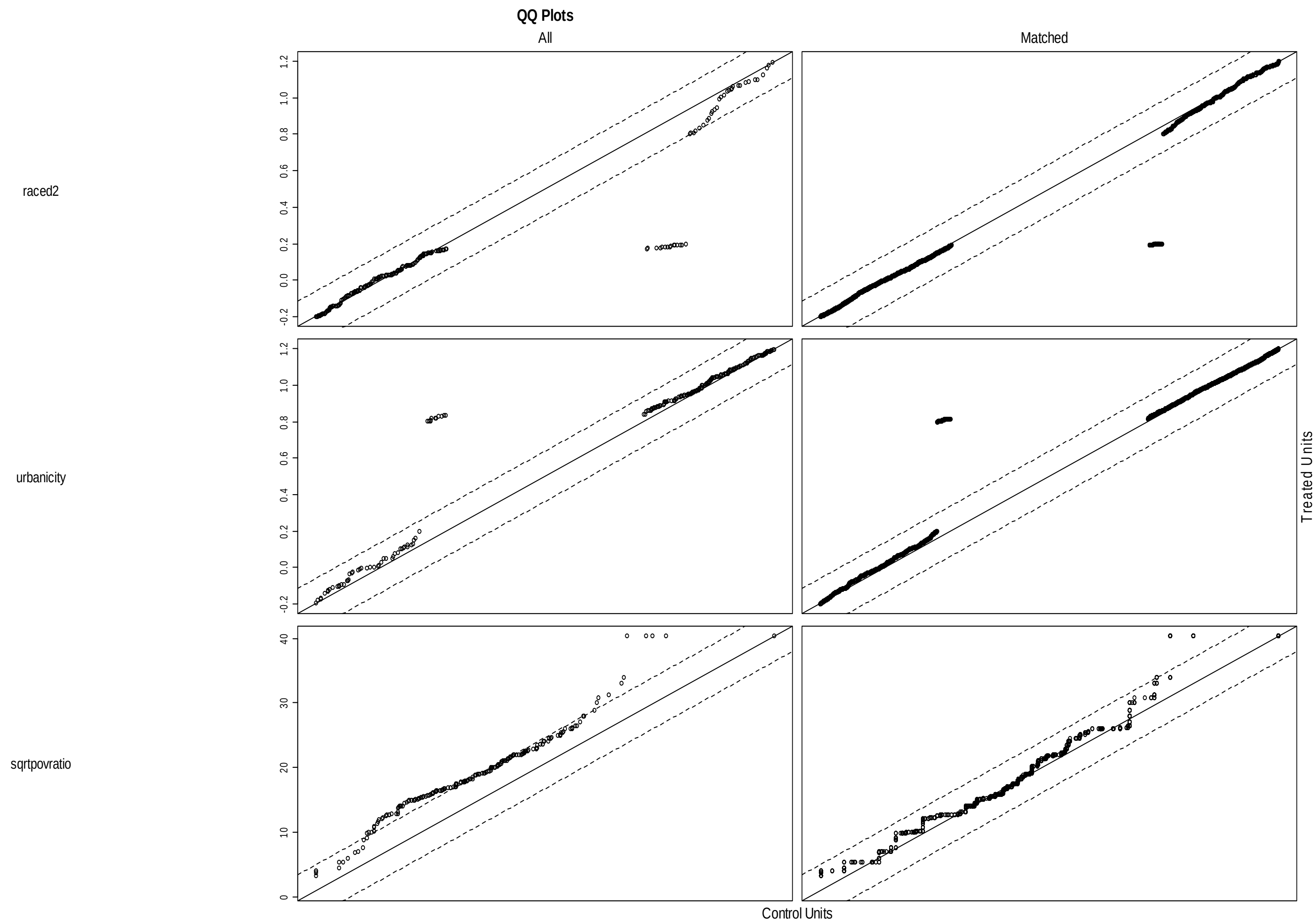

Figure H8. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for race/ethnicity dummy 2 (raced2), urbanicity (urbanicity), and household poverty ratio (sqrtpovratio). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

