ACCELERATION AND WELL-BEING AT AGE 50 IN THE TOP 1% IN MATHEMATICAL ABILITY

By

Stijn Smeets

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
Psychology

December, 2013

Nashville, Tennessee

Approved:

Professor David Lubinski
Professor Camilla P. Benbow
Professor James H. Steiger
Professor Andrew J. Tomarken
Professor Tamra Stambaugh

To my sisters Liesbeth & Marijke,
for their unconditional support.

ACKNOWLEDGEMENTS

I am indebted to the faculty, colleagues, family, and friends who have directly or indirectly contributed to this dissertation.

My co-advisors, David Lubinski and Camilla P. Benbow, for the opportunity to be part of the SMPY lab and their continuous support. My committee members: Andrew Tomarken for statistical and career advice, James Steiger, for statistical insights and carefully reviewing the manuscript for typos, and Tamra Stambaugh, for giving me the opportunity to work with highly intelligent children, their parents, and their teachers.

I thank my SMPY colleagues Greg Park, Kim Robertson, Jon Wai, Harrison Kell, and Rachel Morgan for their help in reviewing my manuscripts, introducing me to R, and their mental support during many enjoyable walks over Peabody campus. My colleagues at the Vanderbilt Programs for Talented Youth, for welcoming me in their offices and sharing with me their insights in talent development.

I am grateful for my Nashvillian friends: Maquita, for being a great roommate and a caring sister, who made Nashville feel like home and whose example helped me persist. Jason, Will, Doris, and Svenja, for their friendship and shared destiny throughout my graduate education and beyond. Carlos, Sylvia, and Sebastian, for their warm and welcoming presence. Alysha, Eric, and Greyson, for welcoming me into their house and making me part of their family. Jonathan and Kelsey, for the great talks, beers, and hospitality. I am grateful for Jj & Sabina, Shilpa, Samyuktha and the Indian crew, my artist friends, Wenjing & Mathijs, Jorien & Bram, Christi, Darren, Cori, Michael, Laura, Michelle, Andrey, Neil, Elizabeth, Taylor & Lindsay, Steve, Juan, Angela, Juanita, Luis, Kelly, Adam, Heath, Rob & Amy, Brenda, Matthew, and many others.

My Belgian friends, who did not forget about me while I was abroad: Dokus & Lies, Sarah & Csaba, Isabel & Hans, Anna & Johan, Patricia & Noël, Robbe & Jo, Kristien, Pieter & Han, Niels & Julie, Barbara & Bram, Katrien & Dimi, Sarah, Goele, Stefan, Kurt & Fieke, Lien & Bart, and other friends not named here.

My father and mother for their emotional and financial support. My wonderful sisters, Liesbeth and Marijke, and their partners Nico and Mieke, for hosting me during my frequent visits to Belgium. Their warm welcome, generosity, and unconditional support kept me going. Claire, for the supportive phone calls, and her exceptional emotional and social sensitivity that always helped me make the better decision, and Ulli, who with her lovely distractions opened my eyes to a whole new world.

Last, I gratefully acknowledge the financial support by a Research and Training Grant from the Templeton Foundation and the National Institute of Child Health and Development Grant P30HD15052 to the John F. Kennedy Center at Vanderbilt University, and Fulbright for sponsoring my visa.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Students of the same age are heterogeneous in acquired knowledge, skills, and learning pace (Learned & Wood, 1938; Lynch, 1992; Stanley, 1973; Stanley, 2000; Stanley & Stanley, 1986; VanTassel-Baska, 1983). In an age-in-grade, lockstep educational system that uses the same curriculum at the same pace for all same-aged students, learners with more knowledge, more skills, and who learn at a higher rate will be bored and frustrated, and are at risk for underachievement, stress-related physical and psychological disorders, and multiple other behavioral problems (e.g., Gallagher, Harradine, & Coleman, 1997; Reis & McCoach, 2000; Rimm, 1995).

Academic acceleration offers a better match between the skills and learning pace of advanced learners, and the level, complexity, and pace of the curriculum presented to them. It provides advanced learners with appropriate developmental placement (Benbow & Stanley, 1996; Bleske-Rechek, Lubinski, & Benbow, 2004; Lubinski & Benbow, 2000). Pressey (1949) defined acceleration as "progress through an educational program at rates faster or at ages younger than conventional" (p. 2). Accelerative interventions can be grouped in two categories (Southern & Jones, 1991, 2004): 1) *Grade-based acceleration* shortens the period that a learner remains in the K-12 system and places learners with peers who are older in chronological age. Examples include early admission to Kindergarten, early admission to first grade, grade skipping, and early entrance to middle school, high school, or college. 2) *Subject-based acceleration* exposes a student to content and skills that are more advanced than expected for their age or grade, while the student stays with chronological-age peers for most of the day. Examples include continuous progress, self-paced instruction, advanced placement, taking a college course while in high school, and accelerated summer courses.

Previous studies indicate that academic acceleration of carefully selected, able and motivated students is positively associated with academic achievement and not harmful for their social and emotional adjustment (Colangelo, Assouline, & Gross, 2004; Olszewski-Kubilius, 2007). Several meta-analyses have corroborated these findings (Kent, 1992; Kulik, 2004; Kulik & Kulik, 1984; National Mathematics Advisory Panel, 2008; Rogers, 1991; Steenbergen-Hu & Moon, 2011).

A decade after the publication of "A Nation Deceived: How Schools Hold Back America's Brightest Students " (Colangelo et al., 2004), a compendium of the international academic acceleration research that reached a remarkable consensus, concern about the social and emotional adjustment of intellectually talented youth experiencing educational acceleration continues (Neihart, 2007; Rambo & McCoach, 2012; Wood, Portman, Cigrand, & Colangelo, 2010). In contrast with what research on acceleration suggests, many school administrators, counselors, and teachers still harbor prejudices about harmful socio-emotional consequences of acceleration (Rambo et al., 2012; Wood et al., 2010).

These prejudices are similar to those reported in studies prior to the publication of "A Nation Deceived" (Jackson, Famiglietti & Robinson, 1981; Southern, Jones, & Fiscus, 1989; Vialle, Ashton, Carlton, & Rankin, 2001) and can be summarized in five categories:

1. Decrease in academic self-concept due to social comparison with older peers may interfere with positive self-efficacy and academic achievement.
2. Special educational opportunities put children at risk for social isolation.
3. Accelerating students will lead them to stress, burnout, and psychological disorders.
4. Accelerants lack the social and emotional maturity to deal with classroom expectations and relationships with older peers.
5. Grouping gifted students will not prepare them for a life in a heterogeneous society.

The next chapter reviews empirical evidence that refutes these prejudices.

CHAPTER II

PREVIOUS RESEARCH ABOUT ACCELERATION AND SOCIAL-EMOTIONAL ADJUSTMENT

This chapter first reviews empirical findings that refute five prejudices about harmful effects of acceleration on social and emotional adjustment. It then summarizes two meta-analyses and previous findings from the Study of Mathematically Precocious Youth (SMPY) about this topic. Last, it critically evaluates the research to date and indicates what the current research aims to add to the existing research base.

None of the reviewed studies clearly define social-emotional adjustment. This review utilizes the following definitions. Social adjustment is “the degree to which children get along with their peers; the degree to which they engage in adaptive, competent social behavior; and the extent to which they inhibit aversive, incompetent behavior.” (Crick & Dodge, 1994, p. 82). Commonly used indicators of social adjustment include peer acceptance, participation in extra-curricular activities, and satisfaction with social life. An emotionally well-adjusted person is able to successfully adapt to internal and external stressors. This involves the ability to "manage their own emotions, think constructively, regulate and directing their behavior, control their autonomic arousal, and act on the social and nonsocial environment to alter or decrease sources of stress" (Compas, Connor-Smith, Saltzman, Thomsen, & Wadsworth, 2001, p. 87). Commonly used indicators of emotional adjustment include positive self-concept, absence of psychological disorders or need for counseling, motivation, liking for school, and educational aspirations.

Refuting five prejudices

Prejudice 1: Decrease in the academic self-concept may harm academic achievement.

A student's perceived academic ability is positively associated with his/her academic abilities and negatively associated with the average school and classroom ability, also known as the *big-fish-little-pond effect* (BFLPE; Marsh, 1991; Marsh & Hau, 2003; Seaton, Marsh, & Craven, 2009). Because academic self-concept (ASC) is related to academic aspirations, effort and achievement, even after controlling for individual ability (Guay, Larose, Boivin, 2004; Marsh, 1987; Richardson, Abraham, & Bond, 2012), this negative correlation has been used to argue against ability grouping and academic acceleration (Marsh, 1991; Marsh, Hau, & Craven, 2004; Zeidner & Schleyer, 1999). Both interventions place a student in a higher average ability environment, which according to the BFLPE will lead to a decrease in ASC. This generalization, however, is not consistently supported by the evidence. Although some studies found a decrease in ASC after acceleration (Gibbons, Benbow, & Gerrard, 1994; Lupowski, Withmore, & Ramsay, 1992; Richardson et al., 1990), other studies failed to find such a decrease (Ma, 2002; Robinson & Janos, 1986; Sayler & Bookshire, 1993) or found this decrease to be only temporary (Gibbons et al., 1994; Makel et al., 2012). Temporary ability grouping (such as a three-week summer course; Makel, Lee, Olszewski-Kubilius, & Putallaz, 2012) may not affect the ASC as much as permanent ability grouping (Marsh, 1991; Marsh et al., 2003). In addition, the BFLPEs are smaller for students that have an ASC at least two standard deviations above the mean, which is the case for many academically gifted students (Marsh & Rowe, 1996; Trautwein, Luedtke, Marsh, & Nagy, 2009). Even if the ASC of accelerated gifted students decreases, it remains well above the average ASC of the general student population (e.g., Gibbons et al., 1994). Gifted education experts claim that this decrease makes the ASC more realistic and should not be considered a harmful effect (Dai, 2004; Plucker et al., 2004). The consistent positive association between academic acceleration and academic achievement in the short (Kulik, 2004; Kulik & Kulik, 1984; National Mathematics Advisory Panel, 2008; Rogers, 1991; Steenbergen-Hu & Moon, 2011) and with academic, occupational, and creative accomplishments in the long term (Park, Lubinski, & Benbow, 2013; Wai, Lubinski, Benbow, Steiger, 2010) supports this claim.

Prejudice 2 : Special educational opportunities may put children at risk for social isolation.

Previous studies find no statistically significant differences in participation in social activities and peer acceptance between accelerants and non-accelerants. Klausmeister and Ripple (1962) compared 52 second graders in the upper half in chronological age of their grade with above average ability and achievement who skipped third grade with comparable non-skippers. Seven months after enrollment, there were no differences between accelerants and non-accelerants in teacher ratings of social adjustment or peer acceptance as measured with a sociometric instrument that asked students to rank order their five best friends.

The Fund for the Advancement of Education (1957) evaluated the social adjustment of 1,350 freshmen who entered 11 US universities and colleges in years 1951 through 1954, two years younger than usual and most of them without finishing high school. Early entrants had participated in extra-class activities at least as extensively as their class mates. Faculty ratings of overall adjustment (including poise and self-confidence in social situations, leadership ability, participation in group activities, gregariousness, personal appearance, degree of dependence on family, adjustment to the opposite sex, ease in conversation) show - both at the end of the freshman year as well as at the end of the senior year - that early entrants were slightly less adjusted than the comparison group, and their adjustment scores were spread out more. But overall, adjustment of more than 90% of the early entrants was considered either *moderately good*, *good*, or *excellent*. Student testimonies reveal that a majority of the early entrants (81% of the 1951 cohort and 63% of the 1952 cohort) experienced difficulties in adjusting to academic or social aspects of college life. Twenty five percent reported to feel considered as members of an "out group" by regular students. This, however, could be due to university policies that kept them separated from other students. After these policies were changed, isolation problems decreased. More often than the comparison group, they reported feeling ‘bashful’, ‘shy’, or ‘immature’. One third reported difficulties making friends, and 25% reported problems dating. However, at the end of their senior year, a large majority (73% for the 1951 cohort, 80% for the 1952 cohort) reported to have overcome those difficulties.

Ten years later (in 1966) these accelerants were followed up by Pressey (1967). The majority reported satisfaction with the amount they were accelerated and were well-adjusted socially and emotionally. Minor adjustment problems reported previously (e.g., dating for males) were resolved at the time of follow-up, as evidenced in the higher proportion of accelerants being married and marrying at an earlier age. Only a few reported lasting harm to their social-emotional adjustment.

Terman and Oden (1947) concluded that the influence of acceleration on social maladjustment has been exaggerated. Participants were grouped based on age at high school graduation: below the age of 15 years 6 months (group I; men: n = 36 or 4.6%, women: n = 26 or 4.3%), between 15 years 6 months and 16 years 6 months (group II; men: n = 181 or 23%, women: n = 151 or 24.9%) and at or above the age of 16 years 6 months (group III; men: n = 568 or 72.5%, women: n = 430 or 70.8%). Field workers rated the accelerated men six years after initial selection (1928) as slightly less well-adjusted. The difference for women was not statistically significant. The parent ratings (1928) were slightly lower for the accelerants, but the difference was not statistically significant. Reports at age 35 show that accelerants preferred older companions in high school. Only males from group I engaged less in extra-curricular activities during high school and stressed the disadvantages of acceleration more often. The opposite was true for females from group I. However, findings were neither controlled for intelligence nor for socio-economical status (SES). Because IQ was positively related to the amount of acceleration, not controlling for intelligence confounded and may have biased these findings.

Swiatek and Benbow (1991a) studied participants from SMPY's cohort 1 (top 1% in cognitive ability) and cohort 2 (top .5% in cognitive ability) who were selected on their SAT-M, SAT-V, or Test of Standard Written English (TSWE) scores. They compared 107 participants who enrolled in college at least one year early with an equal number of participants who entered college at the typical age and were matched with the accelerants with regard to sex and ability. At age 23, no statistically significant differences were found between the accelerants and non-accelerants with regard to participation in extra-curricular activities.

To conclude, previous studies fail to provide evidence that accelerants are more at risk for social isolation.

Prejudice 3: Accelerating students will lead to stress, burnout, and psychological disorders.

In the study of the Fund for the Advancement of Education (1957), a team of psychiatrists observed no more psychological difficulties in the early entrants group compared to a comparison group of freshmen of normative age. They did not visit counseling or medical services more often than the comparison group. Also Terman and Oden (1947) found no harmful effects of acceleration on physical and mental health.

Cornell, Callahan, and Loyd (1991a) studied 44 female liberal arts college students enrolled in an early entrance/acceleration program that allowed them to finish high school and college in 5 years. Overall adjustment as measured with the Jackson Personality Inventory (JPI; Jackson, 1976) did not differ from the

4000 college students enrolled in 43 American and Canadian colleges or universities used to standardize the measure. However, 25 (57%) exhibited depressive behavior (as perceived by residence hall staff) that lasted at least 2 weeks and required staff intervention, 5 (11%) engaged in some form of suicidal behavior (from verbal threat to suicide attempt), 22 (50%) were seen by the (nonacademic) program guidance counselor, 11 (25%) were seen by a local mental health professional (psychologist or psychiatrist), and 13 (30%) dropped out of the program for reasons related to socio-emotional stress.

These results must be evaluated with care. There was no adequate control group and a very high dropout rate (30% for socio-emotional stress related issues, and even more for other reasons). Stanley (1991) argues that the criteria for acceptance to the program were not stringent enough (e.g., WISC-R IQ's ranged 115 to 155, with a mean of 129). Successful early college entrance requires a cognitive ability at least above the average ability of students of the college that one wants to enter (Stanley, 1985). Cornell, Calahan, and Loyd (1991b) react by stating that there is no empirical evidence for the latter claim. In addition, contrary to what would be expected based on Stanley's criticism, intelligence and adjustment problems were unrelated in their sample. However, the absence of a relation between intelligence and adjustment problems may be due to the limited sample size (N = 44).

Jin and Moon (2006) compared second-year Korean students of a residential science high school (N = 111; program characteristics were low student-teacher ratio, emphasis on laboratory and inquiry methods, independent research, highly qualified students can graduate in only 2 years) with second-year students in a regular high school matched on GPA for the last 2 years of junior high school and award experiences in academic competitions. They found no significant differences in psychological well-being (autonomy, environmental mastery, personal growth, positive relations with others, purpose of life, self-acceptance; Ryff & Keyes, 1995) but students from the special science school reported to be more satisfied with their school.

Robinson and Janos (1986) studied 24 students who enrolled in the early entrance program (EEP) at the University of Washington. To enroll in the EEP, students must be 14 years or younger, have a score in the 50th percentile in both verbal and quantitative areas of the Washington Pre-College Test (WPCT; Noeth, 1978), and score in the 85th percentile on either the verbal or quantitative subtests. In addition, previous academic achievement, letters of reference from teachers, and interviews are considered in the selection process. They were compared with three other groups: 27 students who qualified for the EEP program but remained in high school, 24 regular freshmen at UW matched with the EEP students on WPCT scores, sex, year of college entry, and high school recruitment area, and last, 23 University of Washington National Merit Scholarship Scholars, matched with the EEP students on sex and year of college enrollment. After one full-time probationary quarter, they were administered the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940), the California Personality Inventory (CPI; Gough, 1969), and the Tennessee Self-Concept Scale (TSC; Fitts, 1965). Controlling for the number of statistical tests, the EEP students did not differ significantly from the other three groups on these scales. Compared to data on gifted adolescents in the literature, female EEP students scored lower on responsibility and socialization, but higher on achievement via independence and flexibility. Male EEP students scored higher on flexibility and femininity. According to Robinson & Janos (1986), the latter may tap more into intellectual interests rather than conventional sex roles. These results have been replicated by Janos, Robinson, and Lunnenborg (1989) and Noble, Robinson, and Gunderson (1993) with a similar study design, and show that acceleration is not harmful for healthy psychological development.

Whereas stress and burnout are associated with disengagement, if significant differences in engagement were found between accelerants and equally able non-accelerants (outcomes including liking for school and educational aspirations), they favored the accelerants (Kulik, 2004; Pollins, 1983; Swiatek & Benbow, 1991a, 1991b).

To conclude, most of the previous studies fail to find lasting harmful psychological effects for acceleration among carefully selected and highly able students.

Prejudice 4: Accelerants are confronted with emotional issues that they are too young to handle.

The absence of harmful effects of acceleration on social adjustment (see prejudice 2) and emotional adjustment (see prejudice 3) calls into question this claim. Children identified as gifted tend to be more mature socially and emotionally (Noble & Robinson, 1991; Silverman, 1997). They outperform age peers in social cognition (Gallagher, 1985; Miller, 1956) and moral judgment (Terman, 1926; Thorndike, 1940), have more mature concepts of friendship (Gross, 1989; 1993a, 1993b), prefer more abstract and complex play (Hollingworth, 1931; Terman, 1926), have more mature interests (including visiting museums, making puzzles, listening to foreign language records, and playing bridge or chess; Hollingworth, 1942; Kincaid, 1969; Terman, 1926; Thorndike, 1940), tend to prefer older playmates (Painter, 1976; Terman, 1925; Terman & Oden, 1947), are more mature in personality (Lessinger & Martinson, 1961; Robinson & Janos, 1986), start grappling earlier with good and evil and are younger when first attempting to construct a philosophy of life (Hollingworth, 1942). Contrary to this claim, placing students identified as gifted with chronological age peers may create more discrepancy in social and emotional maturity than placing them with older (mental) peers.

Prejudice 5: Grouping gifted students will not prepare them for a life in a heterogeneous society.

Grouping highly able students creates an intellectual peer group and facilitates the formation of deep and satisfying friendships that are based on shared interest and mutual understanding (Gross, 2003; Isaacs & Duffus, 1995; Janos et al., 1988; Janos at al., 1989). It is only in mutually satisfying and safe relationships that children can develop the social and emotional skills necessary to function in a heterogeneous and diverse society (Gross, 2003).

Meta-analytic studies

There are five meta-analyses investigating the effects of acceleration on social-emotional adjustment in students identified as gifted: Kulik and Kulik (1984), Rogers (1991), Kent (1992), Kulik (2004), and Steenbergen-Hu and Moon (2011). Because Kulik (2004) reviewed all studies that were also included in the previous meta-analyses, only the results of Kulik (2004), and Steenbergen-Hu and Moon (2011) will be reviewed.

Kulik (2004) found 13 studies, using as outcomes educational aspirations (6 studies, Cohen's *d*s range from –.05 to .77), liking for school/subject (6 studies, *d*s range from –.24 to .77), participation in activities (3 studies, *d*s range from –.22 to –.10), and self-acceptance/personal adjustment (5 studies, *d*s range from –.41 to –.02). Kulik concludes that accelerants compared to equally able non-accelerants have higher educational aspirations, no decrease in social participation, and have a slightly lower assessment of self. The effects on liking of school/subjects were too heterogeneous to interpret. No combined effect size was calculated because the number of studies was limited, and study outcomes were too heterogeneous.

As noted previously, the big-fish-little-pond effect explains this negative effect on the academic self-concept (Marsh & Hau, 2003). Experts in the field agree that this decrease leads to a more realistic self-concept and is therefore not psychologically maladaptive (e.g., Neihart, 2007).

Steenbergen-Hu and Moon (2011) analyzed 22 studies that compared the social and emotional adjustment of accelerants with non-accelerated peers that were either older, of the same age, or a combination of both age groups. Outcome measures of social and emotional adjustment included social maturity scores, teacher ratings of social skills, participation in extra-curricular activities, self-concept, self-reported mental health and social adjustment. Acceleration included both grade-based acceleration and subject-based acceleration. Nine studies compared accelerants with same-age non-accelerants and found a combined effect size of .141, 95% CI [–.013, .295]. Ten studies compared accelerants with older non-accelerated peers, and found a combined effect size of .052, 95% CI [–.111, .215]. They found no evidence for harmful effects of acceleration on social or emotional adjustment. These results must be interpreted with care, however. The outcome measures included in this meta-analysis are heterogeneous. Combining them in a single analysis can be misleading.

Previous results from the Study of Mathematically Precocious Youth

Fox (1983) followed 26 mathematically gifted (SAT-M > 370) end-of-the-year seventh-grade girls who took an Algebra I class that was taught twice a week, for two hours, during 3 summer months (only 18 girls completed the class). At age 18, their educational aspirations did not differ from a female and male control group matched on SAT-M, SAT-V, occupation of their father and their parental education.

Using data from SMPY's cohort 1, Pollins (1983) compared 23 radically accelerated participants (i.e., at some point at least three years ahead of their age-mates in educational placement) with equally able students of the same age. At age 13, there were no statistically significant differences in California

Psychological Inventory (CPI) profile between accelerants, non-accelerants, and an 8th grade gifted group. They did not differ significantly in their liking for school and math. At age 18, the radical accelerants participated in more types of out of school activities, but in a smaller number of activities, held fewer jobs (the author hypothesizes because they were too young to work in high school), participated somewhat less in college activities, and reported a slightly less liking for college. They had higher academic aspirations (PhD vs. Master's), and reported having used their educational opportunities better. Both groups reported positive effects of acceleration on social and emotional development.

Brody and Benbow (1987) compared 4 groups consisting of SMPY’s cohort 2 participants. group 1 (n = 143) skipped one or more grades, group 2 (n = 277) took AP courses or college courses on a part-time basis while in high school, group 3 (n = 50) participated in subject matter acceleration, special classes or tutoring, and Group 4 (n = 40) included students with no accelerative experiences. At age 18, group 2 had participated in more extracurricular activities than group 1, and had held more leadership or officer positions in those clubs. Groups 3 and 4 often scored in between groups 1 and 2. Students who were accelerated were attending top 50 colleges more often. Groups 1 and 2 aspired more often to obtain a doctorate (law, medicine, PhD) than groups 3 and 4. There were no significant differences between the groups in measures of locus of control, self-esteem, personality traits (measured with the Cattell 16 Personality Factors Questionnaire; Cattell & Butcher, 1968), and scales of the Adjective Check List (ACL; Gough & Heilbrun, 1980) related to social-emotional adjustment, including self-control, self-confidence, and personal adjustment.

Swiatek and Benbow (1991a) studied participants from SMPY's cohort 1 (top 1% in cognitive ability) and cohort 2 (top .5% in cognitive ability) selected on their SAT-M, SAT-V, or TSWE scores. They compared 107 participants who enrolled in college at least one year early with an equal number of participants who entered college at the typical age and were matched with accelerants on sex and ability. At age 23, no differences were found between the accelerants and non-accelerants with regard to liking for college, participation in extra-curricular activities, attitudes towards mathematics and science, locus of control and self-esteem. Both groups scored favorably on all outcomes.

Swiatek and Benbow (1991b) evaluated the effects of two SMPY fast paced mathematics classes in 1972 (N = 16, participants ranged from completion of 6th grade through accelerated completion of 10th grade) and 1973 (N = 28, 7th graders through accelerated 9th graders). Classes gathered on Saturday mornings over the course of approximately 14 months, covering algebra I & II, plane geometry, trigonometry, and analytical geometry. Students unable to keep pace could transfer to a self-paced class that covered algebra I and II in 12 months. Only participants who completed the after college follow-up questionnaire were included in the analyses. The group of accelerants included participants who completed the fast-paced class and those who completed the self-paced class (N = 37). The comparison group consisted of students who dropped out of the fast-paced class before completing it or students who qualified but never enrolled (N=58). The comparison group scored slightly higher on self-esteem than did the accelerants, but the average self-esteem score of both groups was positive. There were no statistically significant differences in attitudes towards mathematics and science.

Lubinski, Webb, Morelock, & Benbow (2001) asked SMPY participants who scored in the top 1 in 10,000 in mathematical or verbal reasoning ability how they felt about their accelerative experience at age 23. More than 70% expressed satisfaction with what they did. More than 10% even wished they had accelerated more. Less than 5% wished they had not accelerated, and approximately 5% wished they had not accelerated as much. When asked about the impact of acceleration on development areas including acceptance of abilities, acceptance of self, interest in math-science-humanities-social science, emotional stability, and social life, their answers ranged from *no effect* to *a favorable effect*. Using data from SMPY Cohorts 1 and 2, Benbow, Lubinski, Shea, and Eftekhari-Sanjani (2000) found that participants at age 35 report the effect of acceleration on their social life to be neutral on average. The far majority did not support eliminating homogeneous ability grouping.

Summary, limitations of previous studies, and aim of the current research

Previous studies fail to find harmful effects of acceleration on outcomes including educational aspirations, social participation, marital life, personality, locus of control, psychological well-being, and popularity. However, an overall conclusion that acceleration of carefully selected students is not harmful for their socio-emotional development may not be justified yet. Support for such a statement is most scientifically compelling when studies have following characteristics (this closely follows Cornell, Callahan, Bassin, & Ramsay, 1991):

1. Social-emotional development must be broadly defined. It must encompass intrapersonal (anxiety, depression, self-concept) as well as interpersonal adjustment constructs (family and peer relationships).
2. Outcomes should measure the specific challenges of acceleration. According to Cornell et al. (1991) these are stress related program attrition, need for mental health or counseling, incidence of depression, loneliness, suicidality, or other signs of emotional distress, and a level of social activity and friendship that is developmentally appropriate for the accelerant.
3. Multiple methods (self-report, observation by parents, teachers, peers) must be used to measure social-emotional adjustment. Relying too much on self-report measures may induce bias. Many studies used only self-report (e.g., Benbow, 1983; Brody & Benbow, 1987; Gagné & Gagnier, 2004; Pollins, 1983; Robinson & Janos, 1986).
4. Measures should have proven reliability and validity. Information about reliability and validity lack in, for example, Brody and Benbow (1987), Fund for the Advancement of Education (1957), and Pollins (1983).
5. Accelerants should be compared with equally able students who were not accelerated. Both groups should be equivalent on variables that are related to both treatment effect and treatment

assignment (including intelligence, motivation, and how much parents value education). Studies with no comparison group, or limited equivalence between the groups included Benbow (1983), Cornell et al., (1991), Gagné and Gagnier (2004), Jin and Moon (2006), Ma (2002), Sayler and Bookshire (1993), Terman and Oden (1947).

6. Measurement must not contain excessive error and sample sizes must be large to minimize Type I and Type II errors. This was not the case in Engle (1938), Janos (1987), Janos et al. (1989), Klausmeister (1963), Klausmeister and Ripple (1962), Pollins (1983), and Robinson and Janos (1986).
7. Dropout must be limited, because students who were negatively affected by acceleration may be more likely to dropout of both the program and the study. Dropout may have biased results of Cornell et al. (1991), Engle (1938), Fox, Benbow, and Perkins, (1983), Lupowski, Whitmore, and Ramsay (1992), Pollins (1983), and Swiatek and Benbow (1991b).

Some studies did not carefully select the students to accelerate (e.g., Engle, 1938); some used inappropriate ability assessments (e.g., Cornell et al., 1991a; see also Stanley, 1991). Additionally, the long term effects of acceleration have not been fully investigated. Small negative effects on socio-emotional adjustment may not significantly influence short-term measures but could accumulate over time and have important consequences at midlife. Conversely, small short term negative effects (for example a lowered self-concept) may disappear (Gibbons et al., 1994) or be associated with improved long-term socio-emotional adjustment.

The aim of the current research was to evaluate the consequences of academic acceleration on social and emotional adjustment at age 50 using two empirical studies that meet the quality criteria outlined above (except that all outcomes were self-reported). Outcome variables related to social-emotional adjustment were diverse including positive affect, negative affect, psychological flourishing, satisfaction with work, self, and life, and physical and mental health. Measures with respectable reliability and validity were used. The current studies control for abilities, motivation, parental education, parental occupational prestige, sex of the participant, number of siblings, birth order, and grades skipped prior to the talent search. A large sample was used to minimize Type I and Type II errors. The long-term (age 50) consequences on social-emotional adjustment evaluated in these studies add to the short (age 18) and long-term (ages 23 and 33) consequences evaluated in previous SMPY studies. Last, to decrease the bias due to missing data, multiple imputation was use.

The next chapter introduces the current studies.

CHAPTER III

THE CURRENT RESEARCH

In the current studies, the effects of grade-based acceleration and subject-based acceleration on social and emotional adjustment at age 50 were evaluated. A valid indicator of social and emotional adjustment at age 50 is *well-being* (Ryan & Deci, 2001). This chapter first defines well-being. Then it describes how educational acceleration in childhood and adolescence can enhance or attenuate well-being among the intellectually talented in midlife. Last, the hypotheses of the current studies are introduced.

Well-being

Two perspectives on well-being

Two different views on well-being have created distinct research traditions (Ryan & Deci, 2001): the hedonic view, also referred to as subjective well-being (SWB) and the eudaimonic view, also referred to as psychological well-being (PWB)

The hedonic view or SWB sees well-being as people's evaluations of their lives, including "emotional reactions to events, their moods, and judgments they form about their life satisfaction, fulfillment, and satisfaction with domains such as marriage and work" (Diener, Oishi, & Lucas, 2003, p. 404). According to Diener, Suh, Lucas and Smith (1999), SWB has four components: positive affect, negative affect, life satisfaction, and domain satisfaction (e.g., work, family, & health). Positive and negative affect are two distinct dimensions of SWB (see Figure 1; Watson & Tellegen, 1985). The positive affect (PA; Watson, Wiese, Vaidya, & Tellegen, 1999) dimension reflects the co-occurrence of positive mood states. A person that reports feeling active and elated is also more likely to feel enthusiastic, excited, peppy, and strong, and less likely to feel drowsy, dull, sleepy, or sluggish. The negative affect (NA) dimension reflects the co-occurrence of distress and dissatisfaction. A person who reports feeling distressed or fearful, is also more likely to feel hostile, jittery, nervous, angry, and sad, but less likely to feel at rest, calm, placid or relaxed. As can be seen on Figure 1, Watson and Tellegen (1985) hypothesized PA and NA to be orthogonal dimensions. Subsequent studies, however, have found a moderate negative correlation between positive and negative affect (estimated correlations range from –.43 to –.46; Diener, Smith, & Fujita, 1995; Barrett & Russell, 1998; Tellegen, Watson, & Clark, 1985).

Despite the moderate correlation between PA and NA, there is strong evidence that they are distinct dimensions. For example, PA and NA are differentially related to the "Big Two" personality traits extraversion and neuroticism, respectively. Watson et al. (1999) found strong positive correlations between NA and neuroticism (r =.58) and between PA and extraversion (r =.51), and moderate, negative correlations between NA and extraversion (r = –.25) and between PA and Neuroticism (r = –.33). These correlations have been replicated by a recent meta-analyses of Steel, Schmidt, and Schulz (2008). When controlling for the overlap between extraversion and neuroticism (r = –.31), the correlations between PA and neuroticism and between NA and extraversion decreases (*r* = –.08 and –.17, respectively; Watson et al., 1999).

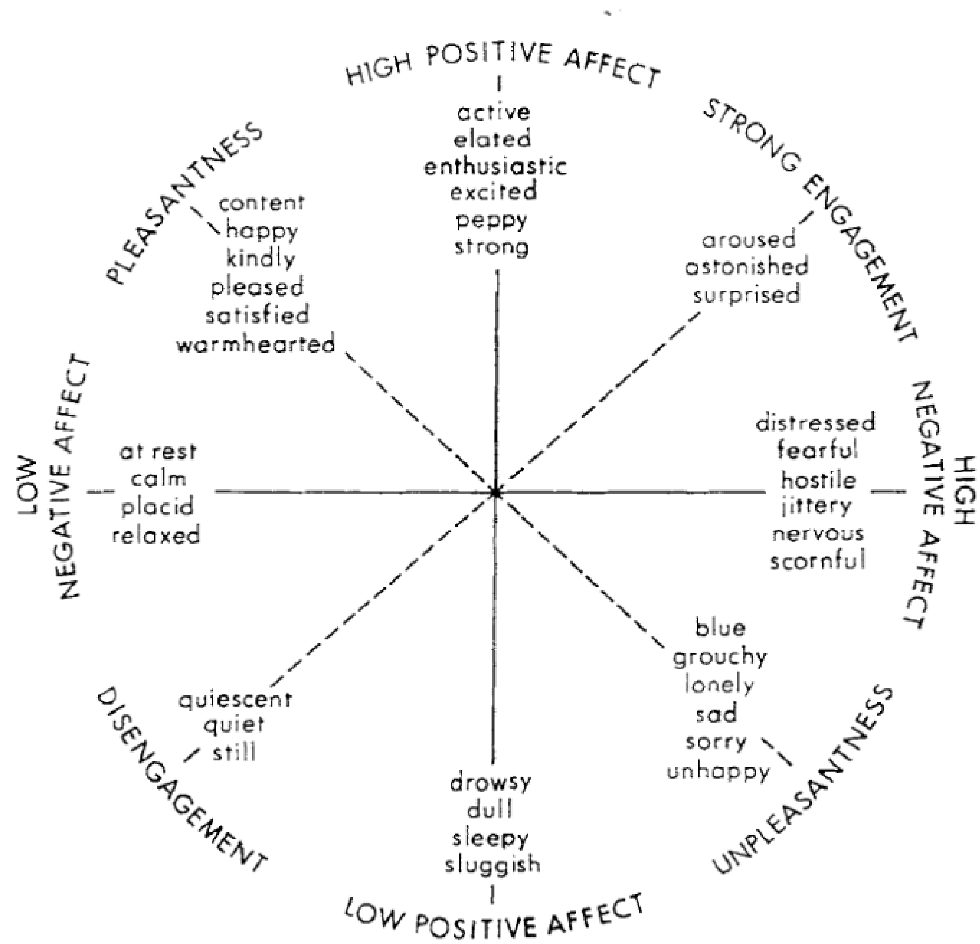

Figure 1: The Two-Factor Structure of Affect.

From "Towards a consensual structure of mood." by D. Watson and A. Tellegen. 1985. Psychological Bulletin, 98, p. 221.

In addition, PA and NA have been differentially linked to two evolutionary adaptive motivational systems that mediate goal-directed approach and avoidance behavior (e.g., Carver & White, 1994; Depue, Luciana, Arbisi, Collins, & Leon, 1994; Tomarken & Keener, 1998). The NA dimension has been linked to the *behavioral inhibition system* (BIS; Fowles, 1987, 1994) which helps the organism to avoid aversive stimuli including pain and punishment. The PA dimension has been associated with the *behavioral activation system* (BAS; Fowles, 1987) that motivates the organism to seek pleasure and reward. Where evolutionary the BIS would keep the organism out of trouble, the BAS would promote survival by ensuring that the organism obtained necessary resources (including food, shelter, & sexual partners; Watson et al., 1999). In addition, the PA dimension is associated with depressed affect and social anxiety, whereas the NA, although generally linked to psychopathology, is more strongly related to distress and dysphoria (Watson, Gamez, & Simms, 2005).

The eudaimonic view, also referred to as psychological well-being (PWB), defines well-being as the development towards a "true self", living life congruent with deeply held values and being fully engaged in life (Ryan & Deci, 2001). There is no consensus about what constitutes PWB. For example, Ryff and Keyes (1995) argue that PWB has six dimensions: positive evaluation of self and life (Self-Acceptance), personal growth and development (Personal Growth), perceiving one's life as meaningful (Purpose in Life), having good relationships with others (Positive Relations with Others), feeling effective in managing one's life and surroundings (Environmental Mastery), and having a sense of self-determination (Autonomy). However, factor analysis provides support for three underlying dimensions, capturing Self-Acceptance, Environmental Mastery, Purpose in Life, and Personal Growth in a single dimension (Ryff & Singer, 2006; Springer & Hauser, 2006; Springer, Hauser, & Freese, 2006).

Explaining differences in well-being

Wilson (1960) explained individual differences in SWB with following postulates: (1) "...satisfaction of needs causes happiness, while the persistence of unfulfilled needs causes unhappiness", and 2) "The degree of fulfillment required to produce satisfaction depends on adaptation or aspiration level, which is influenced by past experience, comparison with others, personal values, and other factors" (Wilson, 1967, p. 302). Differences in SWB are the result of an interaction between external, objective causes (bottom-up) and personality traits that influence how an individual perceives these causes and reacts to them (top-down; Diener, 1984).

Self Determination Theory (Deci & Ryan, 2000) defines a psychological need as "innate psychological nutriments that are essential for ongoing psychological growth, integrity, and well-being" (Deci & Ryan, 2000, p. 229). Three psychological needs have been identified: need for autonomy, need for competence, and need for relatedness. Need for autonomy refers to the need to have choices, to self-initiate, and to be in charge of one's life. Competence refers to a need to feel effective in bringing about desired effects and outcomes. Relatedness refers to the need to care about others, and to be cared about by others.

Fulfillment of these needs is directly related to SWB (e.g., increased positive affect, decreased negative affect, and happiness) and PWB (e.g., integrity, vitality, and self-congruence; Ryan & Deci, 2001; see also Kasser & Ryan, 1993, 1996, 1999; Miquelon & Vallerand, 2008; Reis, Sheldon, Gable, Roscoe, & Ryan, 2000; Sheldon, Ryan & Reis, 1996). For example, Reis, Sheldon, Gable, Roscoe and Ryan (2000) followed 68 students for 14 days. Average levels of autonomy, competence, and relatedness were associated with well-being. After controlling for average levels, daily fluctuations in need fulfillment was related to fluctuations in well-being.

Acceleration and well-being at age 50

Acceleration may affect well-being at age 50 through multiple pathways.

Peer acceptance and the development of social competence

Acceleration may decrease peer acceptance and increase social isolation, resulting in less opportunity to develop social competence (Newcomb & Bagwell, 1995; see also Asher & Gottman, 1981). Low levels of social competence can prohibit the initiation and maintenance of positive social relationships in young adulthood and midlife (e.g., a romantic relationship, friendships, participation in social activities; Hartup, 1983; Haugaard & Tilly, 1988; Parker & Asher, 1987). This may thwart the need for relatedness and decrease well-being.

There may be some initial social adjustment problems (Fund for the Advancement of Education, 1957), especially for males that skip more than one grade (Terman & Oden, 1947). In general, however, accelerants tend to be well accepted by peers, do not show signs of lasting social maladjustment, participate as much in extra-curricular activities, and ten years later are as likely to be married as non-accelerants (Klausmeister & Ripple, 1962; Fund for the Advancement of Education, 1957; Pressey, 1967; Terman & Oden, 1947). Hence, previous empirical studies do not support this hypothesis.

Sustained emotional stress and physical and psychological maladjustment

More challenging academic work and more competitive peers can cause higher stress levels. In addition, older peers may confront the accelerant with social and emotional issues that he/she is not mature enough to handle (e.g., romantic involvement or sexual activity). Sustained high levels of emotional stress have been associated with decreased physical and mental health (Miller, Chen, & Parker, 2011; O'Leary, 1990). In addition, placement with older peers may interfere with making friends and establishing social reciprocity, which is related to psychosocial problems and need for counseling (Parker & Asher, 1987). Psychological disorders can be directly related to well-being, through decreased positive affect, increased negative affect, and decreased life satisfaction (e.g., symptoms of depression; Watson et al., 2005). Physical disorders that limit daily functioning or independence may undermine feelings of autonomy and competence, or pain can cause a decrease in positive affect and an increase in negative affect.

Previous studies fail to find lasting harmful effects on psychological adjustment of carefully selected, able accelerants (Fund for the Advancement of Education, 1957; Janos, Robinson, & Lunnenborg, 1989; Jin & Moon, 2006; Noble, Robinson, & Gunderson, 1993; Robinson & Janos, 1986; Terman & Oden, 1947). They found no signs of increased psychological disorders, neuroticism, narcissism, or need for counseling. In addition, Okun and George (1984) found only a small association between SWB and health as rated by others (r = .16).

Decreased academic self-concept, academic aspirations, achievement and income

Perceived academic ability has a positive association with academic abilities and a negative association with the average school and classroom ability, previously referred to as the big-fish-little-pond-effect (BFLPE; Marsh & Hau, 2003; Seaton, Marsh, & Craven, 2009). Average school ability was negatively related to future achievement, effort, educational aspirations, coursework selection and academic choices (Marsh, 1991). If this effect persists, it may result in lower general self-concept, and a decrease in educational achievement, subsequent job status, and income. A previous meta-analysis found an effect size of .13 between education and SWB (Witter, Okun, Stock, & Haring, 1984). However, the effect of education on SWB may be mediated by income. In a nationally representative US sample, Diener, Sandvik, Seidliz, and Diener (1993) found a correlation of .12 between income and SWB. After controlling for income, the effect of education on SWB disappeared (Campbell, Converse, & Rogers, 1976). In addition, lower self-concept may thwart the need for competence.

Accelerants tend to have a lower academic self-concept than equally able non-accelerants (Gibbons et al., 1994; Kulik, 2004; Richardson & Benbow, 1990). However, there is no negative association between acceleration and either educational achievement or educational aspirations. To the contrary, both are positively related with acceleration (Kulik, 2004). In addition, there is evidence for long term beneficial effects of acceleration on achievement (Park, Lubinski, & Benbow, 2013; Wai, Lubinski, Benbow, & Steiger, 2010). As mentioned before, although the academic self-concept decreases, it remains positive and becomes more realistic (Dai, 2004; Plucker et al., 2004).

Current research

The current research evaluates the effects of grade-based acceleration (skipping one or more grades in high school) and subject-based acceleration (receiving accelerated curriculum without full-time placement with peers older in chronological age) on social and emotional adjustment in midlife in the top 1% in mathematical ability. It builds on the studies of Park et al. (2013) and Wai et al. (2010).

Study 1

Park et al. (2013) provided evidence for the educational efficacy of skipping one or more grades in high school. They concluded that among students in the top 1% in mathematical ability, grade skippers compared to matched controls were more likely to obtain advanced degrees in STEM and have at least one peer reviewed publications in STEM, they earned their degrees and authored their first STEM publication earlier, accrued more total citations and had more highly cited papers. The current research (Study 1) evaluated the effect of this accelerative educational intervention on well-being at age 50. Measures of well-being at age 50 included positive affect, negative affect, psychological flourishing, and satisfaction with their relationship, self, work, and life. Because the effects of grade skipping on well-being may be mediated by mental and physical health, these outcomes were included as secondary indicators of well-being.

In the first phase, the groups of grade skippers and matched controls used by Park et al. (2013) were followed-up and compared on their well-being and health at age 50.

In the second phase, to reduce effects of differential attrition between grade skippers and matched controls, propensity score matching (Ho, Imai, King, & Stuart, 2007; Rubin, 2010) was used to match grade skippers who responded to the age 50 follow-up with controls on background variables including sex, abilities, motivation, grades skipped prior to high school, number of siblings, birth order, parental education, and parental occupation. These newly matched groups were then compared on well-being and health at age 50.

Study 1 tested the hypothesis that skipping one or more grades in high school for intellectually talented youth results in psychological maladjustment or distress at midlife. Based on previous findings, this hypothesis was not expected to be supported by the data.

Study 2

Wai et al. (2010) provided evidence for a positive association between STEM educational dose (i.e., "the advanced and enriching pre-collegiate STEM learning opportunities beyond the norm that students participate in", p. 861) and STEM accomplishments at age 33 in the top .5% in mathematical ability (Wai et al., 2010). They found that STEM educational dose (accelerated educational interventions limited to Science, Technology, Engineering, and Mathematics; STEM) was positively associated with STEM accomplishments, including earning a STEM PhD, having a STEM publication, having a STEM tenure track faculty position, and having a STEM occupation. The current research (Study 2) evaluated the effects of these opportunities on their well-being and health at age 50.