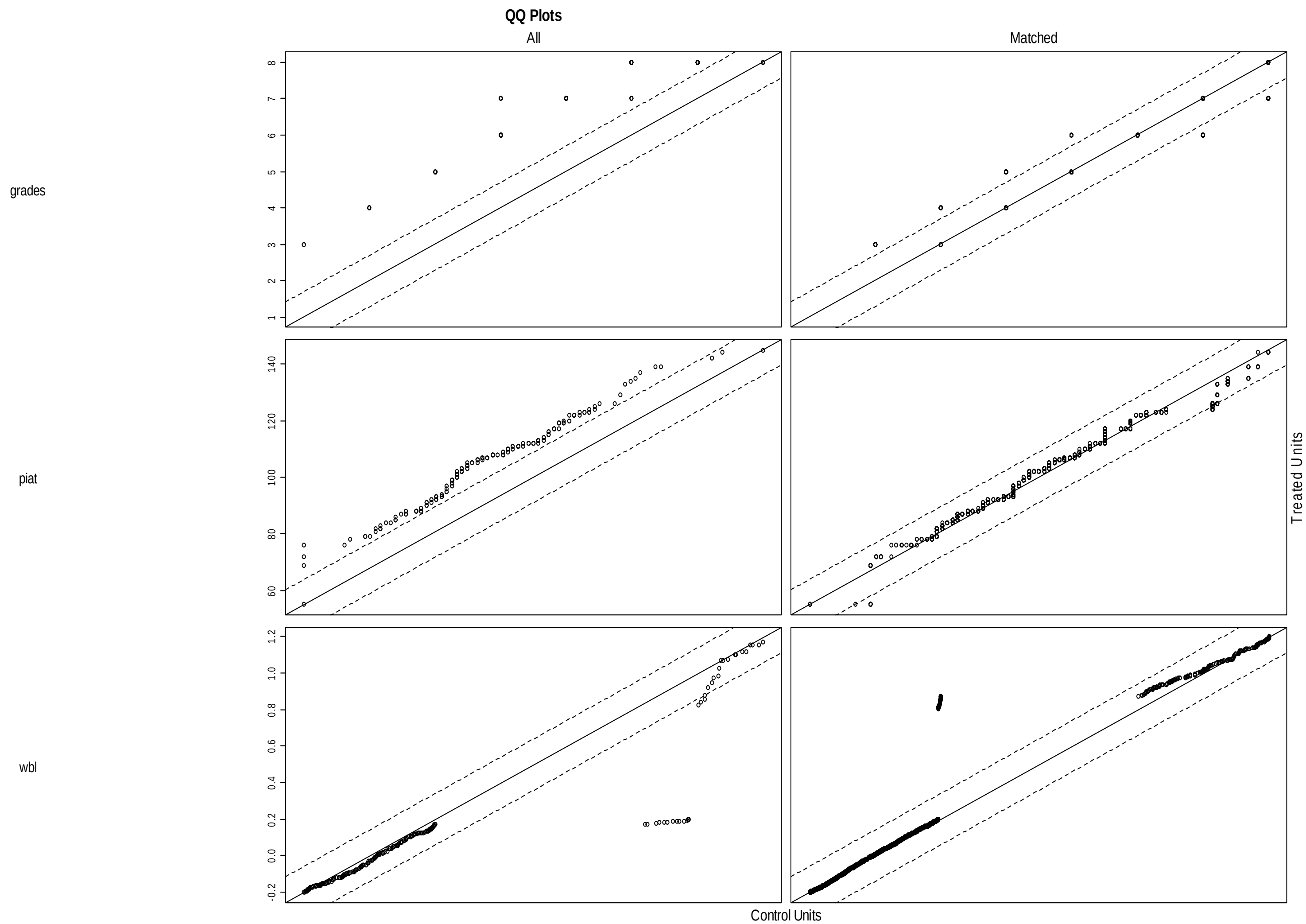

Figure H9. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for grades received in eighth grade (grades), PIAT math standard score (piat), and work-based learning (wbl). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

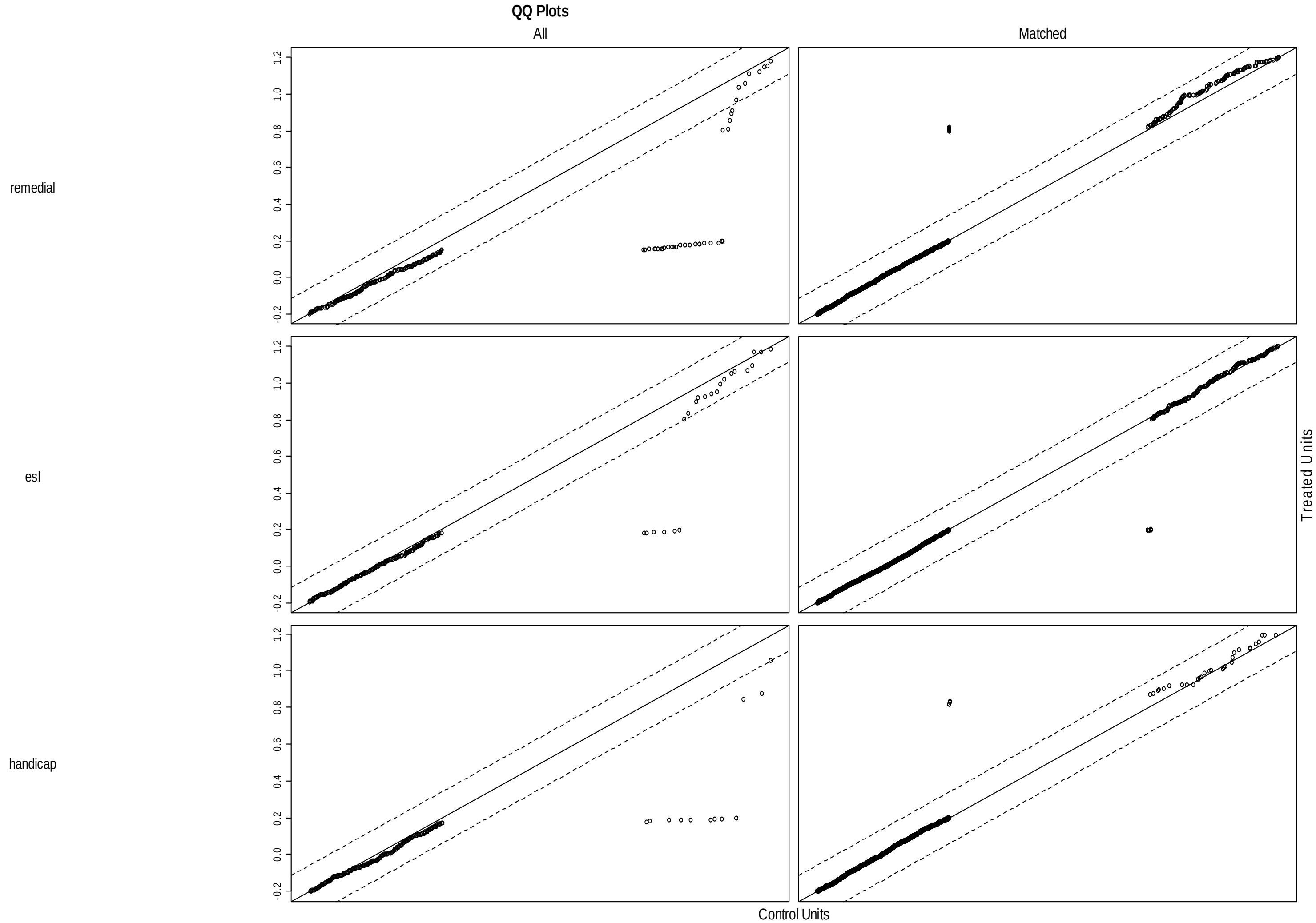

Figure H10. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for remedial English and/or math (remedial), ESL and/or bilingual program (esl), and educational and/or physical handicap (handicap). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