In the first phase, the high and low STEM educational dose groups of Wai et al. (2010) were followed-up and compared on their well-being at age 50. Measures of well-being were identical to Study 1. A comparison of means and proportions between groups that were created using median split (the methodology used by Wai et al., 2010), however, did not control for confounding background variables including ability, SES, and motivation. In addition, attrition may have differentially affected the high and low STEM educational dose groups.

In the second phase, to control for these confounding background variables and differential attrition, the propensity score matching framework was extended to encompass the continuous treatment *STEM educational dose* (Imai et al., 2004). Subgroups were formed based on generalized propensity scores (predicted STEM educational dose from a model that regresses STEM educational dose on the same background variables as in study one). For each subgroup, outcomes related to well-being at age 50 were regressed on STEM educational dose and all background variables using linear and logistic regression.

Study 2 tested the hypothesis that subject-based acceleration among mathematically talented youth results in psychological maladjustment or distress at midlife. This hypothesis was not expected to receive empirical support.

CHAPTER IV

STUDY 1: THE EFFECTS OF SKIPPING ONE OR MORE FULL GRADES IN HIGH SCHOOL ON WELL-BEING AT AGE 50

Introduction

Grade skipping is an educational intervention that targets intellectually precocious students. It provides them with developmentally appropriate content by skipping over what they already know or can easily and rapidly assimilate (Colangelo, Assouline, & Lupkowski-Shoplik, 2004; Stanley, 2000) and places them with peers who are older in chronological age (Southern & Jones, 1991; 2004).

Using the same data set and the same methodology as the current study, Park, Lubinski, and Benbow (2013) concluded that among students in the top 1% in mathematical ability, grade skippers compared to matched controls were more likely to obtain advanced degrees in STEM and have at least one peer reviewed publication in STEM, they earned their degrees and authored their first STEM publication earlier, accrued more total citations and had more highly cited papers. This study evaluated the association between grade-based acceleration (skipping one or more grades in high school) and well-being 35 years later. It tested the hypothesis that skipping one or more grades in high school among mathematically talented youth results in psychological maladjustment or distress at age 50.

First, this chapter introduces the Neyman-Rubin Causal Model (NRCM; Holland, 1986; Neyman, 1990; Rubin, 2004) as a framework for the propensity score matching. Second, it describes the participants, measures, and analyses. Last, it reviews and discusses the results.

Propensity Score Matching and the NRCM

A propensity score is a participant's probability to receive a treatment (in our case skipping one or more grades in high school). It is often estimated using a logistic regression with the confounding background covariates included in the matching procedure as predictors. The result is a predicted probability of receiving the treatment (the propensity score) that ranges from 0 to 1. This propensity score has two functions. First, it facilitates the process of matching participants on multiple background covariates by reducing the dimensionality to a one-dimensional propensity score. Second, given the propensity score, the conditional distribution of the treatment assignment is no longer related to the background covariates. It is the latter feature of the propensity score that makes it useful in the investigation of causal effects in a quasi-experimental design. The NRCM (Holland, 1986; 1990; Rubin, 2004) clarifies how the propensity score can be used in causal inference.

The Neyman-Rubin Causal Model

Rubin (2004) defines causal effects as "comparisons of the potential outcomes that would have been observed under different exposures of *units* to *treatments*" (p. 343–344). A unit (participant) can receive the treatment (skipping a grade in high school) or not receive the treatment. Let $Y_i(1)$ be the potential outcome if unit *i* would receive the treatment, and $Y_i(0)$ the potential outcome if unit *i* would not receive the treatment. The causal effect for unit *i* is the difference between these potential outcomes, or

$$Y_i(1) - Y_i(0) \tag{1}$$

Only one of these potential outcomes can be observed. When the experiment is completed, the participant received the treatment or did not receive the treatment, but never both. This is called the "fundamental problem of causal inference" (Holland, 1986). However, this causal effect can be estimated. The following reasoning closely follows Rubin (1974).

If there are multiple units in an experiment, the subscript *i* can be used to denote the i^{th} unit (i = 1, 2, ... , *N*) out of *N* units. In addition, if the measure of interest is not the causal effect for one particular unit, but the average causal effect for the *N* units, the statistic of interest can be expressed as

$$\frac{1}{N}\sum_{i=1}^{N}[Y_i(1) - Y_i(0)] \tag{2}$$

If there are only two units in our analysis, one exposed to the treatment and the other exposed to the control condition, the average causal effect would be:

$$\mu_c = \frac{1}{2}[Y_1(1) - Y_1(0) + Y_2(1) - Y_2(0)] \tag{3}$$

The best estimate for this effect, depending on what unit gets assigned to the treatment condition, is either

$$\hat{\mu}_{c1} = Y_1(1) - Y_2(0) \tag{4}$$

or

$$\hat{\mu}_{c2} = Y_2(1) - Y_1(0) \tag{5}$$

These equations are not necessarily close to equation (3) or the causal effect of each unit individually. Their equivalence depends on the assignment mechanism, or how units are assigned to the treatment or the control condition. If this assignment is random, both equations (4) and (5) are equally likely to be observed. The expected estimated causal effect becomes

$$\hat{\mu}_c = \frac{1}{2}[Y_1(1) - Y_2(0)] + \frac{1}{2}[Y_2(1) - Y_1(0)] \quad (6)$$

which equals equation (3).

Random assignment is not always possible. It can be unethical to withhold a treatment that is known to be effective from participants who want to be treated. Some randomized studies would take many years to obtain results (e.g., the effect of childhood nutrition on longevity). In addition, by limiting the set of valid empirical study designs to completely randomized trials, useful information from observational studies would be ignored (Rubin, 1974).

To investigate causal effects in studies that do not have completely randomized designs, the concept of 'regular designs' can be used (Rubin, 2004). Contrary to completely randomized designs, in regular designs the treatment assignment is allowed to depend on covariates and can be different from unit to unit.

For a completely randomized design with two conditions, treatment and control, the probability of being assigned to either condition does not depend on any covariates, or

$$Pr[W|X, Y(1), Y(0)] = Pr[W] = .50 \quad (7)$$

where $W = (W_1, W_2, \ldots, W_N)^T$ is a vector containing the treatment assignments for the *N* units, with $W_i = 1$ if the unit is assigned to the treatment and $W_i = 0$ if the unit is not assigned to the treatment, and X is a matrix of covariates. (Note that this implies that in a randomized design, the missing potential outcomes are missing completely at random)

On the other hand, in a regular design, the assignment to the treatment or control group is allowed to be dependent on background covariates, or

$$Pr[W|X, Y(1), Y(0)] = Pr[W|X] \quad (8)$$

The regular design is different from the randomized design in two ways (Rosenbaum & Rubin, 1983). First, the treatment assignment is not determined by a specified random mechanism, but by an unknown mechanism. Therefore, the probability for each unit to be assigned to the treatment condition is unknown [contrary to (7)]. Although this probability can be estimated using for example logistic regression with background covariates as predictors, these estimated probabilities remain an imperfect approximation of the true assignment mechanism. Previous studies have shown that errors in model specifications can induce considerable bias to the estimated treatment effects (Kang & Schafer, 2007; Smith & Todd, 2005). One important aspect of model specification is the selection of covariates to be included as predictors. This is the second difference. In randomized designs, none of the covariates is related to the treatment assignment. That is not the case for a regular design. Still, to make a valid causal inference using a regular design, all covariates that are possibly related to both the treatment assignment as well as the outcomes of interest must be controlled for. Therefore, it has to be assumed that all confounding covariates are included in our observed background covariates. More formally, if U is the matrix of all unobserved covariates, this assumption states that

$$Pr[W|X,U,Y(1),Y(0)] = Pr[W|X,U] = Pr[W|X] \tag{9}$$

The first and last part of this equation also illustrate the assumption that the *treatment assignment is strongly ignorable*, or that given the known covariates X, the treatment assignment is unrelated to the potential outcomes (Ho et al., 2007; Rosenbaum & Rubin, 1983).

These covariates can then be used to construct a paired comparison randomized experiment (Rubin, 2004). Assuming that treatment assignment is strongly ignorable, and if pairs of units are exactly matched on all covariates, it is reasonable to assume that the units will react equally to the treatment and equations (4) and (5) will be equivalent. Let S_E be the subset that includes all indices of the units that are assigned to the experimental group, and let S_C be the subset that includes all indices of units that have been assigned to the control group, then

$$\frac{2}{N}\sum_{i \in S_E} Y_i(1) - \frac{2}{N}\sum_{i \in S_C} Y_i(0) \tag{10}$$

will be an unbiased estimate of the average causal effect in the sample.

Creating pairs of units that are identical on all covariates is not always feasible. As the number of covariates increases, the probability of finding exact matches decreases. However, exact matching is not necessary if there exists a one-dimensional balancing score $e(X)$ for which

$$Pr[W|X] = Pr[W|e(X)] \tag{11}$$

or that given the balancing score, the covariates and the treatment assignment are independent (Rosenbaum & Rubin, 1983). If this holds, it would be sufficient to match units on the one-dimensional balancing score which would greatly reduce the complexity of finding matched units. Of course, in practice $e(X)$ cannot be known, but must be estimated. The balancing score used in this study is the propensity score (p_i), which is defined as the probability of being assigned to the treatment or

$$p_i \equiv Pr[W|X] \tag{12}$$

Conditional on the propensity scores, the background covariates are no longer related to the treatment assignment or

$$Pr[W|p_i, X] = Pr[W|p_i] \tag{13}$$

Pairs can be created that are matched on estimated propensity scores, including one member that received treatment and one that did not. However, because they have the same propensity score, both members were equally likely to receive the treatment. Assuming strong ignorability of the treatment assignment, the assignment of one member to the treatment and one to the control condition was not related to any of the confounding covariates and as close to random as possible in a nonrandomized design.

Propensity score matching has been used in research fields including psychology (e.g., Becker et al., 2012; Park et al., 2013), economics (e.g., Haas et al., 2012), medical science (Redd et al., 2012), and accountancy (Peel et al., 2012).

Average treatment effect on the treated

When treatment assignment depends on covariates, those who receive the treatment may be substantially different from those who did not receive the treatment on one or more covariates. For example, students who have skipped a grade may be considerably more able than those who did not. When grade skippers are matched with equally able non-skippers, the non-skippers may be considerably more able than the population of non-skippers. In the words of Meehl (1971): they become unrepresentative for their respective population. However, this study does not target the overall average treatment effect, but the effect of the treatment on those that have been treated. Grade skippers are compared with equally able participants who did not skip a grade, yielding an average treatment effect on the treated (ATT):

$$\frac{1}{\sum_{i=1}^{N} W_i} \sum_{i=1}^{N} W_i E[Y_i(1) - Y_i(0)|X_i] \tag{14}$$

After matching the grade skippers with control units, equation (10) is an unbiased estimator of the ATT.

Stable unit treatment value assumption

An important assumption in propensity score matching is the *stable unit treatment value assumption* (SUTVA; Rubin, 1980). It has two parts. First, whether a particular unit receives treatment or not is assumed not to affect the potential outcomes of any other unit. Rubin (2004) gives the example of two people participating in a headache experiment. If one person receives a treatment for headaches and as a result of the treatment stops complaining, this may affect the headache of the other person, regardless of whether he receives the treatment or not. In this case the SUTVA is violated. Second, it is assumed that all units receive the same treatment. Skipping a grade in elementary school can have a different effect on social-emotional adjustment than skipping a grade in high school. By restricting our treatment to grades skipped in high school, it remains plausible to accept the SUTVA.

Some reflections about matching

When using matching to control for confounding variables, some pitfalls must be avoided (Bai, 2011; Meehl, 1971; Stigler & Miller, 1993).

First, matching on one covariate may unmatch subjects on other covariates. Matching grade skippers and non-grade skippers on IQ may unmatch them on parental support. Students with high IQ's who did not skip a grade may have had parents who were considerably less invested in their children's education than students with high IQ's who were grade-skipped. Matching on IQ decreases the difference between groups in abilities, but may increase the differences in parental support, thereby introducing bias. This problem can be avoided by matching on all variables that are known to be related to the treatment assignment. This study matches subjects on a large number of covariates, including sex, mathematical ability, verbal ability, parents' highest degree earned, parents' occupational prestige, participant's birth order, number of siblings, the number of grades skipped prior to the talent search, and some variables measuring motivation and achievement in school. It is assumed that these variables include the most important confounding covariates, thereby meeting the requirement of the strong ignorability assumption of the treatment assignment.

Second, matching can create samples that are unrepresentative of their respective populations. For example, students who skipped a grade may be considerably more able than those who did not skip a grade. This study compares grade skippers with equally able participants who did not skip a grade to estimate the effect of grade skipping on those who skipped. That control units are not representative for their respective population has then become irrelevant.

Third, causal inference from matching is as problematic as causal inference from any correlational design (Campbell & Stanley, 1966). The longitudinal nature of the data make the causal direction self-evident, but association between grade skipping and well-being at age 50 may be explained by a third confounding variable. This limitation has been addressed previously in the section about the strong ignorability assumption. It is assumed that all confounding covariates have been included in the set of background covariates.

Another perspective on matching

One does not need to agree with the NRCM to acknowledge the value of propensity score matching (PSM). PSM can be viewed as a method to pre-process the data. Pre-processing creates common support between the groups under comparison and makes results less dependent on the specified model (Ho et al., 2007). PSM is only one way of attaining that result.

Participants

Participants were drawn from Cohorts 1 and 2 of the Study of Mathematically Precocious Youth (SMPY), a 40 year longitudinal study of intellectual talent (Lubinski & Benbow, 2006). Students who scored in the top 3% on conventional achievement tests routinely administered in their school (e.g., the Iowa Test of Basic Skills) were invited to take the College Board Scholastic Aptitude Test (SAT) at or before the age of 13. Initially constructed for college bound high school seniors to assess mathematical and verbal reasoning ability necessary for college work, at age 13 it functions more on an analytical reasoning level because formal instruction has not yet been initiated. At this age, the SAT has sufficient ceiling to differentiate among students in the top 1% in mathematical and verbal abilities (Park et al., 2007, 2008).

Cohort 1 was identified between 1972–1974 and will be referred to as the 1972 Cohort. It includes 2,188 participants (96% Caucasian, 2% Asian, 2% other) who scored in the top 1% on either the mathematical subtest (SAT-M $\geq$ 390) or the verbal subtest (SAT-V $\geq$ 370). The majority is from the State of Maryland, mostly the Baltimore-Washington area.

Cohort 2 was identified between 1976–1979 and will be referred to as the 1976 Cohort. It includes 778 participants (89% Caucasian, 6% Asian, 5% other) who scored in the top 1 in 200 (SAT-M $\geq$ 500 or SAT-V $\geq$ 430). They came mostly from mid-Atlantic states.

Identical to Park et al. (2013), only participants who scored at or above 390 on the SAT mathematics subtest, or the top 1% in mathematical reasoning ability were included.

Participants were surveyed at ages 18, 23, and 33 through phone, mail, internet surveys (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Lubinski, Benbow, Webb, & Bleske-Rechek, 2006) and searches of public internet databases (Proquest Dissertation and Theses database, Google Scholar, and Google patents). Outcome measures in this study come from the age 50 follow-up, an internet based survey conducted from 8 January 2012 to 18 February 2013. Completion of the survey yielded a $20 Amazon gift card or a donation of equal amount to scholarships for Vanderbilt's Programs for Talented Youth.

Table 1 gives an overview of the proportion of grade skippers and matched controls that constituted the comparison groups of Park et al. (2013) and that have completed the age 50 follow-up survey.

Table 1: Number of Grade Skippers and Matched Controls who Constituted the Comparison Groups of Park et al. (2013) and Have Completed the Age 50 Follow-Up Survey

	Grade Skippers			Matched Controls		
	Park et al. (2013)	Age 50 follow up	Proportion	Park et al. (2013)	Age 50 follow up	Proportion
1972 Cohort	179	89	0.50	358	191	0.53
Men	102	47	0.46	204	102	0.50
Women	77	42	0.55	154	89	0.58
1976 Cohort	116	68	0.59	231	143	0.61
Men	81	45	0.56	162	99	0.61
Women	35	23	0.66	69	44	0.64
Both Cohorts	295	157	0.53	589	333	0.57
All Men	183	92	0.50	366	200	0.55
All Women	112	65	0.58	223	133	0.60

Measures

Grade skipping

Participants reported at age 18 and 23 which grades they had skipped in elementary and secondary school. Only interventions that happen after the assessment of the background variables included in the matching procedure, can be evaluated without introducing bias (Ho et al., 2007). Therefore a dichotomous variable was created that reflected whether or not a participant had skipped one or more grades in high school. The number of grades skipped prior to high school was included as a background covariate in the matching procedure.

Background covariates

Background information was retrieved from the age 13 surveys. All background covariates were assessed before the onset of treatment (before high school). Different covariates were obtained for the 1972 and the 1976 Cohorts and are identical to the background covariates used by Park et al. (2013).

The 1972 Cohort background covariates include participant's sex, SAT-mathematics subtest score (the SAT-verbal score could not be included because most students were selected based on their SAT-M scores and lacked SAT-V scores), parents' highest degree earned (1= less than high school, 7=doctoral degree), parents' occupational prestige (coded using the socioeconomic index for occupations of Duncan, 1961), participant's birth order (1= first born, 2=second born, etc.), number of siblings, and the number of grades skipped prior to the talent search. In addition, students responded to the following items:

1. Circle the word that best describes your liking for school (1=positive dislike, through 4=very strong)

2. Circle the word that best describes your liking for arithmetic and mathematics (1=positive dislike, through 4=very strong)
3. Check the one statement that best describes how well you are doing in your mathematics class this year. (1=Less well than the majority of your class mates, 2=About as well as most of your class mates, 3=Better than all but one or two of my class mates, 4=Better than all of your class mates)
4. This school year, how are you learning most of your arithmetic and mathematics? Check only one. (1=In regular class work with other students, 2=In school, but working on your own with some help or direction from your teacher, 3=On your own, outside of school, helped by a tutor or parent, 4=On your own outside of school with little help from anyone)
5. How important do you think math will be for the job you will someday have? Circle one. (1=Not at all, through 5=Very)

This resulted in 14 background covariates.

The 1976 cohort background covariates included participant's sex, both SAT-M and SAT-V, type of school attended at the time of identification (public, private, or parochial), parents' highest degree earned (1= less than high school, 6=more than college), number of siblings, and the number of grades skipped prior to the talent search. In addition, students responded to the following items:

1. Check the box that best describes your liking for school / arithmetic and mathematics / biology / chemistry / physics (each subject was scored separately, 1=positive dislike, through 4=very strong)
2. Check the one statement that best describes how well you are doing in your mathematics class this year. (1=Better than all of your class mates, 2 = Better than all but one or two of my class mates, 3=About as well as most of your class mates, 4=Less well than the majority of your class mates). The same question was asked for the science class.
3. This school year, how are you learning most of your arithmetic and mathematics? Check only one. (1=In regular class work with other students, 2=In school, but working on your own with some help or direction from your teacher, 3=On your own, outside of school, helped by a tutor or parent, 4=On your own outside of school with little help from anyone). The same question was asked about learning science.
4. How important do you think math will be for the job you will have someday? Circle one. (1=Not at all, 4=Very). The same question was asked for biology, chemistry, and physics.

This resulted in 21 background covariates.

All questions and scales below were included in the age 50 questionnaire.

Primary Outcomes: Subjective and Psychological Well-being

Satisfaction with work

How satisfied are you with the current direction of your professional career? Participants answered on a 7-point Likert-type scale (very dissatisfied – very satisfied)

Satisfaction with relationship

Overall, how satisfied are you with your relationship? Participants answered on a 7-point Likert-type scale (very dissatisfied – very satisfied).

Satisfaction with self

The *Core Self-Evaluations Scale* (CSES; Judge, Erez, Bono, & Thoresen, 2003) is a self-report scale that purports to measure a basic, fundamental appraisal of one's worthiness, effectiveness, and capability as a person. It is a latent higher order concept that combines the traits self-esteem, generalized self-efficacy, neuroticism, and locus of control. Participants are asked to indicate their agreement with 12 statements on a 5-point Likert-type scale (strongly disagree – strongly agree). Typical items are: "When I try, I generally succeed", and "Sometimes when I fail, I feel worthless" (reverse scored). Details about validity and reliability can be found in Judge et al. (2003).

Satisfaction with life

The *Satisfaction With Life Scale* (Diener, Emmons, Larsen, & Griffin, 1985) is a self-report scale that purports to measure global life satisfaction. In the age 50 questionnaire, participants were asked to indicate their agreement with 5 statements on a 7-point Likert-type scale (strongly disagree- strongly agree). A typical item is "I am satisfied with my life". Diener et al. (1985) have shown that the scale has respectable validity and reliability.

Positive affect

The *Emotional Wellness Scale* (Positive Affect Only) (Diener & Biswas-Diener, 2008) purports to measure overall emotional well-being. Participants are asked to indicate on a 5-point Likert-type scale (very rarely or never – very often or always) how often they experience following feelings: positive, good, pleasant, contented, interested, happy, loving, and joyful.

The construct measured by this scale is different from the *positive affect* construct as defined by Watson and Tellegen (1985). It coincides more with their *pleasantness* construct. Throughout this study, positive affect is used to maintain consistent with label assigned to this scale by Diener & Biswas-Diener (2008).

Negative affect

The age 50 questionnaire included a 50-item personality inventory that captured the "Big Five" personality traits (Goldberg, 1992). Negative affect was measured with the neuroticism subscale. Participants were asked to indicate on a 7-point Likert-type scale how accurately each statement described themselves. There were 10 statements measuring neuroticism. An example of a typical statement is: "I often feel blue".

The construct measured by this scale is a combination of the *negative affect* and *unpleasantness* constructs as defined by Watson and Tellegen (1985). Throughout this study, the term negative affect will be used to refer to this construct.

Psychological Well-being

The *Psychological Flourishing Scale* (Diener & Biswas-Diener, 2008) is a self-report scale with 12 items covering important aspects of human functioning, including positive relationships, feelings of competence, having meaning and purpose in life, and self-acceptance. Participants are asked to indicate their agreement with each item on a 7-point Likert-type scale (strongly disagree – strongly agree). A typical item is: "I lead a purposeful and meaningful life".

Secondary Outcomes: Physical and Mental health

Two sections of the Health at 40+ module of NLSY79 were included in the age 50 survey: severe and common health problems

Severe health problems

Participants are asked to indicate if they have been told by a doctor that they have any of the disorders listed. Severe health problems include stroke, congestive heart failure, chronic lung disease, heart problems, hypertension, diabetes, arthritis/rheumatism, emotional/nervous/psychiatric problems, and non-skin cancer. Answer options were *yes* or *no*.

Common health problems

Participants were asked to indicate if they had any health problems, including eye trouble, ulcer, epilepsy, severe headaches, dizziness or fainting spells, and depression or excessive worry or nervous trouble. Answer options for the 44 items were *yes* or *no*.

Missing Data

Out of a total of 1,712 respondents, the majority of the respondents (n = 1,375) completed all well-being items. A group of partial respondents (n = 244) did not complete any well-being item. The remaining respondents (n = 93) had at least one item missing. For each scale, the average of a respondent's completed items was used to impute a respondent's missing items.

Hundred fifteen respondents had one or two health items missing. Seventeen respondents had three or more health items missing. Because the list of health items was extensive (44 items), leaving an item open may have been a time saving strategy for some respondents. Therefore, if a respondent completed at least one health items, missing items were interpreted as indicating absence of the health problem and rescored as zero.

Multiple imputation was used to handle missing values in the background covariates (Rubin, 1976, 1987; Schafer, 1997). Assuming multivariate normality and missingness at random (MAR), m values are imputed for each missing value, with the variability between the m imputed values reflecting the uncertainty about the missing data point (Honaker, King, & Blackwell, 2010). This results in m imputed data sets. Honaker et al. (2010) suggest using m = 5 when missingness is not very high. Similar to Park et al. (2013), in this study m equals 10. All datasets are used in parallel in the matching process (hence resulting in 10 matched control groups) and the subsequent analyses (resulting in 10 separate estimates for each parameter). Preliminary results are then combined to obtain the final result in the following way. The point estimate of the parameter ($\bar{q}$) is obtained by averaging the point estimates of that parameter (q_j) across m imputed datasets, or

$$\bar{q} = \frac{1}{m}\sum_{j=1}^{m} q_j \qquad (15)$$

The variance for the resulting point estimate is the sum of the variance of the point estimates within each data set ($SE(q_j)^2$) and the sample variance of the point estimates across datasets, $S_q^2 = \sum_{j=1}^{m}(q_j - \bar{q})^2/(m-1)$, multiplied with a correction factor because $m < \infty$. This yields

$$SE(q)^2 = \frac{1}{m}\sum_{j=1}^{m} SE(q_j)^2 + S_q^2(1 + \frac{1}{m}) \qquad (16)$$

Imputation algorithms assuming multivariate normality perform as well in datasets including both categorical and continuous variables as alternatives specifically tailored to such datasets (the background covariates include the dichotomous variable sex of the participant; Schafer, 1997; Schafer & Olsen, 1998). Previous studies have shown that this procedure reduces bias, increases efficiency, and better recovers the actual data structure compared to list wise deletion (King, Honaker, & Scheve, 2001; Honaker et al., 2010). Multiple imputation was done using the Amelia II package in R (Honaker et al., 2007; Horton & Kleinman, 2007; King et al., 2001). To maximize multivariate normality, Box-Cox transformations were applied to all variables. All background covariates and well-being outcomes were included in the imputation procedure. Health outcomes were excluded because an excess of categorical variables prohibited the imputation algorithm to converge.

Because Study 1 Phase 1 intends to follow-up the grade skippers and matched controls of Park et al. (2013), the same imputed background covariates are used as in Park et al. (2013). These multiply imputed background covariates were used to create 10 matched datasets. Although it is best practice to include outcomes (in this case well-being and health) in the imputation process, redoing imputation would decrease the similarity between the covariate distributions of the grade skippers and the matched controls because they were matched using the by Park et al. (2013) imputed background covariates. For all other analyses, newly imputed datasets were used.

Evaluating Balance

The goal of matching is to increase the similarity between the covariate distributions of the grade skippers and the matched controls, also called "balance". Because of age 50 attrition, balance between grade skippers and matched controls (i.e., from both groups those that responded to the age 50 survey) has to be re-evaluated. No golden standard for evaluating balance exists. Four criteria were used to evaluate balance (Ho et al., 2007; Imai & van Dyck, 2008):

1. Graphical and formal tests of equality between univariate propensity score density estimates across grade skippers and matched controls (method based on Bowman & Azzalini, 1997; R package: sm).
2. Standardized Mean Differences (SMD) of the background covariates, calculated as the difference of the means between grade skippers and matched controls, divided by the square root of their pooled variances. The resulting SMD equals Cohen's *d*.
3. Kolmogorov-Smirnov test of equality of two continuous one-dimensional probability distributions (Kolmogorov, 1933; Smirnov, 1948). For each background covariate, the distance between the empirical distribution functions of grade skippers and matched controls is compared with the test-statistic's null distribution that assumes that both samples were drawn from the same distribution (R package: Matching, function: ks.boot).
4. The Cross-match test (Rosenbaum, 2005; R Package: crossmatch), a distribution-free test for comparing multivariate distributions. Using the distance between subjects in the multidimensional space (with the background covariates as the dimensions), disjoint pairs are formed that minimize the total distance within pairs. The number of cross-match pairs (one grade skipper and one matched control) is compared with the number of expected cross-match pairs if observations were sampled from the same multivariate distribution.

Results Study 1 Phase 1

In Phase 1, the grade skippers and matched controls used by Park et al. (2013) are followed-up and compared on their well-being and health at age 50.

Evaluating balance

Because not all participants included in the Park et al. (2013) groups completed the age 50 questionnaire, balance between accelerants and matched controls (i.e., from both groups those that responded to the age 50 survey) had to be re-evaluated.

Figure 2 shows the propensity score density distributions averaged across imputed datasets for grade skippers and matched controls for both the original matched groups from Park et al. (2013) and those who completed the age 50 survey. Using a Dunn-Šidák (Šidák, 1967) correction for multiple comparisons, none of the imputed datasets (across both cohorts) showed a statistically significant difference between the propensity score density distributions of grade skippers and matched controls. The Dunn-Šidák (Šidák, 1967) procedure keeps the familywise type I error rate at $\alpha = .05$ by requiring every individual test to have an error rate of $1-(1-\alpha)^{1/C}$ This procedure is less conservative than Dunn's multiple comparison procedure (also called the Bonferroni procedure; Dunn, 1961) because $1-(1-\alpha)^{1/C} > \alpha/C$.

Tables 2 and 3, and Figures 3 and 4 show the absolute standardized mean differences (ASMD) on all background covariates between grade skippers and matched controls.

For the 1972 Cohort, the ASMD increases for most background covariates. However, 11 out of 14 ASMD's remain below .10. Largest differences were found for the SAT Math, number of grades skipped prior to high school, and liking for school. None of the ASMD exceeded .25, the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (WWC; 2009). Across imputed datasets, none of the Kolmogorov-Smirnov tests for similarity between univariate distributions was statistically significant. Table 25 (Appendix) shows the results of the cross-match test for the matched groups of Park (2013) of the 1972 cohort before and after attrition. After controlling the family-wise error rate, none of the cross-match tests were significant. The cross-match test failed to provide evidence for non-equivalence of the multivariate background covariate distributions of the grade skippers and matched controls.

For the 1976 Cohort, 11 out of 21 ASMDs decreased. The largest ASMD were the SAT Math, liking for school, and liking for the math class. All ASMDs remained well under .25. Figure 3 shows that these differences were small and always favored the grade skippers. Across multiply imputed datasets, none of the KS-tests were significant after controlling for a family wise error rate (21 background covariates). In addition, none of the cross-match test results were significant (see Table 26).

To conclude, all criteria fail to provide evidence for non-equivalence on background covariates for grade skippers and matched controls after attrition.

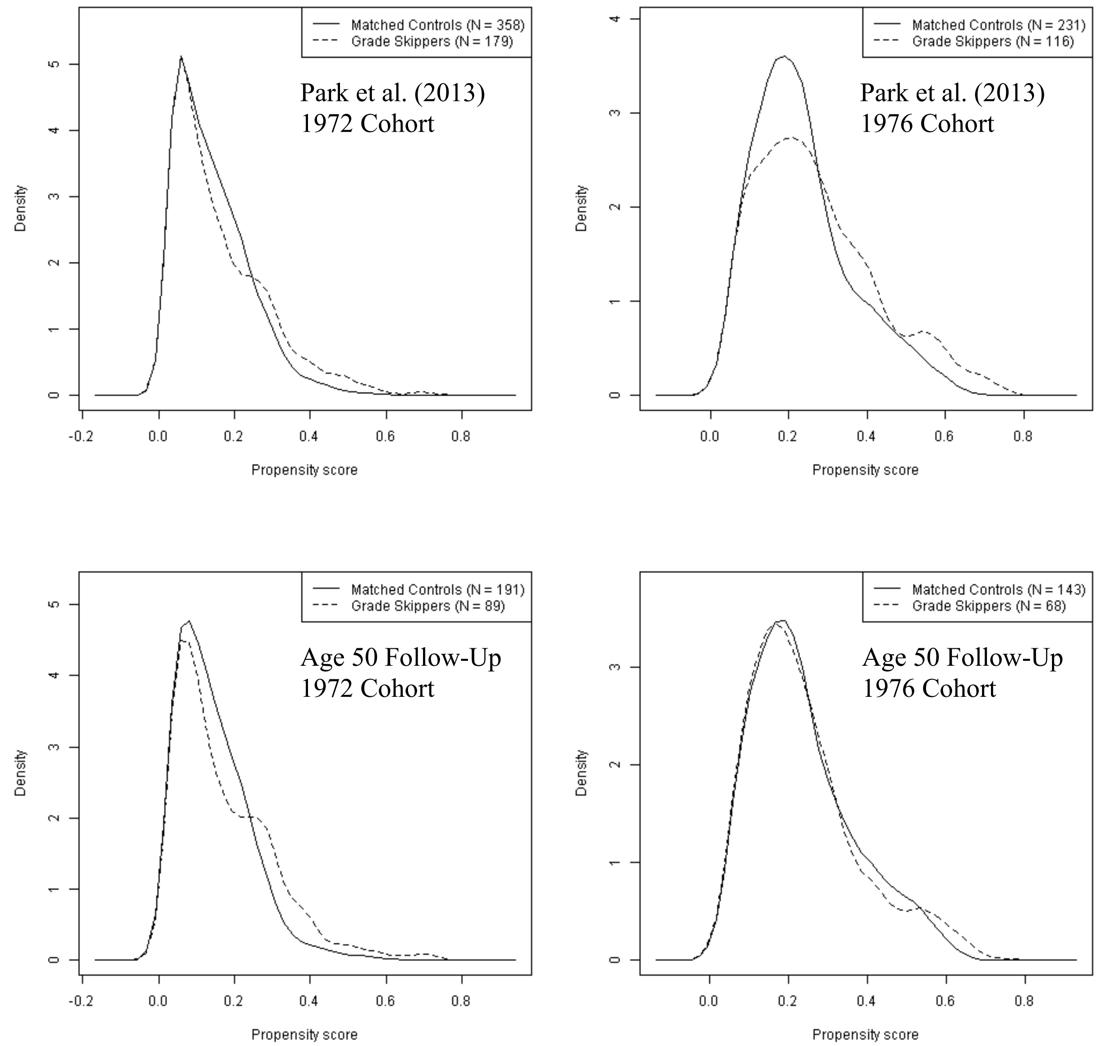

Figure 2: Density Plots Comparing the Propensity Score Distributions for the 1972 and 1976 Cohorts.

Density plots comparing the propensity score distributions of grade skippers (dashed lines) and matched controls (solid lines) for the groups used in Park et al. (2013; upper panel) and of these groups those who completed the age 50 survey (lower panel). The results for the 1972 Cohort are presented in the left column, the results for the 1976 Cohort in the right column.

Table 2: Means and Proportions of 13 Background Covariates Measured at Age 13 for the 1972 Cohort

	Park et al. (2013)			Age 50 follow-up		
	Grade Skippers	Matched Controls	ASMD	Grade Skippers	Matched Controls	ASMD
N	179	358		89	191	
SAT Math	568	559	0.10	577	560	0.21
Mother's highest degree	3.7	3.6	0.03	3.7	3.8	0.05
Father's highest degree	4.5	4.5	0.00	4.7	4.6	0.07
Mother's occupational prestige	74	75	0.00	75	75	0.03
Father's occupational prestige	78	78	0.01	78	79	0.05
Birth order	2.0	2.0	0.03	2.0	2.0	0.06
Number of siblings	2.3	2.2	0.04	2.4	2.3	0.05
Liking for school	3.2	3.1	0.03	3.1	3.2	0.11
Liking for math	3.5	3.5	0.02	3.5	3.5	0.07
Doing well in math class	3.0	3.0	0.04	3.0	3.0	0.01
Learning math	1.4	1.4	0.04	1.4	1.4	0.07
Math importance	4.4	4.4	0.03	4.4	4.3	0.07
Previous grades skipped	0.2	0.2	0.00	0.3	0.2	0.17
Proportion Male	0.57	0.57	0.00	0.53	0.53	0.01

Note. ASMD = Absolute Standardized Mean Difference

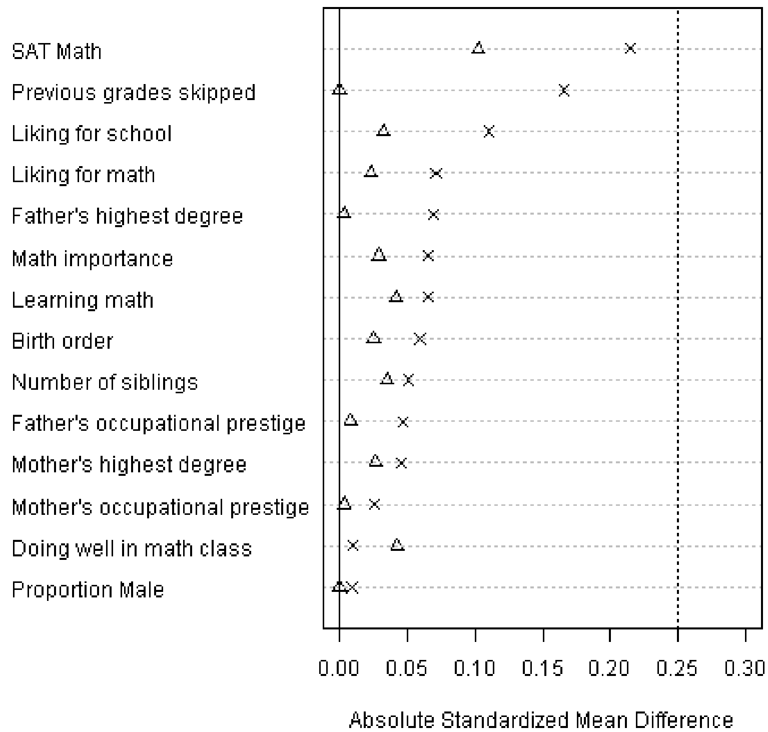

Figure 3: Absolute Standardized Mean Differences Between Grade Skippers and Matched Controls of the 1972 Cohort

The absolute standardized mean differences between the grade skippers (n = 179) and the matched controls (n = 358) of Park et al. (2013; triangles), and for grade skippers (n = 89) and matched controls (n = 191) who completed the age 50 survey (x's) on important background covariates. The dotted vertical line indicates the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (2009). Data come from the 1972 Cohort.

Table 3: Means and Proportions of 21 Background Covariates Measured at Age 13 for the 1976 Cohort

	Park et al. (2013)			Age 50 follow-up		
	Grade skippers	Matched controls	ASMD	Grade skippers	Matched controls	ASMD
N	116	231		68	143	
SAT Math	577	567	0.13	556	570	0.20
SAT Verbal	482	474	0.11	469	475	0.08
Mother's highest degree	4.7	4.6	0.09	4.8	4.7	0.07
Father's highest degree	5.4	5.3	0.06	5.4	5.4	0.07
Number of siblings	1.8	1.7	0.02	1.8	1.7	0.04
Liking for school	3.9	4.0	0.05	3.9	4.0	0.12
Liking for math class	4.4	4.4	0.00	4.3	4.4	0.10
Liking for biology class	3.5	3.5	0.01	3.5	3.6	0.04
Liking for chemistry class	3.9	3.9	0.04	3.8	3.8	0.02
Liking for physics class	3.8	3.8	0.07	3.7	3.7	0.06
Doing well in math class	1.8	1.8	0.06	1.8	1.8	0.01
Doing well in science class	1.9	1.9	0.07	1.9	1.9	0.05
Learning math	1.6	1.5	0.11	1.6	1.5	0.09
Learning science	1.2	1.2	0.05	1.2	1.2	0.02
Math importance	3.6	3.6	0.04	3.6	3.6	0.07
Biology importance	2.4	2.4	0.02	2.4	2.5	0.03
Chemistry importance	2.8	2.8	0.06	2.9	2.8	0.05
Physics importance	3.2	3.1	0.13	3.1	3.1	0.07
Proportion Male	0.70	0.70	0.00	0.66	0.69	0.07
Previous grades skipped	0.19	0.18	0.00	1.23	1.22	0.00
Proportion in public school	0.84	0.84	0.01	0.16	0.16	0.01

Note. ASMD = Absolute Standardized Mean Difference

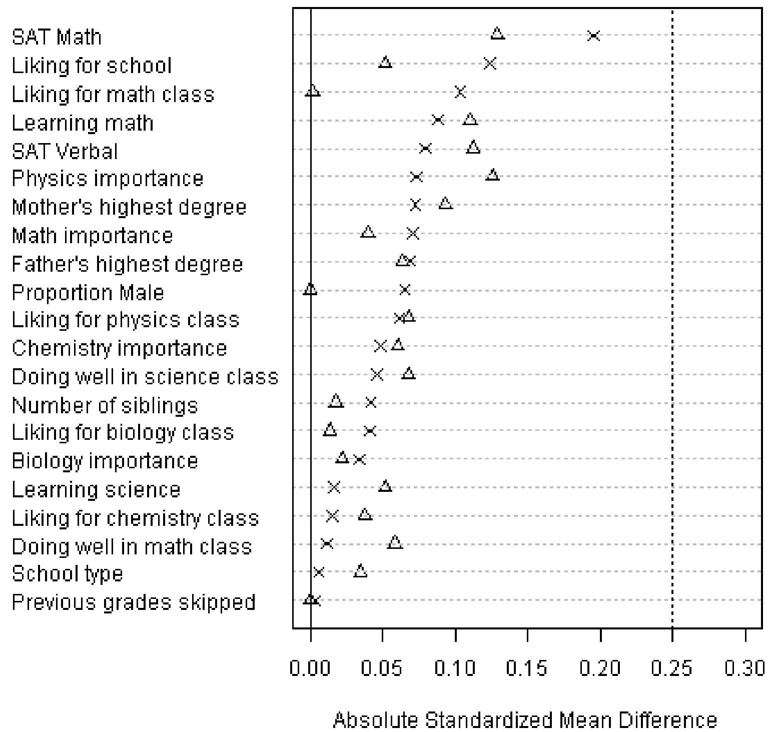

Figure 4: Absolute standardized mean differences between grade skippers and matched controls of the 1976 Cohort

The absolute standardized mean differences between the grade skippers (*n* = 116) and the matched controls (*n* = 231) of Park et al. (2013; triangles), and for grade skippers (*n* = 68) and matched controls (*n* = 143) who completed the age 50 survey (x's) on important background covariates. The dotted vertical line indicates the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (2009). Data come from the 1976 Cohort.

Well-Being

In a first step the mean scores on measures of Positive Affect, Negative Affect, Life Satisfaction, Psychological Flourishing, Core Self-Evaluations, Career satisfaction, and Relationship Satisfaction were compared for grade skippers and matched controls who completed the age 50 survey. The results are presented in Table 4. Mean scores for both groups were very similar. On average, the participants endorsed items indicating positive well-being, and their scores were similar to or better than a normative sample (see Table 27). Only on the Core Self-Evaluations scale the normative sample scored slightly higher..

To interpret the magnitude of the difference between grade skippers and matched controls, it is important to take into account the differences in the number of items and the standard deviations of the total score for each scale. Linear regression was used to estimate the standardized mean difference, adjusted for background covariates. The regression coefficient for the dichotomous variable that indicates whether a participant skipped one or more grades in high school can be interpreted as a SMD if the dependent variable is standardized. Using this procedure, the SMD can be adjusted for background covariates by adding these background covariates as predictors in the linear regression. These Adjusted Standardized Mean Differences (AdSMD) are visually presented in Figure 5. The horizontal lines indicate a 95% confidence interval.

Each well-being outcome has AdSMD estimates for the 1972 cohort, the 1976 cohort, all men from both cohorts, all women from both cohorts, and both cohorts combined. For the 1972 and 1976 cohort estimates, all their respective background covariates were included as predictors in the regression. Because background covariates were different or scored differently across cohorts, results for all men, all women, and both cohorts only included SAT Math, sex of the participant, and grades skipped prior to high school as additional predictors.

Model assumptions of the linear regression were tested using the Breusch-Pagan test for heteroscedasticity (Breusch & Pagan, 1979; R package: lmtest, function bptest), the Shapiro-Wilk test for normality of the residuals (R package: stats, function: Shapiro.test) and a Bonferroni corrected *t*-test for outliers (R package: car, function: outlierTest). Presence of outliers and non-normally distributed residuals signaled the need for robust regression (R package: MASS, function: rlm), which uses iteratively re-weighted least squares to decrease the influence of outliers on the estimated model parameters. A Huber (Huber, 1981) objective function was used. In addition, a Box-Cox power transformation of the well-being outcomes was used prior to the robust linear regression to correct for non-normality of the residuals. Because applying the Box-Cox transformation did not change the results but made straightforward interpretation of the model estimates difficult, results are presented without the Box-Cox transformation.

All estimated AdSMD's in Figure 5 are small and not significantly different from zero. The difference between female grade skippers and female matched controls tends to favor the grade skippers, whereas the opposite pattern can be seen in male grade skippers and male matched controls. This is in line with previous studies that show that girls tend to benefit from grade skipping more than boys (Terman & Oden, 1947; Fund for the Advancement of Education, 1957). Possible explanations for this pattern are an on

Table 4: Means on Well-Being Measures for Grade Skippers and Matched Control from Park et al. (2013) Who Completed the Age 50 Survey.