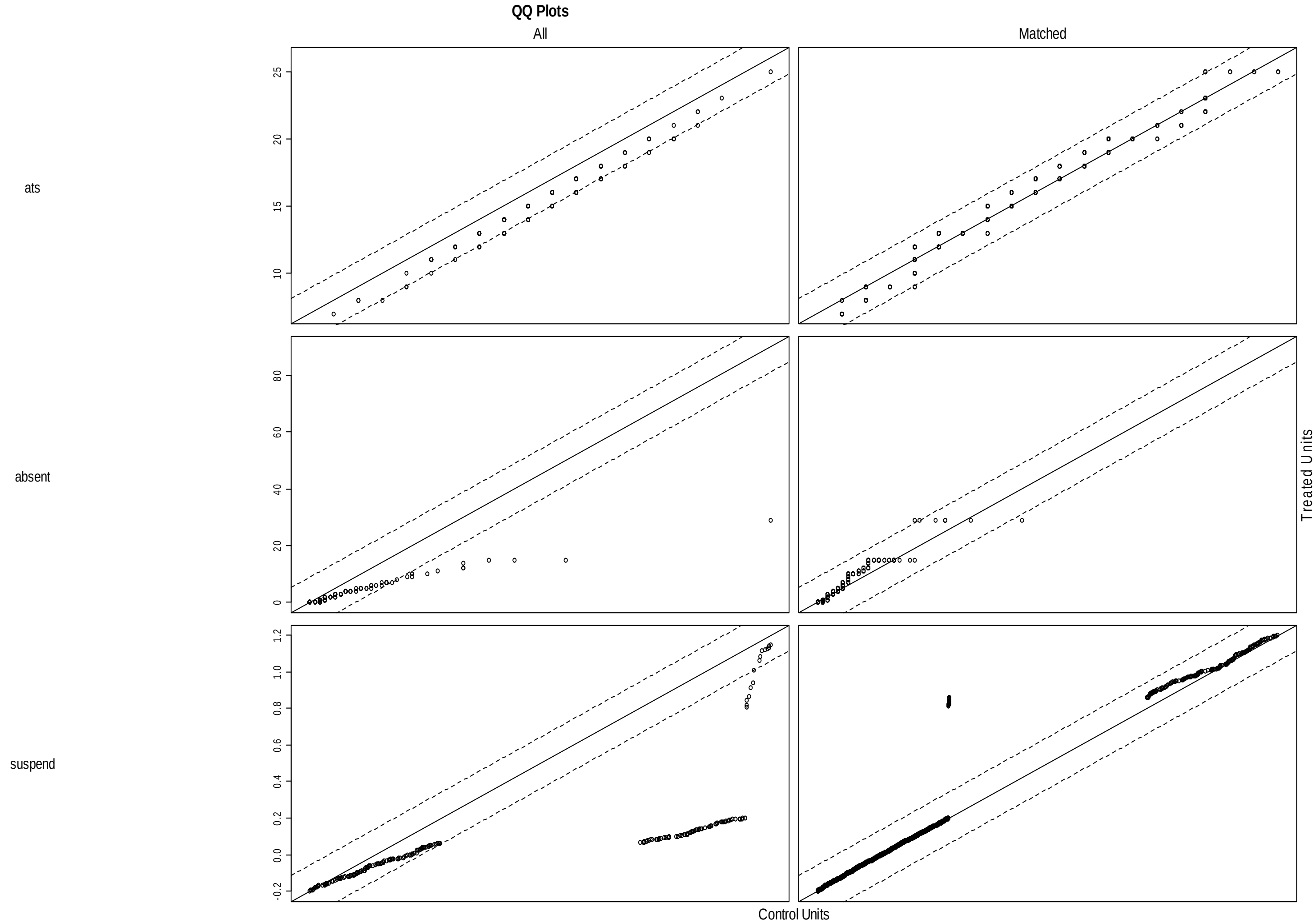

Figure H11. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for attitudes toward school (ats), number of days absent from school (absent), and ever suspended from school (suspend). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