Cohort	Group	N	Positive Affect	Negative Affect (reversed)	Life Satisfaction	Psychological Flourishing	Core Self-Evaluations	Career Satisfaction	Relationship Satisfaction
1972 Cohort	Matched Controls	191	22.2	47.0	25.5	70.1	44.9	5.4	6.1
	Grade Skippers	89	22.6	47.0	24.7	70.0	45.0	5.3	6.2
1976 Cohort	Matched Controls	142	22.0	45.7	25.4	69.6	44.6	5.5	6.3
	Grade Skippers	68	22.0	46.3	25.9	68.7	45.7	5.3	6.1
Both Cohorts	Matched Controls	333	22.1	46.5	25.5	69.9	44.8	5.4	6.2
	Grade Skippers	157	22.4	46.7	25.2	69.5	45.3	5.3	6.2

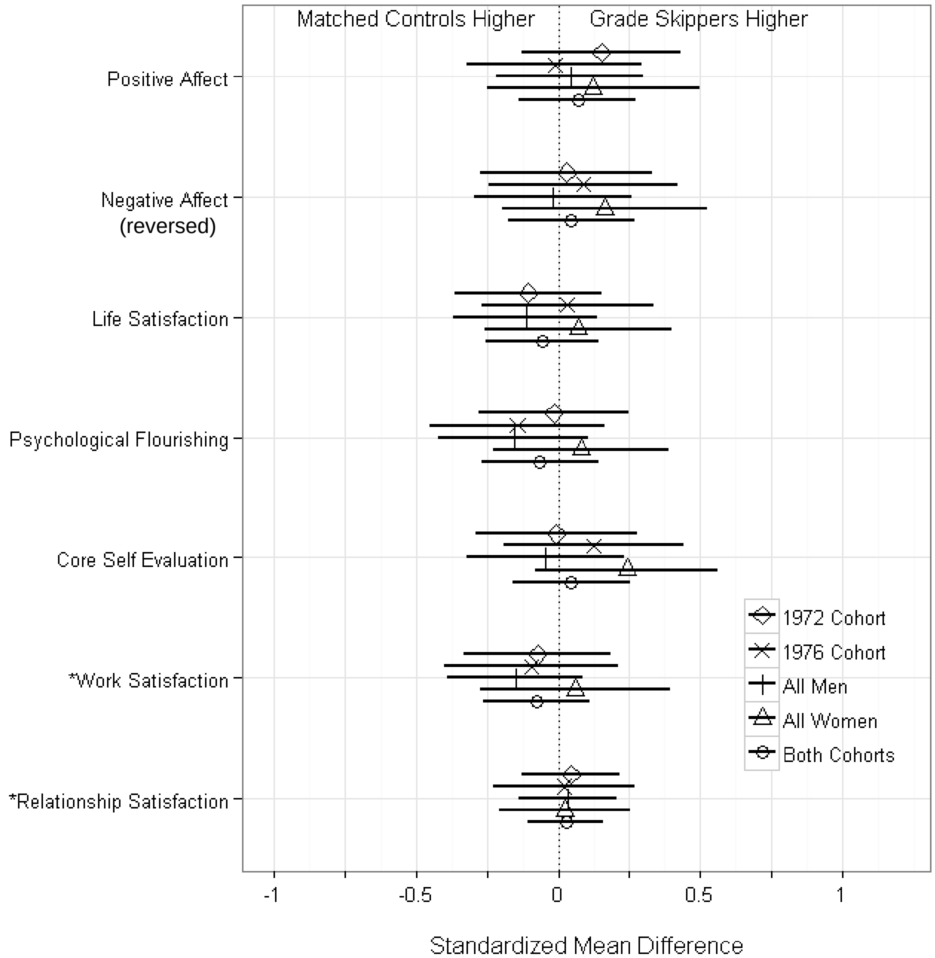

Figure 5: Estimated Adjusted Standardized Mean Differences between Grade Skippers and Matched Controls for Well-Being Outcomes

Adjusted Standardized Mean Differences between grade skippers and matched controls on measures of well-being and Core Self-Evaluations. Point estimates represent regression coefficients for the dichotomous variable that indicates whether the participant skipped at least one grade in high school (receiving a score of 1) or did not skip a grade in high school (receiving a score of 0). For the 1972 Cohort and 1976 Cohort, all background covariates were included as predictors in the linear regression. Because the 1972 and 1976 background covariates were overlapping but not identical, all other analyses only included overlapping background covariates (SAT-M, sex, and previous grades skipped). Horizontal lines represent a 95% confidence interval. Shapes represent different samples (see legend). Results are combined across the 10 imputed data sets. Data come from the 1972 Cohort and the 1976 Cohort. * Outcomes with a star represent a single item. For these items, the regression residuals were not normally distributed, so results must be interpreted with care.

average earlier onset of puberty and higher levels of social and emotional maturity in girls compared to boys. In addition, male freshmen are more willing to date younger accelerated females than female freshmen may be to date younger, accelerated males (Fund for the Advancement of Education, 1957). But neither in girls nor in boys is the difference between grade skippers and matched controls statistically significant.

To explore this difference between men and women further, mean score on well-being scales by sex are presented in Table 5. For each cohort, the mean scores on the well-being scales were calculated for males and females separately. At the bottom of the table, the mean scores for both cohorts combined are presented. In the 1972 cohort and the results for both cohorts combined male matched controls tend to outperform the male grade skippers, and the reverse is true for females. But differences are small and there is no evidence to refute that these patterns originate from random fluctuation.

To conclude, there is no evidence for any harmful effect of grade skipping on well-being at age 50.

Health

Table 6 and 7 give an overview of the prevalence of severe and common health issues across different samples. The first column gives the prevalence among all participants that have responded to the age 50 follow-up. Subsequent columns give the prevalence for grade skippers and matched controls for the 1972 cohort, the 1976 cohort, all males, all females, and both cohorts. Compared to a nationally representative sample at age 50 (National Longitudinal Survey of Youth 1979; see Table 28), the overall prevalence of all severe health issues is equal or lower among the participants of this study, indicating better health. Only one health item (emotional, nervous, or psychiatric problems) was endorsed more by the current sample . These comparisons must be interpreted with care. Without weights, the NLSY79 sample is not necessarily a nationally representative sample because minorities were oversampled. On the other hand, the weights provided for the 2010 survey may not be applicable. The data collection procedure (the 50+ health module was only administered if participants had turned 50 between the last biennial survey and the current survey) would require a complex combination of weights from multiple survey years. In addition, at the time of the last survey (2010), only 24% of the participants had turned 50 and had completed the 50+ health module. Another explanations for the higher prevalence of diagnosed emotional, nervous, or psychiatric problems is the increased access to (mental) health care for more able individuals. The discrepancy between this item and the on average high score on the well-being scales discussed previously warrants a cautious interpretation.

NLSY79 did not survey the common health issues at age 50, which precluded normative comparison for those items.

Figure 6 gives a visual overview of Table 6 and 7. The difference between grade skippers and matched controls in percentage endorsement of each health item are plotted on the x-axis for the 1972 cohort and on the y-axis for the 1976 cohort. Each label represents a health item. The purpose of Figure 6 is to visualize consistent patterns across cohorts (upper panel) and all men and all women (lower panel).

Table 5: Means on Well-Being Measures by Sex for Grade Skippers and Matched Control from Park et al. (2013) who Completed the Age 50 Survey

Cohort	Group	N	Positive Affect	Negative Affect (Reversed)	Life Satisfaction	Psychological Flourishing	Core Self-Evaluations	Career Satisfaction	Relationship Satisfactioon
1972 Cohort	Men	149	22.1	48.1	25.2	68.9	45.3	5.3	6.1
	Matched Controls	102	22.0	48.7	25.5	69.4	45.6	5.4	6.0
	Grade Skippers	47	22.4	46.8	24.4	67.8	44.6	5.2	6.2
	Women	141	22.6	45.7	25.4	71.5	44.6	5.3	6.2
	Matched Controls	89	22.5	45.0	25.5	71.0	44.2	5.3	6.3
	Grade Skippers	42	22.8	47.3	25.0	72.5	45.6	5.5	6.2
1976 Cohort	Men	144	21.6	46.4	25.2	68.7	45.2	5.4	6.2
	Matched Controls	99	21.6	46.0	25.1	68.9	44.9	5.4	6.3
	Grade Skippers	45	21.7	47.1	25.4	68.4	45.7	5.3	6.0
	Women	67	22.8	45.0	26.4	70.6	44.5	5.6	6.3
	Matched Controls	44	22.8	45.2	26.2	71.3	43.9	5.6	6.3
	Grade Skippers	23	22.8	44.6	27.0	69.4	45.8	5.4	6.4
Both Cohorts	Men	292	21.9	47.3	25.2	68.8	45.2	5.4	6.2
	Matched Controls	200	21.8	47.4	25.3	69.1	45.3	5.4	6.2
	Grade Skippers	92	22.0	47.0	24.9	68.1	45.1	5.2	6.1
	Women	198	22.7	45.5	25.7	71.2	44.6	5.4	6.3
	Matched Controls	133	22.6	45.0	25.8	71.1	44.1	5.4	6.3
	Grade Skippers	65	22.8	46.4	25.7	71.4	45.6	5.4	6.3

Health issues that are more prevalent in matched controls for both samples appear in the top right quadrant. The bottom left quadrant shows health issues that are more prevalent among grade skippers for both samples. The top left and bottom right quadrant represent health issues where there is inconsistency among the samples.

The axes have to be interpreted with care. Because they represent the difference in percentage endorsement between grade skippers and matched controls, larger scores do not necessarily indicate more prominent differences. Base rates for health issues have to be taken into account to evaluate the meaning of the difference. Risk ratios would be better measures to investigate meaningful differences. Because for many health issues at least one of the groups had no members endorsing the item, risk ratios could not be calculated for those items. To be able to plot all health issues, difference in percentage endorsement was chosen. Health issues that are more common are printed in a darker color.

Most health issues cluster around zero; grade skippers and matched controls were in both samples equally likely to endorse the item. Only two health items showed a consistent pattern across cohorts: matched controls were more likely to have emotional, nervous, or psychiatric problems, and grade skippers tended to have somewhat more ear, nose, and throat troubles. Neither of these differences between grade skippers and matched controls were statistically significant.

When comparing grade skippers and matched controls across all men and all women, we find that in both sexes matched controls were more likely to have emotional, nervous, or psychiatric problems. This difference is larger for women and without correction for multiple comparisons would have been statistically significantly different from zero. Although most health issues cluster around zero, there are slightly more in the top left quadrant. Those are health issues that were more common in female matched controls compared to female grade skippers and male grade skippers compared to male matched controls. This pattern is consistent with the pattern that emerged in Figure 5, with females benefiting more form grade-skipping than males. None of these differences, however, was statistically significant.

A useful statistic to compare percentages is the risk ratio, obtained by dividing the percentage of grade skippers that endorsed a health item by the percentage of matched controls that endorsed a health item. Because the matched group is not representative of the general population, but similar to the grade skippers on all background covariates, the risk ratio can be interpreted as the increase in the probability of having a particular health issue due to the grade skipping.

Logistic regression was used to adjust the risk ratio for background covariates. (Although conditional logistic regression seems optimal for matched data, in this study many matched pairs were incomplete due to attrition.) Only health items that had sufficient endorsement from both groups were included to avoid complete separation or lack of convergence in the estimation process due to insufficient information in the data. Risk ratio's were obtained by dividing the expected risk for grade skippers by the expected risk for matched controls. A risk ratio of 1 indicates that both groups are equally likely to endorse the health item. A risk ratio higher than 1 indicates a higher prevalence among the grade skippers. If the risk ratio is lower than 1, matched controls were more likely to endorse the health item.

Results are presented in Figure 7. Horizontal lines represent 95% confidence intervals. All horizontal lines include 1, indicating that there is no statistically significant difference in the estimated risks between grade skippers and matched controls even without a correction for multiple comparisons. For all samples, matched controls were more likely to have emotional, nervous, or psychiatric problems. But these differences were not statistically significant. Females tended to benefit more from grade skipping than males. Four out of nine health issues (arthritis, indigestion, allergies, and cholesterol) were less likely to be endorsed by female grade skippers then by female matched controls, whereas the the opposite pattern appeared for males. In addition, female grade skippers were 50% less likely to have emotional, nervous, or psychiatric problems than female matched controls. Male grade skippers were only 10% less likely to have those problems compared to male matched controls.

To conclude, there was no evidence for any harmful effect of grade skipping on health at age 50. Grade skippers were not more likely to be depressed at age 50 and were less likely to have emotional, nervous, or psychiatric problems, although these differences were not significant.

.

Table 6: Prevalence of Severe Health Issues Across Grade Skippers and Matched Controls

	All respondents	Both Cohorts		1972 Cohort		1976 Cohort		All Men		All Women	
Severe Health Issues		MC	GS	MC	GS	MC	GS	MC	GS	MC	GS
N	1474	333	157	191	89	142	68	200	92	133	65
A heart attack or myocardial infarction?	0.5	0.1	0.6	0.2	0.0	0.0	1.5	0.1	1.1	0.0	0.0
Angina or chest pains due to your heart?	1.2	0.8	0.0	1.4	0.0	0.0	0.0	1.0	0.0	0.5	0.0
Congestive heart failure?	0.2	0.4	0.0	0.6	0.0	0.0	0.0	0.5	0.0	0.1	0.0
A stroke?	0.4	0.4	0.0	0.5	0.0	0.2	0.0	0.4	0.0	0.2	0.0
Not including asthma, but chronic lung disease such as chronic bronchitis or emphysema?	0.7	0.7	0.0	1.3	0.0	0.0	0.0	0.4	0.0	1.1	0.0
High blood pressure or hypertension?	18.1	18.2	17.9	21.6	20.5	13.7	14.7	18.8	18.7	17.3	16.9
Diabetes or high blood sugar?	5.7	5.7	7.7	5.5	9.1	5.9	5.9	6.1	8.8	5.1	6.2
Arthritis or rheumatism?	13.1	13.8	13.5	16.3	20.5	10.5	4.4	9.1	11.0	20.9	16.9
Emotional, nervous, or psychiatric problems?	15.3	18.3	12.2	17.5	11.4	19.2	13.2	14.7	12.1	23.6	12.3
Cancer or malignant tumor of any kind except skin cancer?	3.2	3.6	2.6	4.5	4.5	2.5	0.0	1.5	2.2	6.8	3.1
Other heart problems?	6.4	6.1	5.1	8.0	4.5	3.7	5.9	6.3	6.6	5.8	3.1

Note. MC = Matched Controls, GS = Grade Skippers

Table 7: Prevalence of Common Health Issues Across Grade Skippers and Matched Controls

Common Health Issues	All respondents	Both Cohorts		1972 Cohort		1976 Cohort		All Men		All Women	
		MC	GS	MC	GS	MC	GS	MC	GS	MC	GS
N	1474	333	157	191	89	142	68	200	92	133	65
Eye trouble, other than glasses or contacts?	8.1	9.2	6.4	9.8	10.2	8.3	1.5	9.5	4.4	8.7	9.2
Ulcer?	1.5	1.3	1.9	1.0	3.4	1.6	0.0	1.3	2.2	1.3	1.5
Severe tooth or gum trouble?	3.6	4.4	4.5	6.4	4.5	1.7	4.4	4.0	5.5	5.0	3.1
Epilepsy or fits?	0.6	0.8	0.6	1.4	1.1	0.1	0.0	1.3	0.0	0.2	1.5
Stomach or intestinal ulcers?	2.2	1.6	3.8	1.5	6.8	1.7	0.0	1.7	3.3	1.5	4.6
Lameness or paralysis (including polio)?	0.7	0.7	2.6	1.2	3.4	0.0	1.5	0.7	2.2	0.7	3.1
Frequent trouble sleeping?	15.8	17.6	18.6	15.1	23.9	21.0	11.8	16.1	16.5	19.8	21.5
Frequent or severe headaches, dizziness or fainting spells?	5.4	5.4	7.7	4.9	10.2	6.2	4.4	3.5	6.6	8.2	9.2
Pain or pressure in your chest, palpitation or pounding heart, or heart trouble?	4.3	4.2	1.3	5.7	2.3	2.2	0.0	2.3	1.1	6.9	1.5
Anemia?	4.3	4.9	3.8	6.4	6.8	2.9	0.0	0.8	1.1	10.7	7.7
Swollen or painful joints, frequent cramps in your legs or bursitis?	6.3	5.4	7.1	6.2	10.2	4.4	2.9	2.0	4.4	10.4	10.8
Problems with your feet and legs?	11.3	10.7	13.5	10.5	20.5	10.9	4.4	7.4	14.3	15.3	12.3
Neuritis?	1.1	1.2	1.9	1.1	2.3	1.4	1.5	0.4	3.3	2.4	0.0
Asthma? (Shortness of breath or chronic cough?)	7.0	8.1	7.1	8.4	10.2	7.6	2.9	6.4	9.9	10.4	3.1
Depression or excessive worry or nervous trouble of any kind?	17.3	19.5	17.9	18.2	19.3	21.1	16.2	17.3	15.4	22.6	21.5
Kidney or bladder problems?	3.5	3.7	5.1	5.0	8.0	1.9	1.5	4.2	7.7	2.9	1.5
Hardening of the arteries?	0.5	0.4	0.0	0.8	0.0	0.0	0.0	0.6	0.0	0.2	0.0
Frequent urinary tract infections? (other than kidney problems discussed earlier?)	1.1	1.3	1.9	2.1	2.3	0.2	1.5	0.1	0.0	3.0	4.6
Scarlet fever, rheumatic fever, tuberculosis, jaundice or hepatitis?	0.8	0.7	0.6	0.7	1.1	0.7	0.0	0.4	0.0	1.1	1.5
Problems with your back?	23.5	21.7	24.4	23.0	30.7	20.0	16.2	21.3	20.9	22.2	29.2
Osteoporosis?	1.0	1.1	0.0	1.6	0.0	0.3	0.0	0.0	0.0	2.6	0.0
Frequent indigestion, stomach, liver or intestinal trouble, gall bladder trouble or gallstones?	9.6	11.6	10.3	11.2	13.6	12.3	5.9	8.7	8.8	15.9	12.3
Painful or "trick" shoulder or elbow, "trick" or locked knee?	9.4	9.5	5.1	11.0	3.4	7.5	7.4	9.9	4.4	8.9	6.2
Ear, nose, or throat trouble?	6.9	7.5	10.3	8.2	11.4	6.4	8.8	7.4	15.4	7.5	3.1
Low blood pressure?	2.6	2.2	1.9	2.1	2.3	2.4	1.5	0.9	1.1	4.1	3.1
Skin disease?	5.7	5.3	5.8	4.9	6.8	5.8	4.4	4.2	5.5	6.7	6.2
Chronic or frequent colds, sinus problems, hay fever or allergies?	21.4	22.4	24.4	23.0	20.5	21.7	29.4	18.3	25.3	28.2	23.1
Adverse or allergic reaction to any serum, drug or medicine?	13.2	15.2	11.5	16.3	17.0	13.7	4.4	10.1	7.7	22.4	16.9
Bone, joint or other deformity?	2.6	2.5	3.2	3.7	3.4	1.0	2.9	2.7	2.2	2.2	4.6
High cholesterol?	21.7	22.5	21.2	24.1	21.6	20.4	20.6	27.6	28.6	15.2	10.8
Thyroid trouble or goiter?	6.5	5.8	7.7	7.3	10.2	3.8	4.4	1.8	6.6	11.5	9.2
Tumor, growth, or cyst?	5.6	5.5	4.5	5.4	3.4	5.6	5.9	4.2	5.5	7.2	3.1
Loss of finger or toe?	0.3	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0

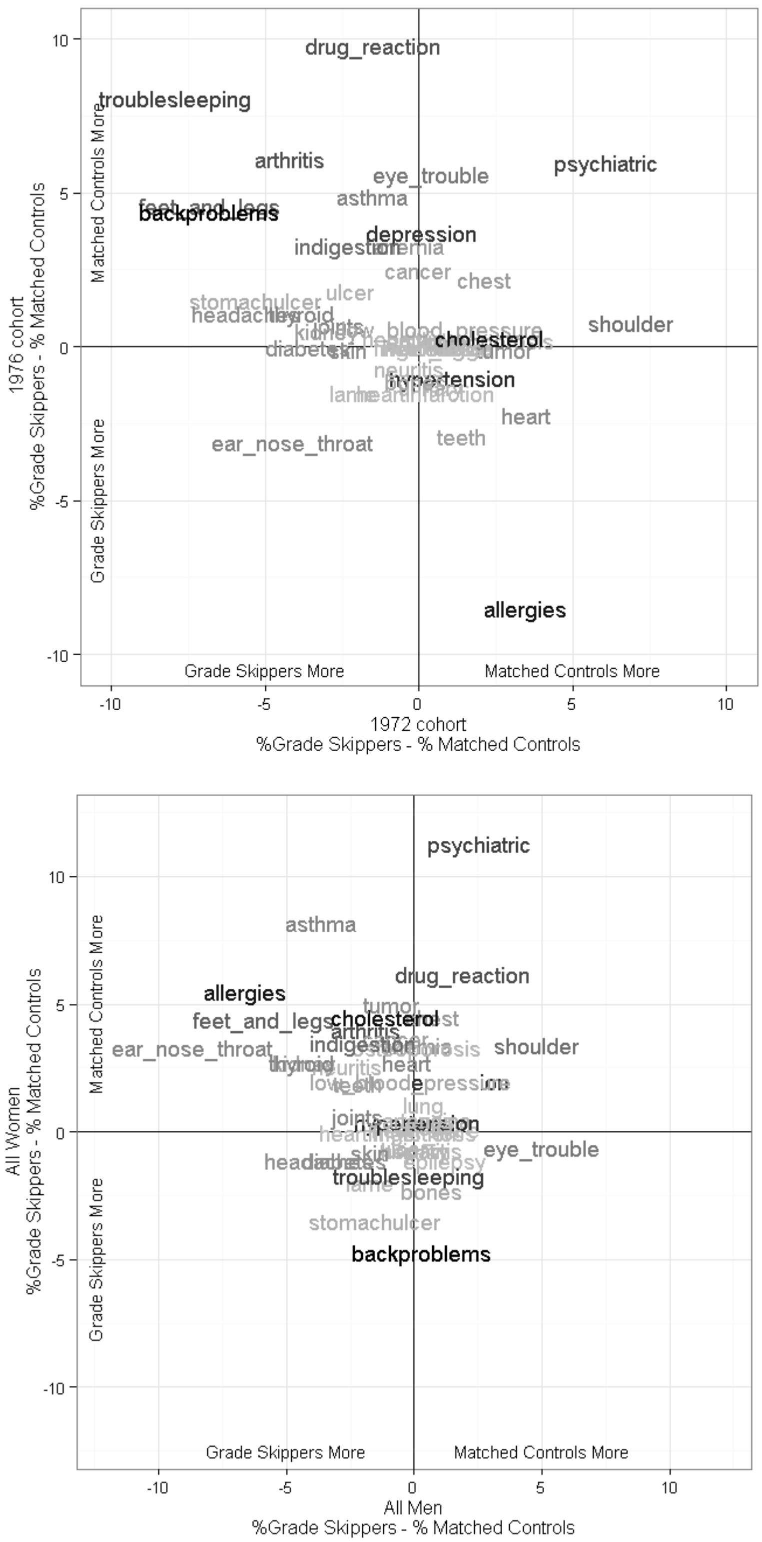

Figure 6: Differences in Percentage Endorsement of Health Items Between Grade Skippers and Matched Controls

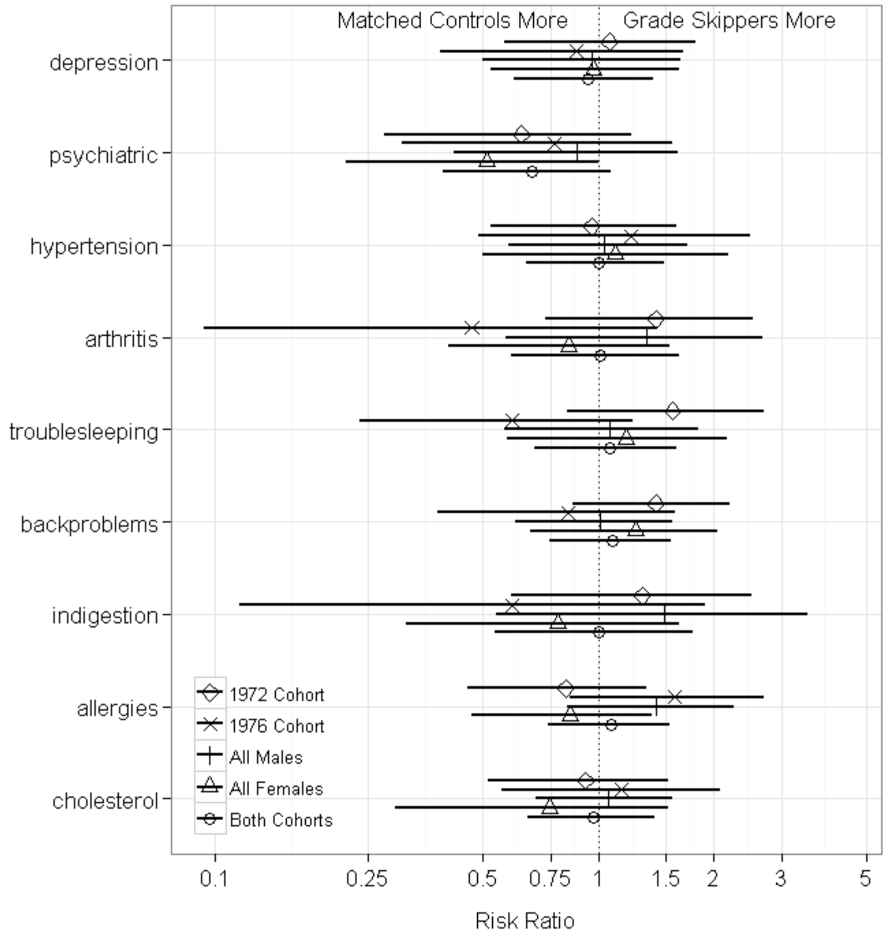

Figure 7: Estimated Risk Ratios for Selected Health Issues Comparing Grade Skippers and Matched Controls Who Completed the Age 50 Follow-Up

Point estimates represent logistic regression coefficients for the dichotomous variable that indicates whether the participant skipped at least one grade in high school (receiving a score of 1) or did not skip a grade in high school (receiving a score of 0). For the 1972 Cohort and 1976 Cohort, all background covariates were included as predictors in the logistic regression. Because the background covariates were different for the 1972 and 1976 Cohorts, other analyses only include background covariates they had in common. Horizontal lines represent a 95% confidence interval. Shapes represent different samples (see legend). Results are combined across the 10 imputed data sets. Data come from the 1972 Cohort and the 1976 Cohort.

Results Study 1 Phase 2

In Study 1 Phase 2 the propensity score matching was redone to optimize equivalence between the grade skippers and matched controls.

Propensity score matching

To maximize the equivalence between the grade skippers and the comparison group, grade skippers that responded to the age 50 follow-up were newly matched to the controls that responded to the age 50 survey using propensity score matching. First, the propensity score had to be estimated. Misspecifications of the propensity score model can lead to biased estimates of the treatment effect (Kang & Schafer, 2007; Smith & Todd, 2005). Park et al. (2013) explored several propensity score models and retained the model that optimized the covariate balance. A limitation of this procedure is that the number of possible propensity score models increases exponentially with the number of background covariates. Recently, Imai et al. (2004) suggested an automated algorithm to estimate the propensity score while simultaneously optimizing the covariate balance, resulting in the *covariate balancing propensity scores* (CBPS). The CBPS makes propensity score matching more robust to misspecifications of the propensity score model. More details about the CBPS estimation procedure can be found in Imai et al. (2004). The same propensity score model as Park et al. (2013) was used, followed by nearest neighbor matching on the covariate balanced propensity score.

Because of lower sample size due to age 50 attrition, the stringent exact matching on sex and grades skipped prior to high school used in Park et al. (2013) sets an upper boundary to the possible improvement in balance on the other background covariates. For example, in the 1972 cohort, for the seven female grade skippers who also skipped one grade prior to high school, there were only 16 possible exact matches. Maintaining a matching ratio of 2:1 implied that regardless of their scores on the other background covariates, at least 14 out of these 16 had to be included in the matched control group. Reducing the matching ratio to 1:1 or omitting the exact matching restriction reduces the ASMD on the other background covariates. Because the results for subsequent analyses using these different matching options were close to identical, only results are presented for a model with exact matching on both sex and grades skipped prior to high school, using a matching ratio of 2:1. For the 1976 Cohort, one female participant that skipped 2 grades prior to high school could not be exactly matched on sex and number of grades skipped prior to high school, because none of the female controls had skipped 2 years prior to high school. To avoid omitting her from the analyses, she was matched with two females that only skipped one grade prior to high school.

Evaluating balance

Figure 8 shows the propensity score density plots comparing grade skippers and newly matched controls. Density distributions for grade skippers are not identical to the density plots of Study 1 Phase 1 because new propensity scores were estimated. Plots show no clear improvement in the similarities

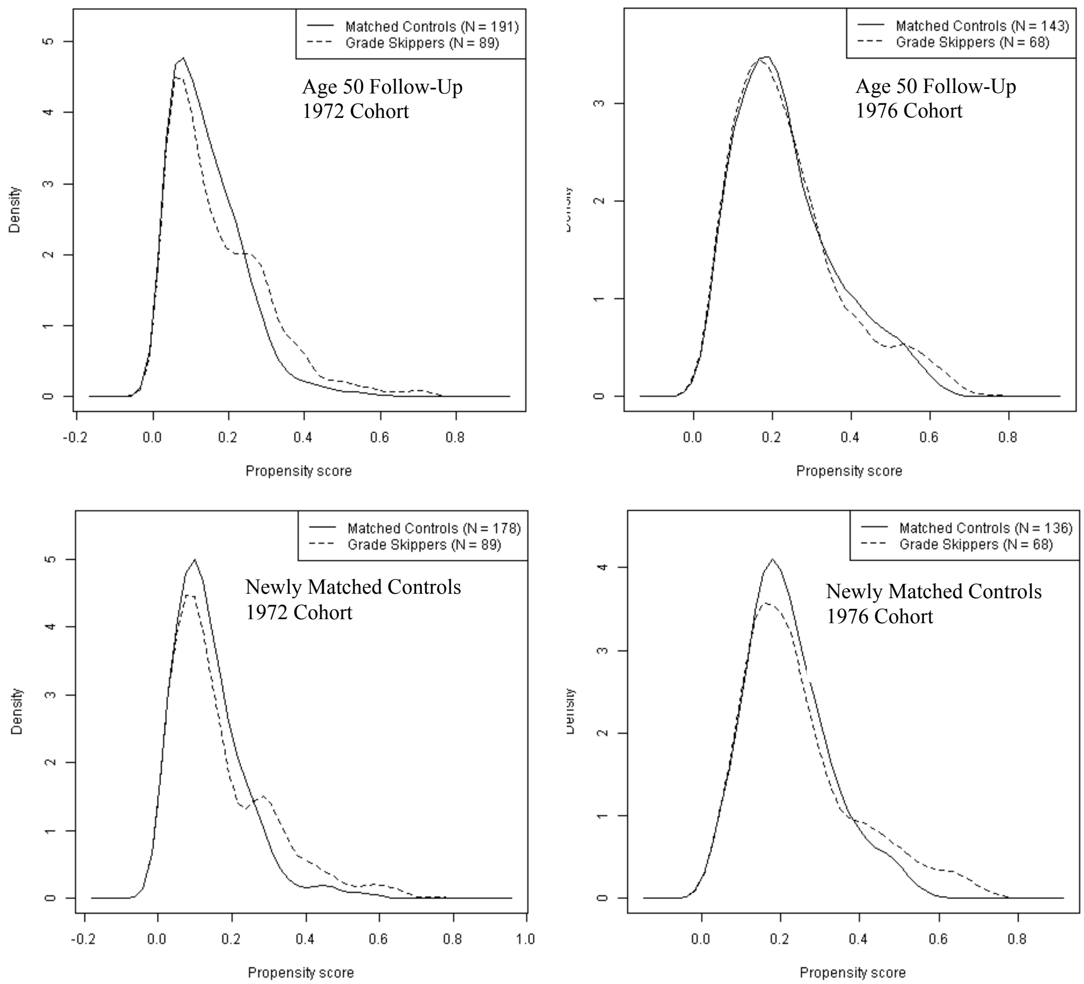

Figure 8: Density Plots of the Propensity Score Comparing Groups of Study 1 Phase 1 with Grade Skippers and Newly Matched Controls of Phase 2.

Density plots comparing the propensity score distributions of grade skippers (dashed lines) and matched controls (solid lines) for participants in the groups used in Park et al. (2013) who completed the age 50 survey (upper panel) and newly matched controls . The results for the 1972 Cohort are presented in the left column, the results for the 1976 Cohort in the right column.

between propensity score distributions for grade skippers and newly matched controls. For the 1976 Cohort, the similarity seems to decrease. However, distributional equivalence of propensity score distributions is not the criterion; it is an indicator of equivalence of the multivariate distributions of the background covariates. Tests on distributional equivalence on the individual imputed datasets yielded no statistically significant results, providing evidence for equivalence between the propensity score density distributions of grade skippers and matched controls.

Tables 8 and 9 show the means and absolute standardized mean differences for the grade skippers and the newly matched controls for the 1972 and 1976 Cohorts, respectively. These results are visually presented in Figure 26 and Figure 27.

For the 1972 cohort, the ASMD decreased for important background covariates including previous grades skipped, SAT-Math, parent's highest degree, liking for school and liking for math. All background covariates had an ASMD lower than 0.1, except for SAT-Math (ASMD = .17). After controlling for multiple comparisons, none of the KS-tests or cross-match tests for individual imputed datasets were significant, failing to refute univariate and multivariate distributional equivalence between grade skippers and newly matched controls.

For the 1976 cohort, the ASMD decreased on important covariates including the SAT-Math, SAT-Verbal, previous grades skipped, proportion male, and mother's highest degree. The ASMD increased for some variables, including doing well in the science class and learning science, but never exceeded .15. Similar to the 1972, the KS-tests and the cross-match tests for individual imputed datasets provided no evidence for the non-equivalence of the univariate and multivariate distributions for the grade skippers and matched controls.

Based on these results it is reasonable to assume equivalence between the grade skippers and newly matched controls for the 1972 and 1976 cohorts on important background covariates. Compared to the matched controls of Study 1 Phase 1, the newly matched controls are more equivalent to the grade skippers on important background covariates including previous grades skipped, sex of the participants, and SAT-Math (for the 1976 Cohort).

Table 8: Comparison of the 1972 Cohort Grade Skippers and Newly Matched Control on Background Covariates

	Park et al. (2013)			Follow-up of Park et al. (2013)			Newly matched groups		
	Grade skippers	Matched controls	ASMD	Grade skippers	Matched controls	ASMD	Grade skippers	Matched controls	ASMD
SAT Math	568	559	0.10	577	560	0.21	586	572	0.17
Mother's highest degree	3.7	3.6	0.03	3.7	3.8	0.05	3.7	3.7	0.03
Father's highest degree	4.5	4.5	0.00	4.7	4.6	0.07	4.7	4.6	0.04
Mother's occupational prestige	74.5	74.5	0.00	74.9	75.0	0.03	75.2	74.9	0.06
Father's occupational prestige	78.4	78.4	0.01	78.3	78.6	0.05	78.3	78.2	0.01
Birth order	2.0	2.0	0.03	2.0	2.0	0.06	2.0	2.0	0.03
Number of siblings	2.3	2.2	0.04	2.4	2.3	0.05	2.4	2.4	0.02
Liking for school	3.2	3.1	0.03	3.1	3.2	0.11	3.1	3.1	0.02
Liking for math	3.5	3.5	0.02	3.5	3.5	0.07	3.5	3.5	0.08
Doing well in math class	3.0	3.0	0.04	3.0	3.0	0.01	3.0	2.9	0.08
Learning math	1.4	1.4	0.04	1.4	1.4	0.07	1.3	1.3	0.03
Math importance	4.4	4.4	0.03	4.4	4.3	0.07	4.4	4.4	0.05
Previous grades skipped	0.20	0.20	0.00	0.26	0.19	0.17	0.26	0.26	0.00
Proportion Male	0.57	0.57	0.00	0.53	0.53	0.01	0.53	0.53	0.00

Note. ASMD = Absolute Standardized Mean Difference. Results for the mean scores of the grade skippers in the follow-up of Park et al. are not identical to the mean scores of the grade skippers of the newly matched groups because the missing values were imputed using information about the age 50 follow-up.

Table 9: Comparison of the 1976 Cohort Grade Skippers and Newly Matched Control on Background Covariates

	Park et al. (2013)			Follow-up of Park et al. (2013)			Newly matched groups		
	Grade skippers	Matched controls	ASMD	Grade skippers	Matched controls	ASMD	Grade skippers	Matched controls	ASMD
SAT Math	577	567	0.13	556	570	0.20	556	555	0.02
SAT Verbal	482	474	0.11	469	475	0.08	469	468	0.02
Mother's highest degree	4.7	4.6	0.09	4.8	4.7	0.07	4.8	4.8	0.02
Father's highest degree	5.4	5.3	0.06	5.4	5.4	0.07	5.4	5.4	0.07
Number of siblings	1.8	1.7	0.02	1.8	1.7	0.04	1.8	1.8	0.01
Liking for school	3.9	4.0	0.05	3.9	4.0	0.12	3.9	3.9	0.04
Liking for math class	4.4	4.4	0.00	4.3	4.4	0.10	4.3	4.3	0.07
Liking for biology class	3.5	3.5	0.01	3.5	3.6	0.04	3.5	3.5	0.05
Liking for chemistry class	3.9	3.9	0.04	3.8	3.8	0.02	3.8	3.8	0.06
Liking for physics class	3.8	3.8	0.07	3.7	3.7	0.06	3.7	3.7	0.06
Doing well in math class	1.8	1.8	0.06	1.8	1.8	0.01	1.8	1.9	0.06
Doing well in science class	1.9	1.9	0.07	1.9	1.9	0.05	1.9	1.9	0.07
Learning math	1.6	1.5	0.11	1.6	1.5	0.09	1.6	1.5	0.12
Learning science	1.2	1.2	0.05	1.2	1.2	0.02	1.2	1.2	0.04
Math importance	3.6	3.6	0.04	3.6	3.6	0.07	3.6	3.6	0.00
Biology importance	2.4	2.4	0.02	2.4	2.5	0.03	2.5	2.5	0.04
Chemistry importance	2.8	2.8	0.06	2.9	2.8	0.05	2.9	2.8	0.01
Physics importance	3.2	3.1	0.13	3.1	3.1	0.07	3.2	3.1	0.03
Proportion Male	0.7	0.7	0.00	0.7	0.7	0.07	0.7	0.7	0.00
Previous grades skipped	0.2	0.2	0.00	0.2	0.2	0.01	0.2	0.1	0.04
Proportion in public school	0.8	0.8	0.01	0.8	0.8	0.02	0.8	0.8	0.02

Note. ASMD = Absolute Standardized Mean Difference. Results for the mean scores of the grade skippers in the follow-up of Park et al. are not identical to the mean scores of the grade skippers of the newly matched groups because the missing values were imputed using information about the age 50 follow-up.

Well-Being

Results for the well-being outcomes were close to identical to the results of Study 1 Phase 1. Table 10 shows mean scores for grade skippers and newly matched controls. Table 11 present mean scores on well-being outcomes by sex. Figure 9 shows the estimated adjusted standardized mean differences between grade skippers and newly matched controls. Results are close to identical to the results of Study 1 Phase 1. None of the adjusted standardized mean differences is significantly different from zero. Compared to Study 1 Phase 1, the sex pattern (with female grade skippers scoring higher on well being outcomes than female matched controls and the reverse for males) has become less prominent or absent.

To conclude, even after redoing the matching to optimize the equivalence between grade skippers and newly matched controls, no evidence was found for any harmful effects of grade skipping.

Health

The prevalence of severe and common health issues among grade skippers and matched controls is presented in Table 12 and Table 13. These tables are visually presented in Figure 10. This figure is similar to Figure 6. Most health issues cluster around zero. Only emotional, nervous, or psychiatric problems were across cohorts consistently more prevalent among matched controls. Ear, nose, and throat troubles that in Study 1 Phase 1 were more prevalent among grade skippers compared to matched controls now were equally prevalent among both groups.

If we did not control for multiple comparisons, grade skippers form the 1972 cohort had significantly more stomach or intestinal ulcers, and problems with feet and legs. In addition, male grade skippers would have more ear, nose, and throat troubles. None of these differences remained significant when controlling for multiple comparisons.

Figure 11 shows the estimated risk ratios for health issues that were sufficiently prevalent to be included in the logistic regression analyses. The pattern is close to identical to the results of Study 1 Phase 1 (see Figure 7). None of the risk ratios is significantly different from 1, indicating that all health issues were as prevalent among grade skippers and newly matched controls.

Using conditional logistic regression yielded similar results.

Table 10: Mean Scores on Measures of Well-Being for Grade Skippers and Newly Matched Controls

Cohort	Group	N	Positive Affect	Negative Affect (reversed)	Life Satisfaction	Psychological Flourishing	Core Self-Evaluations	Career Satisfaction	Relationship Satisfaction
1972 Cohort	Matched Controls	178	22.4	46.9	25.7	70.3	44.9	5.4	6.2
	Grade Skippers	89	22.6	47.0	24.7	70.0	45.0	5.3	6.2
1976 Cohort	Matched Controls	136	22.2	46.1	25.7	69.5	44.9	5.5	6.3
	Grade Skippers	68	22.0	46.3	25.9	68.7	45.7	5.3	6.1
Both Cohorts	Matched Controls	314	22.3	46.5	25.7	69.9	44.9	5.4	6.3
	Grade Skippers	157	22.4	46.7	25.2	69.5	45.3	5.3	6.2

Table 11: Means on Well-Being Measures by Sex for Grade Skippers and Newly Matched Control

Cohort	Group	N	Positive Affect	Negative Affect (Reversed)	Life Satisfaction	Psychological Flourishing	Core Self-Evaluations	Career Satisfaction	Relationship Satisfactioon
1972 Cohort	Men	141	22.2	47.6	25.1	68.7	44.7	5.3	6.2
	Matched Controls	94	22.1	48.0	25.4	69.2	44.8	5.4	6.2
	Grade Skippers	47	22.4	46.8	24.4	67.8	44.6	5.2	6.2
	Women	126	22.8	46.1	25.7	71.8	45.2	5.4	6.3
	Matched Controls	84	22.8	45.6	26.0	71.5	45.0	5.4	6.3
	Grade Skippers	42	22.8	47.3	25.0	72.5	45.6	5.5	6.2
1976 Cohort	Men	135	21.7	46.4	25.4	68.6	45.4	5.4	6.2
	Matched Controls	90	21.7	46.0	25.4	68.7	45.3	5.5	6.3
	Grade Skippers	45	21.7	47.1	25.4	68.4	45.7	5.3	6.0
	Women	69	22.9	45.7	26.5	70.6	44.7	5.5	6.3
	Matched Controls	46	23.0	46.3	26.2	71.2	44.1	5.6	6.3
	Grade Skippers	23	22.8	44.6	27.0	69.4	45.8	5.4	6.4
Both Cohorts	Men	276	22.0	47.0	25.2	68.7	45.1	5.4	6.2
	Matched Controls	184	21.9	47.0	25.4	68.9	45.0	5.4	6.2
	Grade Skippers	92	22.0	47.0	24.9	68.1	45.1	5.2	6.1
	Women	195	22.8	46.0	25.9	71.4	45.0	5.5	6.3
	Matched Controls	130	22.9	45.8	26.1	71.4	44.7	5.5	6.3
	Grade Skippers	65	22.8	46.4	25.7	71.4	45.6	5.4	6.3

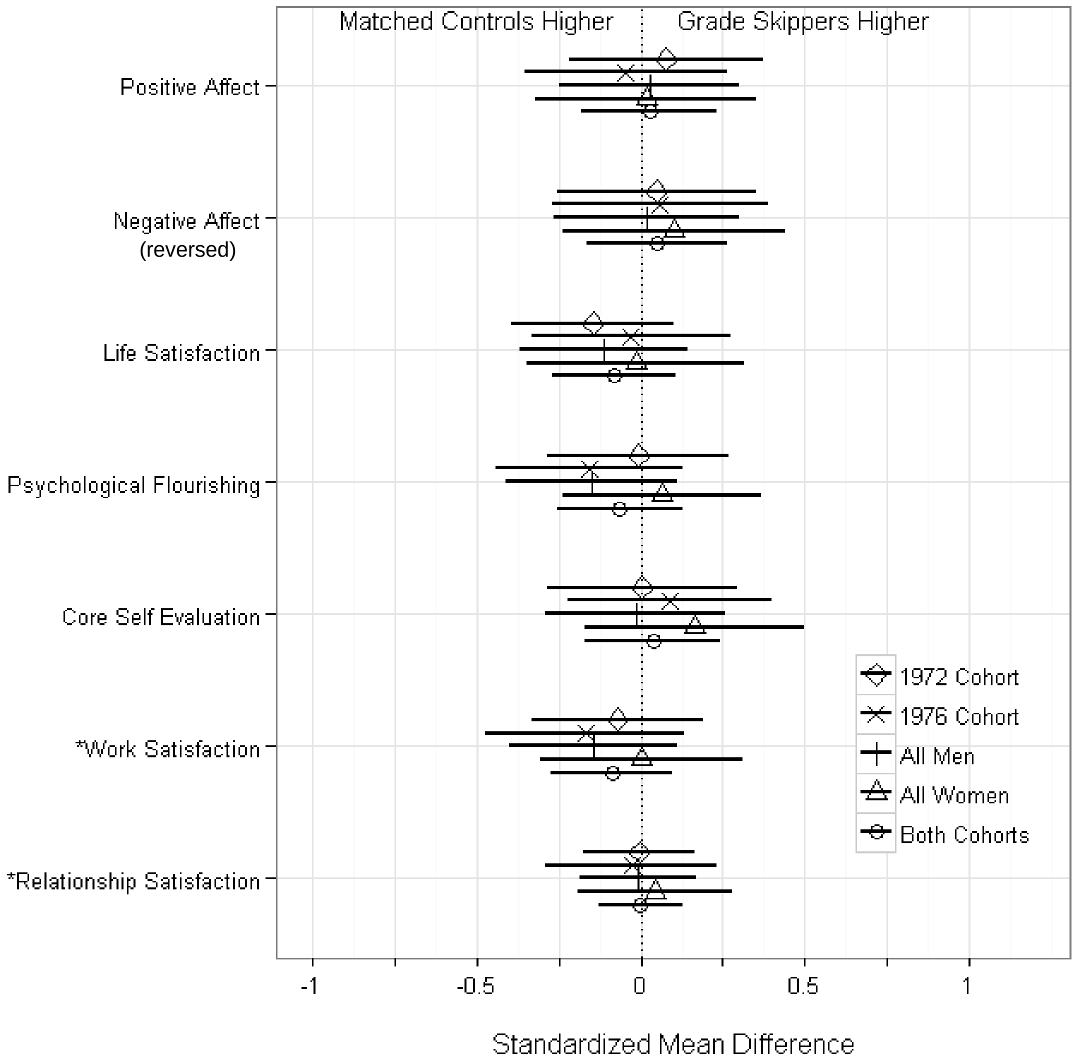

Figure 9: Estimated Adjusted Standardized Mean Differences Between Grade Skippers and Newly Matched Controls for Well-Being Outcomes

Adjusted Standardized Mean Differences between grade skippers and newly matched controls on measures of well-being and Core Self-Evaluations. Point estimates represent regression coefficients for the dichotomous variable that indicates whether the participant skipped at least one grade in high school (receiving a score of 1) or did not skip a grade in high school (receiving a score of 0). For the 1972 Cohort and 1976 Cohort, all background covariates were included as predictors in the linear regression. Because the 1972 and 1976 background covariates were overlapping but not identical, all other analyses only included overlapping background covariates (SAT-M, sex, and previous grades skipped). Horizontal lines represent a 95% confidence interval, not adjusted for multiple comparisons. Shapes represent different samples (see legend). Results are combined across the 10 imputed data sets. Data come from the 1972 Cohort and the 1976 Cohort. * Outcomes with a star represent a single item. For these outcomes regression residuals were not normally distributed, so results must be interpreted with care.