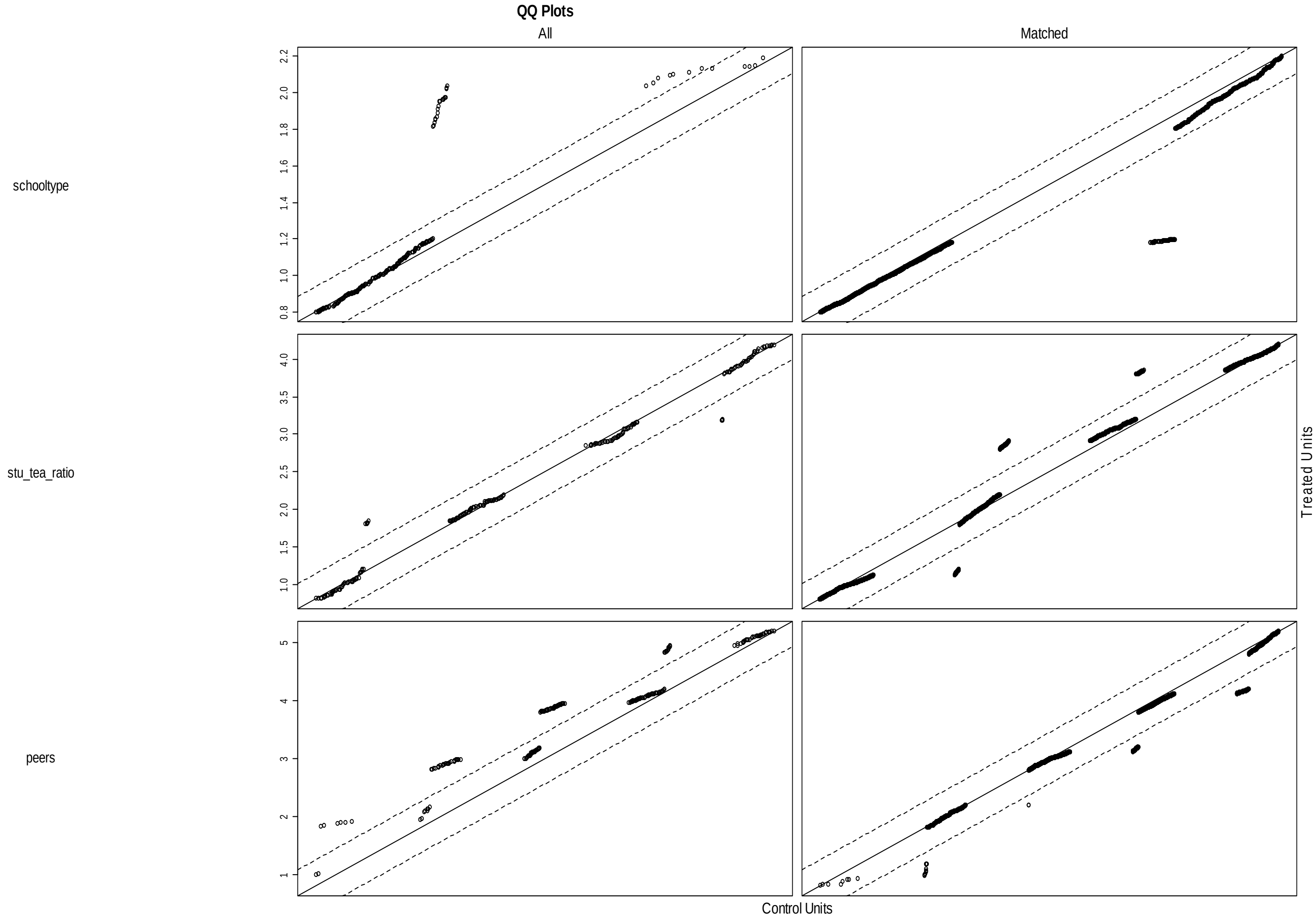

Figure H12. Imputation 1 QQ-plots for the sample of college-preparatory (treatment) and general-track students (control) using full matching. The plots illustrate the pre and post-matching covariate balance for school type (schooltype), student-teacher ratio (stu_tea_ratio), and percent peers college-bound (peers). Increasing proximity of treatment and control units to the 45-degree line is indicative of increasing covariate balance.

APPENDIX I

CHI-SQUARE CONTINGENCY TABLES

FOR

CTE AND GENERAL-TRACK STUDENTS

Table I1

Imputation 1 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	57	65
	Expected frequency	26	39	65
	Adjusted residual	-4.79	4.79	
GED	Observed frequency	5	58	63
	Expected frequency	25	38	63
	Adjusted residual	-5.46	5.46	
Regular HS diploma	Observed frequency	189	206	395
	Expected frequency	158	237	395
	Adjusted residual	5.22	-5.22	
Two-year college degree	Observed frequency	18	18	36
	Expected frequency	14	22	36
	Adjusted residual	1.27	-1.27	
Four-year college degree	Observed frequency	35	45	80
	Expected frequency	32	48	80
	Adjusted residual	.75	-.75	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 255$ [b] $n = 384$

Table I2

Imputation 2 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	50	58
	Expected frequency	23	35	58
	Adjusted residual	-4.28	4.28	
GED	Observed frequency	5	60	65
	Expected frequency	26	39	65
	Adjusted residual	-5.61	5.61	
Regular HS diploma	Observed frequency	191	215	406
	Expected frequency	163	243	406
	Adjusted residual	4.76	-4.76	
Two-year college degree	Observed frequency	18	18	36
	Expected frequency	14	22	36
	Adjusted residual	1.26	-1.26	
Four-year college degree	Observed frequency	35	42	77
	Expected frequency	31	46	77
	Adjusted residual	1.04	-1.04	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 257$ [b] $n = 385$

Table I3

Imputation 3 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	48	56
	Expected frequency	23	33	56
	Adjusted residual	-4.30	4.30	
GED	Observed frequency	5	53	58
	Expected frequency	24	34	58
	Adjusted residual	-5.30	5.30	
Regular HS diploma	Observed frequency	190	202	392
	Expected frequency	162	230	392
	Adjusted residual	4.77	-4.77	
Two-year college degree	Observed frequency	18	15	33
	Expected frequency	14	19	33
	Adjusted residual	1.59	-1.59	
Four-year college degree	Observed frequency	36	48	84
	Expected frequency	35	49	84
	Adjusted residual	.32	-.32	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 257$ [b] $n = 366$

Table I4

Imputation 4 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	47	55
	Expected frequency	23	32	55
	Adjusted residual	-4.20	4.20	
GED	Observed frequency	6	74	80
	Expected frequency	33	47	80
	Adjusted residual	-6.55	6.55	
Regular HS diploma	Observed frequency	191	197	388
	Expected frequency	160	228	388
	Adjusted residual	5.24	-5.24	
Two-year college degree	Observed frequency	18	14	32
	Expected frequency	13	19	32
	Adjusted residual	1.78	-1.78	
Four-year college degree	Observed frequency	35	37	72
	Expected frequency	30	42	72
	Adjusted residual	1.37	-1.37	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 258$ [b] $n = 369$

Table I5

Imputation 5 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	64	72
	Expected frequency	29	43	72
	Adjusted residual	-5.27	5.27	
GED	Observed frequency	5	59	64
	Expected frequency	25	39	64
	Adjusted residual	-5.51	5.51	
Regular HS diploma	Observed frequency	188	196	384
	Expected frequency	153	231	384
	Adjusted residual	5.87	-5.87	
Two-year college degree	Observed frequency	18	20	38
	Expected frequency	15	23	38
	Adjusted residual	.99	-.99	
Four-year college degree	Observed frequency	33	43	76
	Expected frequency	30	46	76
	Adjusted residual	.70	-.70	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 252$ [b] $n = 382$

Table I6

Imputation 1 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	60	68
	Expected frequency	25	43	68
	Adjusted residual	-4.54	4.54	
GED	Observed frequency	6	74	80
	Expected frequency	30	50	80
	Adjusted residual	-5.81	5.81	
Regular HS diploma	Observed frequency	191	240	431
	Expected frequency	159	272	431
	Adjusted residual	5.07	-5.07	
Two-year college degree	Observed frequency	18	17	35
	Expected frequency	13	22	35
	Adjusted residual	1.81	-1.81	
Four-year college degree	Observed frequency	36	50	86
	Expected frequency	32	54	86
	Adjusted residual	1.00	-1.00	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 259$ [b] $n = 441$

Table I7

Imputation 2 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	52	60
	Expected frequency	22	38	60
	Adjusted residual	-3.96	3.96	
GED	Observed frequency	6	61	67
	Expected frequency	25	42	67
	Adjusted residual	-4.99	4.99	
Regular HS diploma	Observed frequency	193	262	455
	Expected frequency	168	287	455
	Adjusted residual	4.04	-4.04	
Two-year college degree	Observed frequency	18	19	37
	Expected frequency	14	23	37
	Adjusted residual	1.51	-1.51	
Four-year college degree	Observed frequency	36	51	87
	Expected frequency	32	55	87
	Adjusted residual	.91	-.91	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 261$ [b] $n = 445$

Table I8

Imputation 3 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	59	67
	Expected frequency	25	42	67
	Adjusted residual	-4.49	4.49	
GED	Observed frequency	6	64	70
	Expected frequency	26	44	70
	Adjusted residual	-5.21	5.21	
Regular HS diploma	Observed frequency	193	250	443
	Expected frequency	164	279	443
	Adjusted residual	4.61	-4.61	
Two-year college degree	Observed frequency	18	16	34
	Expected frequency	13	21	34
	Adjusted residual	1.96	-1.96	
Four-year college degree	Observed frequency	36	53	89
	Expected frequency	33	56	89
	Adjusted residual	.69	-.69	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 261$ [b] $n = 442$

Table I9

Imputation 4 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	56	64
	Expected frequency	24	40	64
	Adjusted residual	-4.27	4.27	
GED	Observed frequency	6	98	104
	Expected frequency	39	65	104
	Adjusted residual	-7.17	7.17	
Regular HS diploma	Observed frequency	192	227	419
	Expected frequency	155	264	419
	Adjusted residual	5.84	-5.84	
Two-year college degree	Observed frequency	18	18	36
	Expected frequency	13	23	36
	Adjusted residual	1.65	-1.65	
Four-year college degree	Observed frequency	36	42	78
	Expected frequency	29	49	78
	Adjusted residual	1.76	-1.76	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 260$ [b] $n = 441$

Table I10

Imputation 5 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	8	65	73
	Expected frequency	27	46	73
	Adjusted residual	-4.85	4.85	
GED	Observed frequency	6	60	66
	Expected frequency	24	42	66
	Adjusted residual	-4.92	4.92	
Regular HS diploma	Observed frequency	192	247	439
	Expected frequency	162	277	439
	Adjusted residual	4.85	-4.85	
Two-year college degree	Observed frequency	18	22	40
	Expected frequency	15	25	40
	Adjusted residual	1.09	-1.09	
Four-year college degree	Observed frequency	35	49	84
	Expected frequency	31	53	84
	Adjusted residual	.97	-.97	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 259$ [b] $n = 443$

APPENDIX J

CHI-SQUARE CONTINGENCY TABLES

FOR

COLLEGE-PREPARATORY AND GENERAL-TRACK STUDENTS

Table J1

Imputation 1 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	8	11
	Expected frequency	5	6	11
	Adjusted residual	-1.31	1.31	
GED	Observed frequency	1	10	11
	Expected frequency	5	6	11
	Adjusted residual	-2.53	2.53	
Regular HS diploma	Observed frequency	74	108	182
	Expected frequency	85	97	182
	Adjusted residual	-2.16	2.16	
Two-year college degree	Observed frequency	12	10	22
	Expected frequency	10	12	22
	Adjusted residual	.76	-.76	
Four-year college degree	Observed frequency	106	88	194
	Expected frequency	91	103	194
	Adjusted residual	3.03	-3.03	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 196$ [b] $n = 224$

Table J2

Imputation 2 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	9	12
	Expected frequency	6	6	12
	Adjusted residual	-1.49	1.49	
GED	Observed frequency	1	8	9
	Expected frequency	4	5	9
	Adjusted residual	-2.13	2.13	
Regular HS diploma	Observed frequency	73	116	189
	Expected frequency	87	102	189
	Adjusted residual	-2.77	2.77	
Two-year college degree	Observed frequency	12	9	21
	Expected frequency	10	11	21
	Adjusted residual	1.04	-1.04	
Four-year college degree	Observed frequency	107	87	194
	Expected frequency	89	105	194
	Adjusted residual	3.42	-3.42	

Note. Frequency counts are rounded to the nearest integer. Two cells (20%) have an expected frequency count of less than 5.
[a] $n = 196$ [b] $n = 229$

Table J3

Imputation 3 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	7	10
	Expected frequency	5	5	10
	Adjusted residual	-1.06	1.06	
GED	Observed frequency	1	10	11
	Expected frequency	5	6	11
	Adjusted residual	-2.53	2.53	
Regular HS diploma	Observed frequency	75	116	191
	Expected frequency	89	102	191
	Adjusted residual	-2.73	2.73	
Two-year college degree	Observed frequency	12	13	25
	Expected frequency	12	13	25
	Adjusted residual	.15	-.15	
Four-year college degree	Observed frequency	106	80	186
	Expected frequency	87	99	186
	Adjusted residual	3.80	-3.80	

Note. Frequency counts are rounded to the nearest integer. One cell (10%) has an expected frequency count of less than 5.
[a] $n = 197$ [b] $n = 226$