Table 12: Prevalence of Health Issues Across Grade Skippers and Newly Matched Controls

	All respondents	Both Cohorts		1972 Cohort		1976 Cohort		All Men		All Women	
Severe Health Issues		MC	GS	MC	GS	MC	GS	MC	GS	MC	GS
N	1474	313	156	177	88	136	68	184	91	129	65
A heart attack or myocardial infarction?	0.5	0.4	0.6	0.6	0.0	0.0	1.5	0.6	1.1	0.0	0.0
Angina or chest pains due to your heart?	1.2	1.1	0.0	2.0	0.0	0.0	0.0	1.3	0.0	0.9	0.0
Congestive heart failure?	0.2	0.3	0.0	0.5	0.0	0.0	0.0	0.4	0.0	0.0	0.0
A stroke?	0.4	0.4	0.0	0.5	0.0	0.3	0.0	0.4	0.0	0.4	0.0
Not including asthma, but chronic lung disease such as chronic bronchitis or emphysema?	0.7	0.7	0.0	1.2	0.0	0.0	0.0	0.5	0.0	1.0	0.0
High blood pressure or hypertension?	18.1	19.0	17.9	21.4	20.5	15.8	14.7	21.2	18.7	15.7	16.9
Diabetes or high blood sugar?	5.7	5.4	7.7	5.1	9.1	5.7	5.9	5.3	8.8	5.5	6.2
Arthritis or rheumatism?	13.1	14.1	13.5	16.4	20.5	11.0	4.4	9.4	11.0	20.7	16.9
Emotional, nervous, or psychiatric problems?	15.3	18.4	12.2	16.8	11.4	20.6	13.2	14.7	12.1	23.7	12.3
Cancer or malignant tumor of any kind except skin cancer?	3.2	3.4	2.6	4.4	4.5	2.0	0.0	1.4	2.2	6.2	3.1
Other heart problems?	6.4	7.0	5.1	8.6	4.5	4.8	5.9	6.8	6.6	7.2	3.1

Note. MC = Matched Controls, GS = Grade Skippers

Table 13: Common Health Issues among Grade Skippers and Newly Matched Controls

Common Health Issues	All respondents	Both Cohorts MC	Both Cohorts GS	1972 Cohort MC	1972 Cohort GS	1976 Cohort MC	1976 Cohort GS	All Men MC	All Men GS	All Women MC	All Women GS
N	1474	313	156	177	88	136	68	184	91	129	65
Eye trouble, other than glasses or contacts?	8.1	9.2	6.4	9.8	10.2	8.3	1.5	9.5	4.4	8.7	9.2
Ulcer?	1.5	1.3	1.9	1.0	3.4	1.6	0.0	1.3	2.2	1.3	1.5
Severe tooth or gum trouble?	3.6	4.4	4.5	6.4	4.5	1.7	4.4	4.0	5.5	5.0	3.1
Epilepsy or fits?	0.6	0.8	0.6	1.4	1.1	0.1	0.0	1.3	0.0	0.2	1.5
Stomach or intestinal ulcers?	2.2	1.6	3.8	1.5	6.8	1.7	0.0	1.7	3.3	1.5	4.6
Lameness or paralysis (including polio)?	0.7	0.7	2.6	1.2	3.4	0.0	1.5	0.7	2.2	0.7	3.1
Frequent trouble sleeping?	15.8	17.6	18.6	15.1	23.9	21.0	11.8	16.1	16.5	19.8	21.5
Frequent or severe headaches, dizziness or fainting spells?	5.4	5.4	7.7	4.9	10.2	6.2	4.4	3.5	6.6	8.2	9.2
Pain or pressure in your chest, palpitation or pounding heart, or heart trouble?	4.3	4.2	1.3	5.7	2.3	2.2	0.0	2.3	1.1	6.9	1.5
Anemia?	4.3	4.9	3.8	6.4	6.8	2.9	0.0	0.8	1.1	10.7	7.7
Swollen or painful joints, frequent cramps in your legs or bursitis?	6.3	5.4	7.1	6.2	10.2	4.4	2.9	2.0	4.4	10.4	10.8
Problems with your feet and legs?	11.3	10.7	13.5	10.5	20.5	10.9	4.4	7.4	14.3	15.3	12.3
Neuritis?	1.1	1.2	1.9	1.1	2.3	1.4	1.5	0.4	3.3	2.4	0.0
Asthma? (Shortness of breath or chronic cough?)	7.0	8.1	7.1	8.4	10.2	7.6	2.9	6.4	9.9	10.4	3.1
Depression or excessive worry or nervous trouble of any kind?	17.3	19.5	17.9	18.2	19.3	21.1	16.2	17.3	15.4	22.6	21.5
Kidney or bladder problems?	3.5	3.7	5.1	5.0	8.0	1.9	1.5	4.2	7.7	2.9	1.5
Hardening of the arteries?	0.5	0.4	0.0	0.8	0.0	0.0	0.0	0.6	0.0	0.2	0.0
Frequent urinary tract infections? (other than kidney problems discussed earlier?)	1.1	1.3	1.9	2.1	2.3	0.2	1.5	0.1	0.0	3.0	4.6
Scarlet fever, rheumatic fever, tuberculosis, jaundice or hepatitis?	0.8	0.7	0.6	0.7	1.1	0.7	0.0	0.4	0.0	1.1	1.5
Problems with your back?	23.5	21.7	24.4	23.0	30.7	20.0	16.2	21.3	20.9	22.2	29.2
Osteoporosis?	1.0	1.1	0.0	1.6	0.0	0.3	0.0	0.0	0.0	2.6	0.0
Frequent indigestion, stomach, liver or intestinal trouble, gall bladder trouble or gallstones?	9.6	11.6	10.3	11.2	13.6	12.3	5.9	8.7	8.8	15.9	12.3
Painful or "trick" shoulder or elbow, "trick" or locked knee?	9.4	9.5	5.1	11.0	3.4	7.5	7.4	9.9	4.4	8.9	6.2
Ear, nose, or throat trouble?	6.9	7.5	10.3	8.2	11.4	6.4	8.8	7.4	15.4	7.5	3.1
Low blood pressure?	2.6	2.2	1.9	2.1	2.3	2.4	1.5	0.9	1.1	4.1	3.1
Skin disease?	5.7	5.3	5.8	4.9	6.8	5.8	4.4	4.2	5.5	6.7	6.2
Chronic or frequent colds, sinus problems, hay fever or allergies?	21.4	22.4	24.4	23.0	20.5	21.7	29.4	18.3	25.3	28.2	23.1
Adverse or allergic reaction to any serum, drug or medicine?	13.2	15.2	11.5	16.3	17.0	13.7	4.4	10.1	7.7	22.4	16.9
Bone, joint or other deformity?	2.6	2.5	3.2	3.7	3.4	1.0	2.9	2.7	2.2	2.2	4.6
High cholesterol?	21.7	22.5	21.2	24.1	21.6	20.4	20.6	27.6	28.6	15.2	10.8
Thyroid trouble or goiter?	6.5	5.8	7.7	7.3	10.2	3.8	4.4	1.8	6.6	11.5	9.2
Tumor, growth, or cyst?	5.6	5.5	4.5	5.4	3.4	5.6	5.9	4.2	5.5	7.2	3.1
Loss of finger or toe?	0.3	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0

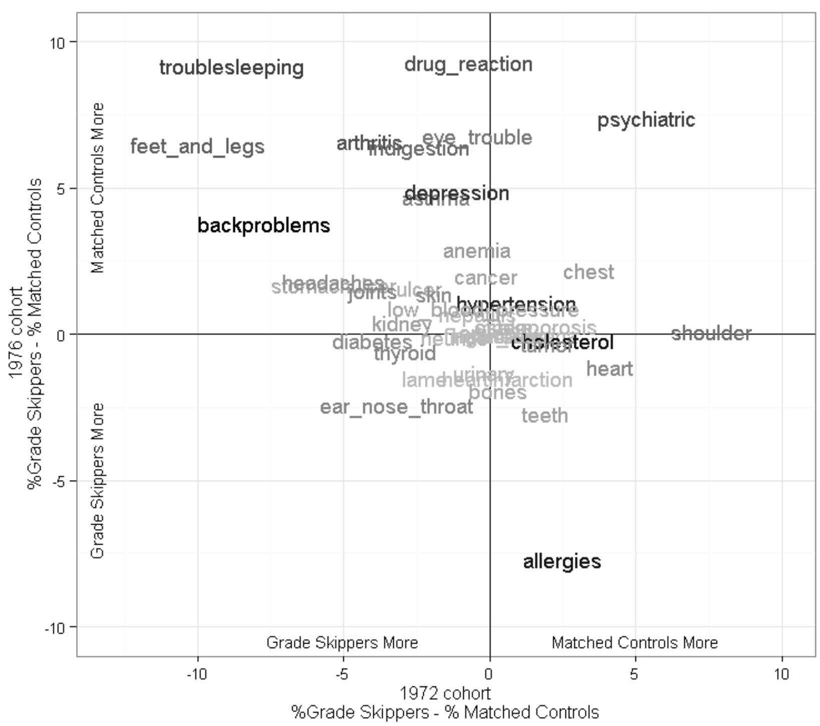

Figure 10: Differences in Percentage Endorsement of Health Items Between Grade Skippers and Newly Matched Controls

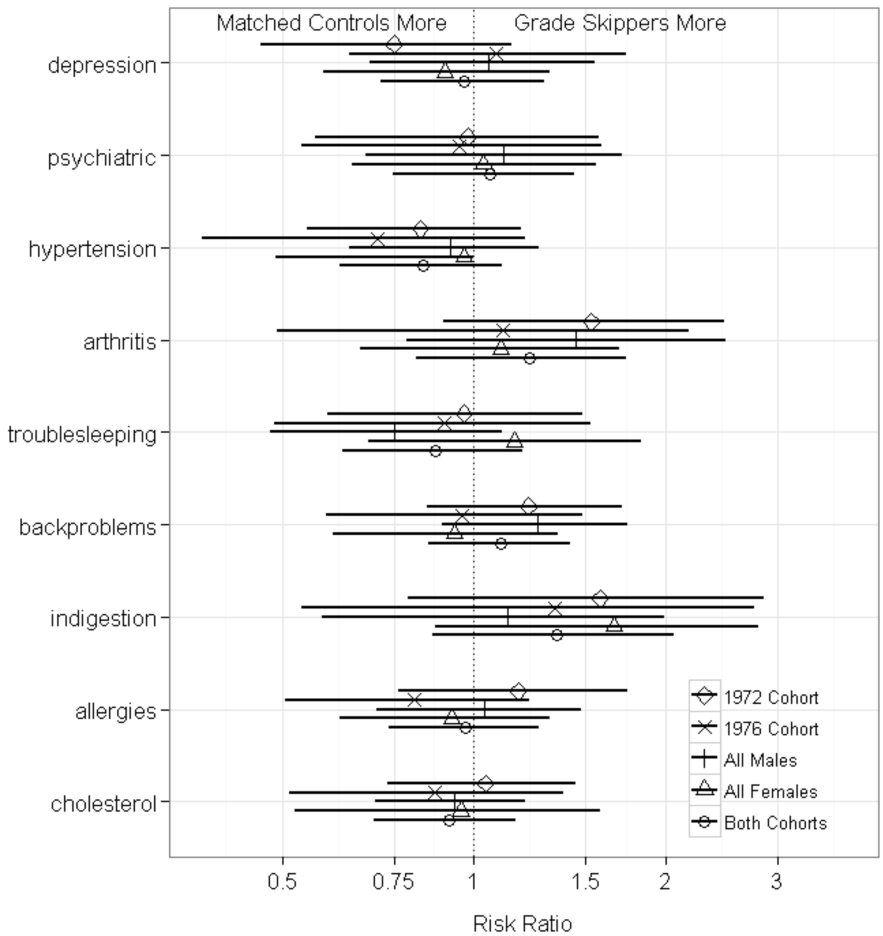

Figure 11: Estimated Risk Ratios for Selected Health Issues Comparing Grade Skippers and Newly Matched Controls

Point estimates represent logistic regression coefficients for the dichotomous variable that indicates whether the participant skipped at least one grade in high school (receiving a score of 1) or did not skip a grade in high school (receiving a score of 0). For the 1972 Cohort and 1976 Cohort, all background covariates were included as predictors in the logistic regression. Because background covariates were different for the 1972 and 1976 Cohorts, other analyses only include background covariates they have in common. Horizontal lines represent a 95% confidence interval. Shapes represent different samples (see legend). Results are combined across the 10 imputed data sets. Data come from the 1972 Cohort and the 1976 Cohort.

Discussion and limitations

This study evaluated the association between skipping one or more grades in high school and well-being 35 years later in the top 1% in mathematical ability. It tests the hypothesis that skipping one or more grades in high school among intellectually talented youth results in psychological maladjustment or distress at age 50.

Overall participants in this study report experiencing positive emotions frequently. They were satisfied with their lives, themselves, their work and their relationships. Compared to a nationally representative sample, they did not report having more severe health issues. When the proportions differed, they almost always favored the current sample. Only one item (emotional, nervous, or psychiatric problems) was endorsed more often in the current sample. Because this observation is inconsistent with the on average higher scores on all but one of the well-being scales, a cautious interpretation is warranted.

In a first phase, the grade skippers and matched controls of Park et al. (2013) were surveyed. No significant differences were found in positive affect, negative affect, life satisfaction, psychological flourishing, core self evaluations, career satisfaction and relationship satisfaction between grade skippers and controls matched on background covariates including abilities, sex, number of grades skipped prior to high school, motivation, and SES. Even after optimizing the equivalence among grade skippers and controls by redoing the propensity score matching in a second phase, the results remained the same. No evidence was found for any harmful effect of grade skipping on well-being at age 50. In addition, across both phases of Study 1, no evidence was found for differences in the prevalence of physical and mental health issues (including emotional, nervous or psychiatric problems, depression, and hypertension). To the contrary, emotional, nervous, and psychiatric problems were consistently more prevalent among matched controls.

Although this study meets most of the quality criteria as specified by Cornell et al. (1991), it is limited in some ways. First, like all research that uses matching, the validity of the findings depends on the assumption that all confounding variables were included in the observed background covariates. Although the study included many purported confounding variables in the matching procedure (including abilities, motivation, SES), matching covariates did not include, for example, measures of personality traits. Extraversion may be positively related to both the likelihood of skipping a grade (extraverts may be perceived as socially more mature) and well-being (DeNeve & Cooper, 1998; Steel et al., 2008). In addition, spatial ability was not included in the matching procedure.

Second, as in most longitudinal studies, the high amount of attrition may have caused bias. It could be that grade skippers that are high in well-being are more likely to respond to the survey than grade-skippers low in well-being, hence introducing bias. However, by using matching, the risk of bias due to differential attrition among grade skippers and matched controls (related to any of the background variables included in the matching procedure) was reduced.

Despite these limitations, the current study was the first study to look at social and emotional adjustment in the top 1% in mathematical ability 35 years after they skipped one or more grades in high school. In addition, it was the first study that investigated the effects of grade skipping on social and emotional adjustment using propensity score matching.

Future studies could investigate how the effects of grade skipping are moderated by the characteristics of the student, the characteristics of the environment, and their interaction. At what age is grade skipping most effective? What is the minimum ability level a student is required to have to benefit from grade skipping? How can teachers and parents facilitate the transition period immediately after skipping a grade? In addition, more fine-grained, observational research is needed about the social dynamics of entering a new class-room with older peers.

Implications of this study will be further discussed in the general discussion at the end.

CHAPTER V

STUDY 2: THE EFFECTS OF EDUCATIONAL DOSE IN HIGH SCHOOL ON WELL-BEING AT AGE 50

Introduction

Educational dose is defined as "the density of advanced and enriching pre-collegiate learning opportunities beyond the norm that students have participated in" (Wai et al., 2010, p. 861). It is based on the idea that the exact combination of accelerated interventions (advanced placement, subject matter acceleration, or taking college courses while in high school) students receive is not important, as long as they receive enough interventions and opportunities that are intellectually stimulating.

Using cohorts 1 through 3 and 5 of SMPY, Wai et al. (2010) found that STEM educational dose (accelerated educational interventions limited to Science, Technology, Engineering, and Mathematics; STEM) was positively associated with STEM accomplishments, including earning a STEM PhD, having a STEM publication, having a STEM tenure track faculty position, and having a STEM occupation. This study evaluated the association between STEM educational dose and well-being at age 50. The hypothesis that subject-based acceleration among intellectually talented youth results in psychological maladjustment or distress at age 50 was tested.

In the first phase, the high and low STEM educational dose groups of cohorts 1 and 2 (Wai et al., 2010) were followed-up and compared on their well-being at age 50. In a second phase, the propensity score framework was extended to encompass continuous treatments. Matching participants on the generalized propensity score made it possible to control for confounding background variables such as ability, SES, and motivation when evaluating the association between well-being and educational dose.

The first section introduces the generalized propensity score. The second section describes the participants and the measures. The last section reviews and discusses the results.

Generalized Propensity Score

Imai and van Dijck (2004) propose an extension of the propensity score method to encompass non-binary treatments. In our second study, the treatment variable, educational dose, is a continuous treatment. The following notation and reasoning is based on Imai and van Dijck (2004). For continuous treatments, the set of potential outcomes becomes

$$\{Y_i(t^p), t^p \epsilon\, T \; for \; i = 1, \dots, n\} \tag{17}$$

where T is a set of potential treatment values, and $Y_i(t^p)$ is a random variable that maps a particular potential treatment t^p to a potential outcome. The number of potential outcomes is equal to the number of potential treatments. The average treatment effect becomes

$$E\{Y(t_1^p) - Y(t_2^p)|X\} = E\{Y(t_1^p)|T^A = t_1^p, X\} - E\{Y(t_2^p)|T^A = t_2^p, X\} \tag{18}$$

where $t_1^p \neq t_2^p$, and X is a matrix of confounding background covariates. As is the case for a binary treatment, confounding covariates must be controlled for. The way the propensity score has been defined until now (the probability of receiving a dichotomous treatment) would not work in the case of a continuous treatment. Therefore, the propensity score must be generalized to encompass a continuous treatment. This can be done using the propensity function, defined as "the conditional probability of the actual treatment given the observed covariates" (Imai et al., 2004, p. 856),

$$e_\varphi(T^A|X) \tag{19}$$

where φ parameterizes the distribution. The dimensionality of X can be reduced by specifying a function $\theta_\varphi(X)$ that is parameterized by θ, for which holds that the propensity function depends on X only through θ, yielding

$$e_\varphi\{T^A|X, \theta_\varphi(X)\} = e_\varphi\{T^A|\theta_\varphi(X)\} = e_\varphi\{T^A|\theta\} \tag{20}$$

Applied to our continuous treatment 'educational dose', the assumption can be made that the conditional distribution of the treatment given the covariates is Gaussian, yielding

$$T^A|X \sim N(\mathrm{X}^\mathrm{T}\beta, \sigma^2) \tag{21}$$

where σ^2 is a scalar and β is a $(p \times 1)$ vector with p being the number of background covariates included in the analyses. The propensity function $e_\varphi\{T^A|\theta_\varphi(X)\}$ is the Gaussian density function and $\theta_\varphi(X) = \mathrm{X}^\mathrm{T}\beta$. Assuming that σ^2remains equal for all values of X, this function is uniquely characterized by the scalar θ, representing the mean expected value of the treatment given the covariates.

In practice, the following steps can be followed.

First, a model is specified that predicts the treatment dose from the background variables, or $p_\varphi(T^A|X)$. Second, the estimated model parameters $\hat{\varphi}$ are used to obtain the expected treatment for each participant, $\hat{\theta} = \theta_{\hat{\varphi}}(X)$. Third, subclasses of equal size are created based on $\hat{\theta}$. Fourth, a parametric model $p_\omega\{Y_i(t^p)|T^A = t^p, \theta\}$ is fitted to each subclass, with ω the unknown parameter. Fifth, the distribution of potential outcomes is computed by taking the weighted average over all subclasses, with weights being the relative size of the subclasses, or

$$p\{Y_i(t^p)\} = \int p\{Y_i(t^p)|T^A = t^p, \theta\}p(\theta)d\theta \quad (22)$$

$$\approx \sum_{j=1}^{J} p_{\widehat{\omega}_j}\{Y_i(t^p)|T^A = t^p\}W_j \quad (23)$$

where $\widehat{\omega}_j$ is the estimate of the model parameter of subclass j and W_j is the relative weight of subclass j. Formula (23) provides us with the distribution of the potential outcomes. In this study this distribution is summarized by the causal effect that can be directly estimated using

$$\approx \sum_{j=1}^{J} \widehat{\omega}_j\{Y_i(t^p)|T^A = t^p, X\}W_j \quad (24)$$

where $\widehat{\omega}_j\{Y_i(t^p)|T^A = t^p, X\}$ is the estimated model parameter of subclass j and W_j is estimated as the relative proportion of the observations included in subclass j. As suggested by Imai and van Dyck (2004) and Robins and Rotnitzky (2001), background covariates (X) are included as covariates to account for the within class variability.