Table J4

Imputation 4 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	9	12
	Expected frequency	6	6	12
	Adjusted residual	-1.54	1.54	
GED	Observed frequency	1	7	8
	Expected frequency	4	4	8
	Adjusted residual	-1.97	1.97	
Regular HS diploma	Observed frequency	75	107	182
	Expected frequency	85	97	182
	Adjusted residual	-2.05	2.05	
Two-year college degree	Observed frequency	12	11	23
	Expected frequency	11	12	23
	Adjusted residual	.52	-.52	
Four-year college degree	Observed frequency	108	91	199
	Expected frequency	93	106	199
	Adjusted residual	2.85	-2.85	

Note. Frequency counts are rounded to the nearest integer. Two cells (20%) have an expected frequency count of less than 5.
[a] $n = 199$ [b] $n = 225$

Table J5

Imputation 5 – Contingency Table for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	8	11
	Expected frequency	5	6	11
	Adjusted residual	-1.35	1.35	
GED	Observed frequency	1	9	10
	Expected frequency	5	5	10
	Adjusted residual	-2.40	2.40	
Regular HS diploma	Observed frequency	75	107	182
	Expected frequency	86	96	182
	Adjusted residual	-2.22	2.22	
Two-year college degree	Observed frequency	12	8	20
	Expected frequency	9	11	20
	Adjusted residual	1.16	-1.16	
Four-year college degree	Observed frequency	109	90	199
	Expected frequency	94	105	199
	Adjusted residual	2.87	-2.87	

Note. Frequency counts are rounded to the nearest integer. One cell (10%) has an expected frequency count of less than 5.
[a] $n = 200$ [b] $n = 222$

Table J6

Imputation 1 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	13	16
	Expected frequency	5	11	16
	Adjusted residual	-1.17	1.17	
GED	Observed frequency	1	20	21
	Expected frequency	7	14	21
	Adjusted residual	-2.74	2.74	
Regular HS diploma	Observed frequency	75	196	271
	Expected frequency	87	184	271
	Adjusted residual	-2.15	2.15	
Two-year college degree	Observed frequency	12	18	30
	Expected frequency	10	20	30
	Adjusted residual	.93	-.93	
Four-year college degree	Observed frequency	109	173	282
	Expected frequency	91	191	282
	Adjusted residual	3.11	-3.11	

Note. Frequency counts are rounded to the nearest integer. Zero cells (0%) have an expected frequency count of less than 5.
[a] $n = 200$ [b] $n = 420$

Table J7

Imputation 2 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	12	15
	Expected frequency	5	10	15
	Adjusted residual	-1.06	1.06	
GED	Observed frequency	1	14	15
	Expected frequency	5	10	15
	Adjusted residual	-2.17	2.17	
Regular HS diploma	Observed frequency	75	195	270
	Expected frequency	88	182	270
	Adjusted residual	-2.30	2.30	
Two-year college degree	Observed frequency	12	20	32
	Expected frequency	10	22	32
	Adjusted residual	.60	-.60	
Four-year college degree	Observed frequency	109	171	280
	Expected frequency	91	189	280
	Adjusted residual	3.03	-3.03	

Note. Frequency counts are rounded to the nearest integer. Two cells (20%) have an expected frequency count of less than 5.
[a] $n = 200$ [b] $n = 412$

Table J8

Imputation 3 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	10	13
	Expected frequency	4	9	13
	Adjusted residual	-.74	.74	
GED	Observed frequency	1	13	14
	Expected frequency	5	9	14
	Adjusted residual	-2.06	2.06	
Regular HS diploma	Observed frequency	75	209	284
	Expected frequency	93	191	284
	Adjusted residual	-3.06	3.06	
Two-year college degree	Observed frequency	12	16	28
	Expected frequency	9	19	28
	Adjusted residual	1.18	-1.18	
Four-year college degree	Observed frequency	108	163	271
	Expected frequency	88	183	271
	Adjusted residual	3.41	-3.41	

Note. Frequency counts are rounded to the nearest integer. Two cells (20%) have an expected frequency count of less than 5.
[a] $n = 199$ [b] $n = 411$

Table J9

Imputation 4 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	19	22
	Expected frequency	7	15	22
	Adjusted residual	-1.93	1.93	
GED	Observed frequency	1	14	15
	Expected frequency	5	10	15
	Adjusted residual	-2.16	2.16	
Regular HS diploma	Observed frequency	75	181	256
	Expected frequency	83	173	256
	Adjusted residual	-1.44	1.44	
Two-year college degree	Observed frequency	12	14	26
	Expected frequency	8	18	26
	Adjusted residual	1.52	-1.52	
Four-year college degree	Observed frequency	109	187	296
	Expected frequency	96	200	296
	Adjusted residual	2.19	-2.19	

Note. Frequency counts are rounded to the nearest integer. One cell (10%) has an expected frequency count of less than 5.
[a] $n = 200$ [b] $n = 415$

Table J10

Imputation 5 – Contingency Table for College-preparatory (Treatment) and General-track Students (Control) based on Full Matching

		Curriculum		
		Treatment[a]	*Control*[b]	*Total*
No HS diploma or GED	Observed frequency	3	20	23
	Expected frequency	7	16	23
	Adjusted residual	-2.02	2.02	
GED	Observed frequency	1	14	15
	Expected frequency	5	10	15
	Adjusted residual	-2.16	2.16	
Regular HS diploma	Observed frequency	75	184	259
	Expected frequency	84	175	259
	Adjusted residual	-1.56	1.56	
Two-year college degree	Observed frequency	12	26	38
	Expected frequency	12	26	38
	Adjusted residual	-.11	.11	
Four-year college degree	Observed frequency	109	173	282
	Expected frequency	91	191	282
	Adjusted residual	3.04	-3.04	

Note. Frequency counts are rounded to the nearest integer. One cell (10%) has an expected frequency count of less than 5.
[a] $n = 200$ [b] $n = 417$

APPENDIX K

SENSITIVITY ANALYSIS

FOR

CTE AND GENERAL-TRACK STUDENTS

Table K1

Imputation 1 – Rosenbaum Bounds for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	1.4e-09	1.4e-09	.3	.3	.2	.4
1.1	6.1e-08	2.1e-11	.2	.3	.2	.4
1.2	1.3e-06	2.7e-13	.2	.3	.1	.5
1.3	.000015	3.2e-15	.2	.4	.1	.5
1.4	.000108	0	.2	.4	.1	.5
1.5	.000569	0	.2	.4	.1	.6
1.6	.00224	0	.1	.5	3.5e-07	.6
1.7	.006987	0	.1	.5	3.5e-07	.6
1.8	.017966	0	.1	.5	-3.5e-07	.7
1.9	.03933	0	.1	.5	-3.5e-07	.7
2.0	.075226	0	.1	.6	-3.5e-07	.7
2.1	.128423	0	3.5e-07	.6	-3.5e-07	.8
2.2	.199167	0	3.5e-07	.6	-.1	.8
2.3	.284809	0	3.5e-07	.6	-.1	.8
2.4	.380321	0	-3.5e-07	.7	-.1	.9
2.5	.479435	0	-3.5e-07	.7	-.1	.9

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table K2

Imputation 2 – Rosenbaum Bounds for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	4.7e-09	4.7e-09	.3	.3	.2	.4
1.1	1.8e-07	7.2e-11	.3	.3	.2	.5
1.2	3.5e-06	9.9e-13	.2	.4	.1	.5
1.3	.000038	1.2e-14	.2	.4	.1	.5
1.4	.000261	1.1e-16	.2	.4	.1	.6
1.5	.001273	0	.2	.5	.1	.6
1.6	.004679	0	.1	.5	3.7e-07	.7
1.7	.013645	0	.1	.5	3.7e-07	.7
1.8	.032866	0	.1	.6	-3.7e-07	.7
1.9	.067535	0	.1	.6	-3.7e-07	.8
2.0	.121542	0	.1	.6	-3.7e-07	.8
2.1	.19575	0	3.7e-07	.6	.-1	.8
2.2	.287259	0	3.7e-07	.7	.-1	.9
2.3	.389991	0	3.7e-07	.7	.-1	.9
2.4	.496245	0	-3.7e-07	.7	.-1	.9
2.5	.598509	0	-3.7e-07	.7	.-1	.9

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table K3

Imputation 3 – Rosenbaum Bounds for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	4.0e-10	4.0e-10	.3	.3	.2	.4
1.1	1.9e-08	4.9e-12	.3	.4	.2	.5
1.2	4.5e-07	5.4e-14	.2	.4	.2	.5
1.3	5.8e-06	5.6e-16	.2	.4	.1	.5
1.4	.000048	0	.2	.4	.1	.6
1.5	.000272	0	.2	.5	.1	.6
1.6	.001157	0	.2	.5	.1	.7
1.7	.003878	0	.1	.5	3.9e-07	.7
1.8	.010649	0	.1	.6	3.9e-07	.7
1.9	.024763	0	.1	.6	-3.9e-07	.8
2.0	.050078	0	.1	.6	-3.9e-07	.8
2.1	.089995	0	.1	.6	-3.9e-07	.8
2.2	.146324	0	.1	.7	.-1	.9
2.3	.21851	0	3.9e-07	.7	.-1	.9
2.4	.303552	0	3.9e-07	.7	.-1	.9
2.5	.396605	0	-3.9e-07	.7	.-1	.9

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table K4

Imputation 4 – Rosenbaum Bounds for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	2.0e-08	2.0e-08	.3	.3	.2	.4
1.1	6.5e-07	3.8e-10	.2	.3	.1	.4
1.2	.000011	6.3e-12	.2	.3	.1	.4
1.3	.000099	9.7e-14	.2	.4	.1	.5
1.4	.000606	1.4e-15	.2	.4	.1	.5
1.5	.00266	0	.1	.4	3.5e-07	.6
1.6	.00889	0	.1	.4	3.5e-07	.6
1.7	.023803	0	.1	.5	-3.5e-07	.6
1.8	.0531	0	.1	.5	-3.5e-07	.7
1.9	.101841	0	.1	.5	-3.5e-07	.7
2.0	.172258	0	3.5e-07	.5	-.1	.7
2.1	.262418	0	3.5e-07	.6	-.1	.8
2.2	.366443	0	3.5e-07	.6	-.1	.8
2.3	.4761	0	-3.5e-07	.6	-.1	.8
2.4	.582931	0	-3.5e-07	.6	-.1	.9
2.5	.680047	0	-3.5e-07	.7	-.1	.9

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table K5
Imputation 5 – Rosenbaum Bounds for CTE (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	2.8e-12	2.8e-12	.3	.3	.2	.4
1.1	2.2e-10	2.0e-14	.3	.4	.2	.5
1.2	7.7e-09	1.1e-16	.3	.4	.2	.5
1.3	1.5e-07	0	.2	.4	.2	.6
1.4	1.7e-06	0	.2	.5	.1	.6
1.5	.000013	0	.2	.5	.1	.6
1.6	.000072	0	.2	.5	.1	.7
1.7	.000313	0	.2	.5	.1	.7
1.8	.00109	0	.1	.6	.1	.7
1.9	.003158	0	.1	.6	4.1e-07	.8
2.0	.007843	0	.1	.6	4.1e-07	.8
2.1	.017079	0	.1	.6	-4.1e-07	.8
2.2	.033229	0	.1	.7	-4.1e-07	.9
2.3	.058682	0	.1	.7	-4.1e-07	.9
2.4	.09531	0	.1	.7	-4.1e-07	.9
2.5	.143978	0	4.1e-07	.7	-4.1e-07	.9

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

APPENDIX L

SENSITIVITY ANALYSIS

FOR

COLLEGE-PREPARATORY AND GENERAL-TRACK STUDENTS

Table L1

Imputation 1 – Rosenbaum Bounds for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	1.2e-08	1.2e-08	.5	.5	.4	.7
1.1	2.8e-07	3.5e-10	.4	.6	.3	.7
1.2	3.6e-06	9.2e-12	.4	.6	.3	.8
1.3	.000029	2.3e-13	.4	.6	.2	.8
1.4	.000165	5.6e-15	.4	.7	.2	.8
1.5	.000695	1.1e-16	.3	.7	.1	.8
1.6	.00231	0	.3	.8	.1	.9
1.7	.00634	0	.2	.8	4.3e-07	.9
1.8	.014837	0	.2	.8	-4.3e-07	1
1.9	.030392	0	.2	.8	-4.3e-07	1
2.0	.055654	0	.2	.8	-4.3e-07	1
2.1	.092688	0	.1	.8	-4.3e-07	1
2.2	.142392	0	.1	.9	-.1	1
2.3	.204188	0	4.3e-07	.9	-.1	1.1
2.4	.276057	0	4.3e-07	.9	-.1	1.1
2.5	.354883	0	-4.3e-07	.9	-.2	1.1

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table L2

Imputation 2 – Rosenbaum Bounds for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	2.7e-07	2.7e-07	.5	.5	.3	.6
1.1	4.6e-06	1.1e-08	.4	.6	.2	.7
1.2	.000044	4.2e-10	.4	.6	.2	.8
1.3	.000276	1.5e-11	.3	.6	.1	.8
1.4	.001237	4.9e-13	.3	.6	.1	.8
1.5	.004237	1.6e-14	.2	.7	4.1e-07	.9
1.6	.011699	4.4e-16	.2	.7	-4.1e-07	.9
1.7	.027083	0	.2	.8	-4.1e-07	1
1.8	.054217	0	.1	.8	-4.1e-07	1
1.9	.09619	0	.1	.8	-.1	1
2.0	.154284	0	.1	.8	-.1	1
2.1	.227398	0	4.1e-07	.9	-.1	1
2.2	.312184	0	4.1e-07	.9	-.2	1
2.3	.403768	0	-4.1e-07	.9	-.2	1.1
2.4	.496757	0	-4.1e-07	1	-.2	1.1
2.5	.586179	0	-4.1e-07	1	-.2	1.1

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table L3

Imputation 3 – Rosenbaum Bounds for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	9.9e-09	9.9e-09	.5	.5	.4	.6
1.1	2.4e-07	2.7e-10	.5	.6	.3	.7
1.2	3.1e-06	6.9e-12	.4	.6	.3	.8
1.3	.000026	1.7e-13	.4	.6	.2	.8
1.4	.000149	3.9e-15	.4	.7	.2	.8
1.5	.000636	1.1e-16	.3	.7	.1	.9
1.6	.002143	0	.3	.7	.1	.9
1.7	.005947	0	.2	.8	.1	1
1.8	.014053	0	.2	.8	4.5e-07	1
1.9	.029028	0	.2	.8	4.5e-07	1
2.0	.053542	0	.2	.8	4.5e-07	1.1
2.1	.089725	0	.1	.9	-.1	1.1
2.2	.13858	0	.1	.9	-.1	1.1
2.3	.19964	0	.1	.9	-.2	1.2
2.4	.270975	0	.1	1	-.2	1.2
2.5	.349525	0	4.5e-07	1	-.2	1.2

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table L4

Imputation 4 – Rosenbaum Bounds for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	2.2e-08	2.2e-08	.5	.5	.3	.7
1.1	5.1e-07	6.5e-10	.4	.5	.3	.7
1.2	6.5e-06	1.7e-11	.4	.6	.2	.8
1.3	.000051	4.4e-13	.3	.6	.2	.8
1.4	.000277	1.1e-14	.3	.7	.1	.8
1.5	.001133	2.2e-16	.3	.7	.1	.9
1.6	.003657	0	.2	.7	.1	.9
1.7	.009733	0	.2	.8	4.3e-07	.9
1.8	.022082	0	.2	.8	4.3e-07	1
1.9	.043855	0	.2	.8	-4.3e-07	1
2.0	.07788	0	.1	.8	-4.3e-07	1
2.1	.125839	0	.1	.9	-.1	1
2.2	.187691	0	.1	.9	-.1	1
2.3	.261544	0	.05	.9	-.1	1.1
2.4	.343992	0	4.3e-07	.9	-.2	1.1
2.5	.430758	0	4.3e-07	1	-.2	1.1

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)

Table L5

Imputation 5 – Rosenbaum Bounds for College-preparatory (Treatment) and General-track Students (Control) based on 5:1 Nearest-neighbor Matching

Gamma	*Sig+*	*Sig-*	*t-hat+*	*t-hat-*	*CI+*	*CI-*
1.0	4.7e-08	4.7e-08	.5	.5	.3	.6
1.1	9.9e-07	1.5e-09	.4	.5	.2	.7
1.2	.000012	4.4e-11	.4	.6	.2	.7
1.3	.000085	1.2e-12	.3	.6	.2	.8
1.4	.000438	3.2e-14	.3	.6	.1	.8
1.5	.001701	7.8e-16	.2	.7	.1	.8
1.6	.005249	0	.2	.7	4.1e-07	.8
1.7	.013415	0	.2	.7	4.1e-07	.9
1.8	.029339	0	.2	.8	-4.1e-07	.9
1.9	.056357	0	.2	.8	-4.1e-07	1
2.0	.097086	0	.1	.8	-4.1e-07	1
2.1	.152589	0	.1	.8	-.1	1
2.2	.221921	0	.1	.8	-.1	1
2.3	.302235	0	4.1e-07	.9	-.2	1
2.4	.389343	0	4.1e-07	.9	-.2	1
2.5	.478517	0	-4.1e-07	.9	-.2	1.1

Note. Gamma = log odds of differential assignment due to unobserved variables (i.e., hidden bias)
sig+ = upper bound significance level (assumption: overestimation of treatment effect)
sig- = lower bound significance level (assumption: underestimation of treatment effect)
t-hat+ = upper bound Hodges-Lehman point estimate
t-hat- = lower bound Hodges-Lehman point estimate
CI+ = upper bound confidence interval (α=.95)
CI- = lower bound confidence interval (α=.95)