Propensity function sub-classification reduces bias, improves efficiency of a parametric model and is more robust to model misspecifications than linear regression without preprocessing of the data (Imai et al., 2004).

Generalized propensity score methodology has been used in psychological science (e.g., Bachman, Staff, O'Malley, Schulenberg, & Freedman-Doan, 2011; Han, Miller, & Waldfogel, 2010), medicine (e.g., Feng, Zhou, Zou, Fan, & Li, 2012; Moodie & Stephens, 2012; Rowan et al., 2012), criminology (e.g., Snodgrass, Blokland, Haviland, Nieuwbeerta, & Nagin, 2011), economy (e.g., Becker, Luedtke, Trautwein, Koeller, & Baumert, 2012; Flores, Flores-Lagunes, Gonzalez, & Neuman, 2012; Kluve, Schneider, Uhlendorff, & Zhao, 2012;), and ecology (e.g., Yuan, 2010).

Participants

Participants were drawn from the 1972 and 1976 Cohorts (see Study 1) who scored at or above 500 on the SAT-math subtest were included, or the top .5% in mathematical reasoning ability. This data sample was identical to the 1972 and 1976 Cohorts of Wai et al. (2010). Table 14 gives an overview of the proportion of participants that have completed the age 50 follow-up survey.

Table 14: Number of Participants Who Constituted the High and Low STEM Educational Dose Groups of Wai et al. (2010) and Have Completed the Age 50 Follow-Up Survey

	High Dose			Low Dose		
	Wai et al. (2010)	Age 50 follow up	Proportion	Wai et al. (2010)	Age 50 follow up	Proportion
1972 Cohort	341	246	0.72	435	271	0.62
Men	249	173	0.69	269	156	0.58
Women	92	73	0.79	166	115	0.69
1976 Cohort	262	192	0.73	205	138	0.67
Men	200	145	0.72	141	94	0.67
Women	62	47	0.76	64	44	0.69
Both Cohorts	603	438	0.73	640	409	0.64
All Men	449	318	0.71	410	250	0.61
All Women	154	120	0.78	230	159	0.69

Measures

STEM educational dose was operationalized in two ways. In the first phase of the study, we closely followed the procedure used by Wai et al. (2010), resulting in the SED1. In the second phase of the study, some refinements to the operationalization of SED1 were suggested, resulting in SED2. The next section describes the difference between SED1 and SED2.

STEM Educational Dose 1 (SED1; Wai et al., 2010)

Wai et al. (2010) operationalized STEM educational dose as the number of different types of accelerative STEM educational interventions a student received. Following variables constituted the SED1.

Participants of the 1972 Cohort were asked following questions at age 18 (unless noted otherwise):

<u>Special academic training:</u>

From the age 13 survey: This school year, how are you learning most of your arithmetic and mathematics? Check only one. a) In regular class work, with other students, b) In school, but working on your own with some help or direction from your teacher, c) On your own outside of school, helped by a tutor or parent, d) On your own outside of school with little help from anyone. Options b, c, and d were considered special academic training.

From the age 23 survey: As a child before talent search participation, did you receive any special academic training in science or mathematics from your parents, relatives, other adults, school, or other?

College courses while in high school.

List all the courses you took for credit at a college before becoming a full-time college student, as well as the name of the institution, the year you took the course, the grade you were in at the time, the final (overall) grade you received in the course, and the number of credits.

Advanced Placement.

List in the appropriate spaces below the exact name and level (such as, Calculus AB or BC, or Physics C Mechanics) of all Advanced Placement Program (APP) examinations you have taken. (Omit those subjects for which you took APP courses but did not take the APP exams.) Show the year(s) you took the exam(s) and the school grade(s) you were in at the time.

Science fair projects and math competitions

List all the national, regional, or state mathematics contests in which you have competed. Please indicate which contest, your score, and awards you received.

List all the science fair projects you submitted to science fairs in your school, state, region, or nation. Please indicate the title of the project, science area (e.g., biology, chemistry, physics), year, the school grade you were in at the time, and any prizes you received.

For each of the four categories, (special academic training, college courses, AP, science fair project/competitions), a participant was awarded a score of 1 if he/she participated in at least one of the STEM-related interventions included in the respective category. Otherwise, the participant received a score of 0 for that category. To calculate the SED1, the scores for all categories were summed. For example, a participant who took one AP course in STEM, and participated in one regional science competition would receive a SED1 of 2. A participant who took four AP courses in STEM, participated in two science contests and two science fairs, would also have an SED1 of 2.

Participants of the 1976 Cohort were asked following questions at age 18:

Special academic training:

From the age 13 survey: This school year, how are you learning most of your science/arithmetic and mathematics? Check only one. a) In regular class work, with other students, b) In school, but working on your own with some help or direction from your teacher, c) On your own outside of school, helped by a tutor or parent, d) On your own outside of school with little help from anyone. Options b, c, and d were considered special academic training.

Check each of the following statements that apply to you. Work on science projects as part of other group work such as scouting, or on your own.

From the age 23 survey: As a child before talent search participation, did you receive any special academic training in science or mathematics from your parents, relatives, other adults, school, or other?

College courses while in high school

Please list all college courses you took on a part-time basis while a high school student.

Advanced Placement

Have you taken any Advanced Placement Program (AP) courses or examinations? If 'Yes', indicate which ones, giving course title, level (e.g., Math AB or BC), course grade, score on the exam if taken, and the grade you were in when taken.

Contests

Have you participated in any mathematics or science contests, or been awarded entry to a special honorary program (such as NSF workshop)?

Special classes (age 13 questionnaire):

Check each of the following statements that apply to you. Have taken special courses or participated in programs given at places other than your regular school (such as the Maryland Academy of Sciences).

Check each of the following statements that apply to you. Participated in the Maryland Summer Program for the Gifted and Talented.

Check each of the following statements that apply to you. Participated in science fairs.

Research/invention and projects

Indicate Which of the Following You Have Accomplished. (Do Not Include Normal Class Assignments): a) created own invention/process, b) worked on special project(s) in mathematics/science, c) contributed importantly to a research project

A similar procedure as in the 1972 cohort was used to calculate the SED1. Because there are six categories for the 1976 cohort, the SED1 ranged from 0 to 6.

STEM Educational Dose 1I

Building forth on the concept of STEM educational dose as described by Wai et al. (2010), the following refinements are proposed.

First, only interventions that took place after the talent search were included in the educational dose variable. The confounding background covariates included in the propensity score estimation must be measured prior to the onset of treatment (Imai et al., 2004; Ho et al., 2007). Because the treatment could influence some of the background covariates (e.g., abilities as measured by the SAT), including background covariates measured after the onset of treatment can introduce bias. Some of the intervention components of Wai et al. (2010) were administered before or during the talent search.

Second, Wai et al., (2010) operationalized educational dose as the number of different types of STEM educational interventions a student received, without taking into account the frequency of these interventions. Whether a student took one or 10 STEM AP courses is not reflected in their educational dose score; both would count for one educational dose unit (EDU). In the current study, a value of one EDU is assigned to every STEM AP course taken. The same approach is used for STEM college courses taken while being in high school, number of competitions, and number of science fairs.

For the 1972 Cohort, this results in an SED2 that does not include the special academic training category because these interventions happened before the assessment of the background variables. In addition, when calculating the SED2, the frequency of educational interventions will be taken into account. For example, a participant who took four AP courses in STEM, participated in two science contests and two science fairs, would have an SED1 of 2 , but an SED2 of 8 (4+2+2).

For the 1976 Cohort, the SED2 did not include the special academic training and special classes categories. Both reflected interventions from before the assessment of the background variables. In addition, it did not include the "research/inventions and projects" categories. The subjective nature of these items makes straightforward interpretations of the responses difficult. For example, participants may have different judgment of what constitutes an "invention" or what it means to "contribute importantly to a research project". Last, similar to the 1972 Cohort, when calculating the SED2, the frequency of educational interventions will be taken into account.

Background covariates

Background covariates used in the estimation of the generalized propensity score were the same as in Study 1.

Outcome measures

Outcome measures were the same as in Study 1.

Results Study 2 Phase 1

In the first phase, the high and low STEM educational dose groups of Wai et al. (2010) are followed-up and compared on their well-being at age 50. MacCallum, Zhang, Preacher, and Rucker (2002) show that median split leads to a loss of information about individual differences, loss of effect size and power, and loss of measurement reliability. In case of small effect sizes and small samples, median split can lead to increased estimates of effect size due to sampling error. However, in the case of Wai et al. (2010), median split was used because of the small within cohort sample sizes, the low base rates of the STEM outcome criteria, and the heterogeneity in the rigor of the educational interventions included in their STEM educational dose variable. Results from Study 2 Phase 1 must be interpreted with this in mind. Study 2 Phase 2 addresses these limitations.

Missing Data

Multiple imputation was used to handle missing data (see Study 1)

Evaluating Balance

Not all participants that constituted the high and low STEM educational dose groups of Wai et al. (2010) completed the age 50 survey. To investigate whether high and low SED1 groups had been differentially affected by attrition, these groups were compared on important background covariates. A comparison of the high and low STEM educational groups on important background covariates for the 1972 and 1976 cohorts can be found in Table 15 and 16, respectively.

The absolute standardized mean differences are presented in Figure 12 and 13. For the 1972 cohort, there were large differences between the high and low dose groups on variables including SAT-Math, parent's highest degree, learning math, previous grades skipped and the proportion male. For the 1976 cohort there were large differences on variables including SAT-Math, previous grades skipped, learning math, number of siblings. These groups were not equivalent. A crude comparison of means on well-being or health variables must be interpreted with care. Differences can be caused by either differences in STEM educational dose, or differences in any of the background covariates that show a difference between high and low STEM educational dose participants.

Well-Being

Table 17 compares high and low STEM educational dose groups on measures of positive affect, negative affect, life satisfaction psychological flourishing, core self-evaluations, career satisfaction, and relationship satisfaction. As in study 1, both groups on average endorsed positive adjustment on all of these outcomes. Table 27 shows that the sample in this study scores comparable to the normative samples used in the test construction and validation, with all but one of the outcomes (Core Self-Evaluations scale) being higher for the sample in this study.

Table 18 shows the same outcomes separated by sex. The 1976 cohort shows a consistent pattern of the high dose group outperforming the low dose group on all but one of the well-being outcomes, for both males and females. The results for the 1972 Cohort are mixed, with no clear pattern emerging. Figure 14 shows the adjusted standardized mean differences. Even after adjusting for confounding background covariates, the adjusted SMD is positive for all but one outcome (relationship satisfaction in the 1976 cohort). Limiting the analyses to only males or females results in larger differences for females on positive affect and negative affect, and larger differences for males on life satisfaction and psychological flourishing, all favoring grade skippers.

Table 15: Comparing Participants form Wai et al.'s High and Low STEM Educational Dose Groups that Completed the Age 50 Follow-Up on Important Background Covariates for the 1972 Cohort.

	Wai et al. (2010)			Age 50 follow-up		
	High Dose	Low Dose	ASMD	High Dose	Low Dose	ASMD
N	341	435		246	271	
SAT Math	59.4	56.7	0.47	59.3	56.7	0.46
Mother's highest degree	3.7	3.2	0.36	3.8	3.3	0.35
Father's highest degree	4.9	4.3	0.29	4.9	4.4	0.30
Mother's occupational prestige	75.1	74.3	0.17	75.4	74.5	0.16
Father's occupational prestige	79.8	77.5	0.30	79.8	78.3	0.22
Birth order	2.0	2.1	0.12	2.0	2.2	0.16
Number of siblings	2.3	2.5	0.10	2.4	2.5	0.05
Liking for school	3.2	3.1	0.20	3.2	3.1	0.19
Liking for math	3.5	3.4	0.15	3.5	3.5	0.11
Doing well in math class	3.1	2.9	0.25	3.1	2.9	0.21
Learning math	1.5	1.2	0.59	1.5	1.1	0.62
Math importance	4.4	4.4	0.06	4.4	4.4	0.04
Previous grades skipped	0.18	0.06	0.36	0.17	0.06	0.32
Proportion Male	0.73	0.62	0.24	0.70	0.57	0.27

Note. High & Low Dose = High & Low STEM Educational Dose, ASMD = Absolute Standardized Mean Difference

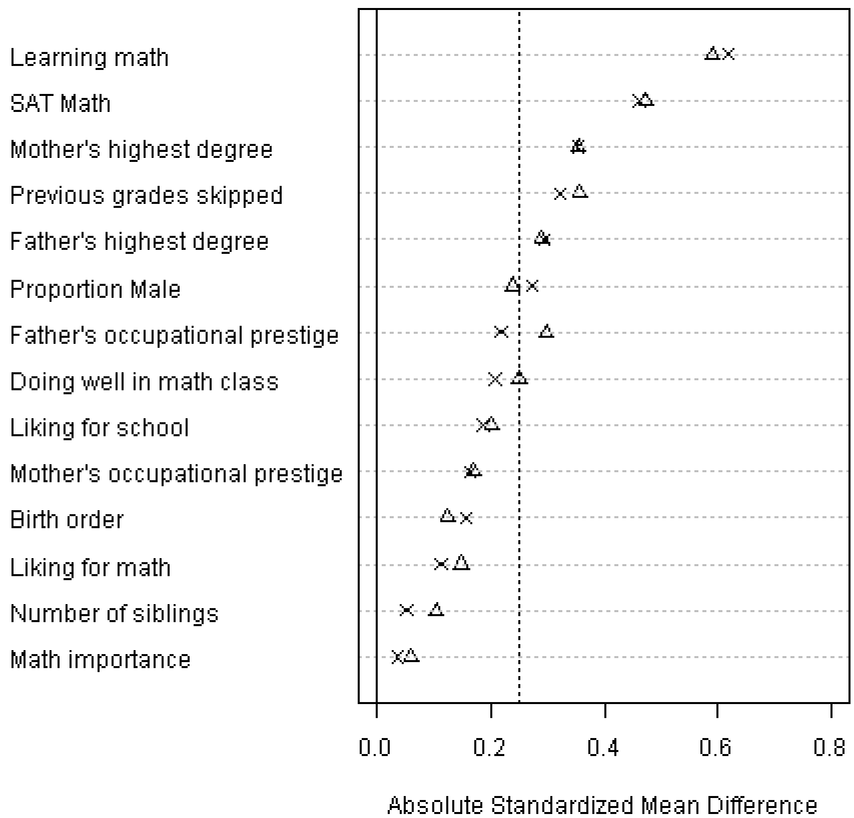

Figure 12: Absolute Standardized Mean Differences Between High and Low STEM Educational Dose Groups of the 1972 Cohort

The absolute standardized mean differences between the low (n = 435) and high (n = 341) STEM educational dose groups of Wai et al. (2010; triangles), and for low (n = 271) and high (n = 246) STEM educational dose groups that completed the age 50 survey (x's) on important background covariates. The dotted vertical line indicates the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (2009). Data come from the 1972 Cohort.

Table 16: Comparing Participants form Wai et al.'s High and Low STEM Educational Dose Groups that Completed the Age 50 Follow-Up on Important Background Covariates for the 1976 Cohort.

	Wai et al. (2010)			Age 50 follow-up		
	High Dose	Low Dose	ASMD	High Dose	Low Dose	ASMD
N	262	205		192	138	
SAT Math	581	557	0.45	578	557	0.41
SAT Verbal	459	463	0.06	454	461	0.10
Mother's highest degree	4.8	4.6	0.16	4.8	4.6	0.21
Father's highest degree	5.5	5.3	0.12	5.5	5.3	0.17
Number of siblings	1.7	1.9	0.17	1.6	2.1	0.32
Liking for school	4.0	4.0	0.01	4.0	4.0	0.06
Liking for math class	4.5	4.4	0.19	4.5	4.3	0.22
Liking for biology class	3.6	3.6	0.02	3.6	3.6	0.01
Liking for chemistry class	3.9	3.8	0.13	3.9	3.8	0.17
Liking for physics class	3.9	3.7	0.23	3.8	3.6	0.22
Doing well in math class	3.2	3.1	0.18	3.2	3.1	0.26
Doing well in science class	3.1	2.9	0.26	3.1	2.9	0.31
Learning math	1.5	1.2	0.38	1.5	1.2	0.32
Learning science	1.3	1.2	0.19	1.2	1.1	0.21
Math importance	3.6	3.6	0.10	3.6	3.5	0.08
Biology importance	2.6	2.6	0.06	2.6	2.7	0.12
Chemistry importance	2.8	2.8	0.00	2.8	2.8	0.08
Physics importance	3.0	2.9	0.09	3.0	2.9	0.08
Proportion Male	0.76	0.69	0.17	0.76	0.68	0.16
Previous grades skipped	0.24	0.07	0.39	0.20	0.07	0.31
Proportion in public school	0.82	0.76	0.16	0.81	0.77	0.10

Note. High & Low Dose = High & Low STEM Educational Dose, ASMD = Absolute Standardized Mean Difference

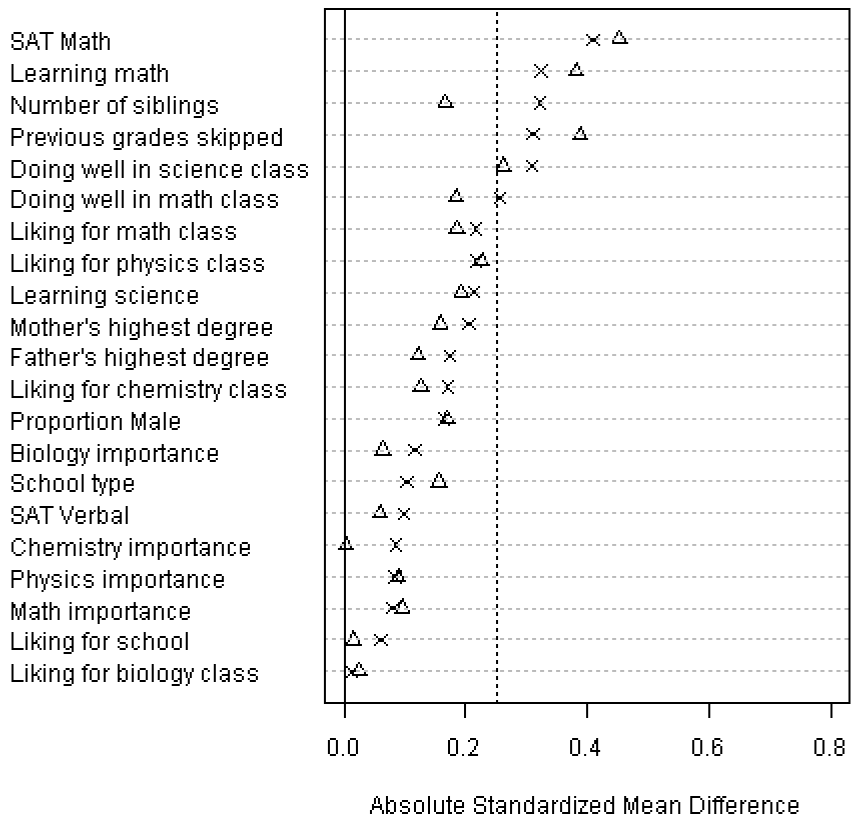

Figure 13: Absolute Standardized Mean Differences Between High and Low STEM Educational Dose Groups of the 1976 Cohort

The absolute standardized mean differences between the low (*n* = 205) and high (*n* = 262) STEM educational dose groups of Wai et al. (2010; triangles), and for low (*n* = 138) and high (*n* = 192) STEM educational dose groups that completed the age 50 survey (x's) on important background covariates. The dotted vertical line indicates the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (2009). Data come from the 1976 Cohort.

Table 17: Comparing High and Low STEM Educational Dose Groups from Wai et al. (2010) Who Completed the Age 50 Follow-Up on Measures of Well-Being

Cohort	Group	N	Positive Affect	Negative Affect (reversed)	Life Satisfaction	Psychological Flourishing	Core Self-Evaluations	Career Satisfaction	Relationship Satisfaction
1972 Cohort	Low SED	271	22.4	47.3	25.7	70.5	45.6	5.5	6.1
	High SED	246	22.3	47.3	25.7	70.0	45.2	5.4	6.3
1976 Cohort	Low SED	138	21.5	44.7	24.6	67.8	43.6	5.3	6.1
	High SED	192	22.5	47.2	26.6	70.7	45.7	5.5	6.2
Both Cohorts	Low SED	409	22.1	46.4	25.3	69.6	44.9	5.4	6.1
	High SED	438	22.4	47.3	26.1	70.3	45.4	5.5	6.2

Note. Low SED = Low STEM Educational Dose I, High SED = High STEM Educational Dose I

Table 18: Comparing High and Low STEM Educational Dose Groups from Wai et al. (2010) Who Completed the Age 50 Follow-Up on Measures of Well-Being By Sex

Cohort	Group	N	Positive Affect	Negative Affect (Reversed)	Life Satisfaction	Psychological Flourishing	Core Self-Evaluations	Career Satisfaction	Relationship Satisfactioon
1972 Cohort	Men	329	22.2	48.5	25.4	69.2	45.7	5.4	6.2
	Low Dose	156	22.4	49.0	25.1	69.4	46.0	5.3	6.0
	High Dose	173	22.1	48.1	25.6	69.1	45.4	5.4	6.3
	Women	188	22.6	45.2	26.2	72.1	44.9	5.5	6.1
	Low Dose	115	22.4	45.1	26.4	72.1	45.2	5.7	6.1
	High Dose	73	23.0	45.3	25.9	72.0	44.5	5.3	6.2
1976 Cohort	Men	239	21.8	46.8	25.7	69.1	45.1	5.4	6.2
	Low Dose	94	21.3	45.8	24.8	67.4	44.1	5.4	6.2
	High Dose	145	22.1	47.4	26.3	70.3	45.8	5.5	6.1
	Women	91	22.8	44.4	25.9	70.5	44.0	5.4	6.2
	Low Dose	44	21.8	42.1	24.1	68.7	42.6	5.2	6.0
	High Dose	47	23.7	46.5	27.6	72.2	45.4	5.6	6.3
Both Cohorts	Men	568	22.1	47.8	25.5	69.2	45.4	5.4	6.2
	Low Dose	250	22.0	47.8	25.0	68.6	45.2	5.4	6.1
	High Dose	318	22.1	47.8	25.9	69.7	45.6	5.5	6.2
	Women	279	22.7	44.9	26.1	71.6	44.6	5.5	6.1
	Low Dose	159	22.2	44.3	25.7	71.2	44.4	5.5	6.1
	High Dose	120	23.3	45.8	26.5	72.1	44.8	5.5	6.2

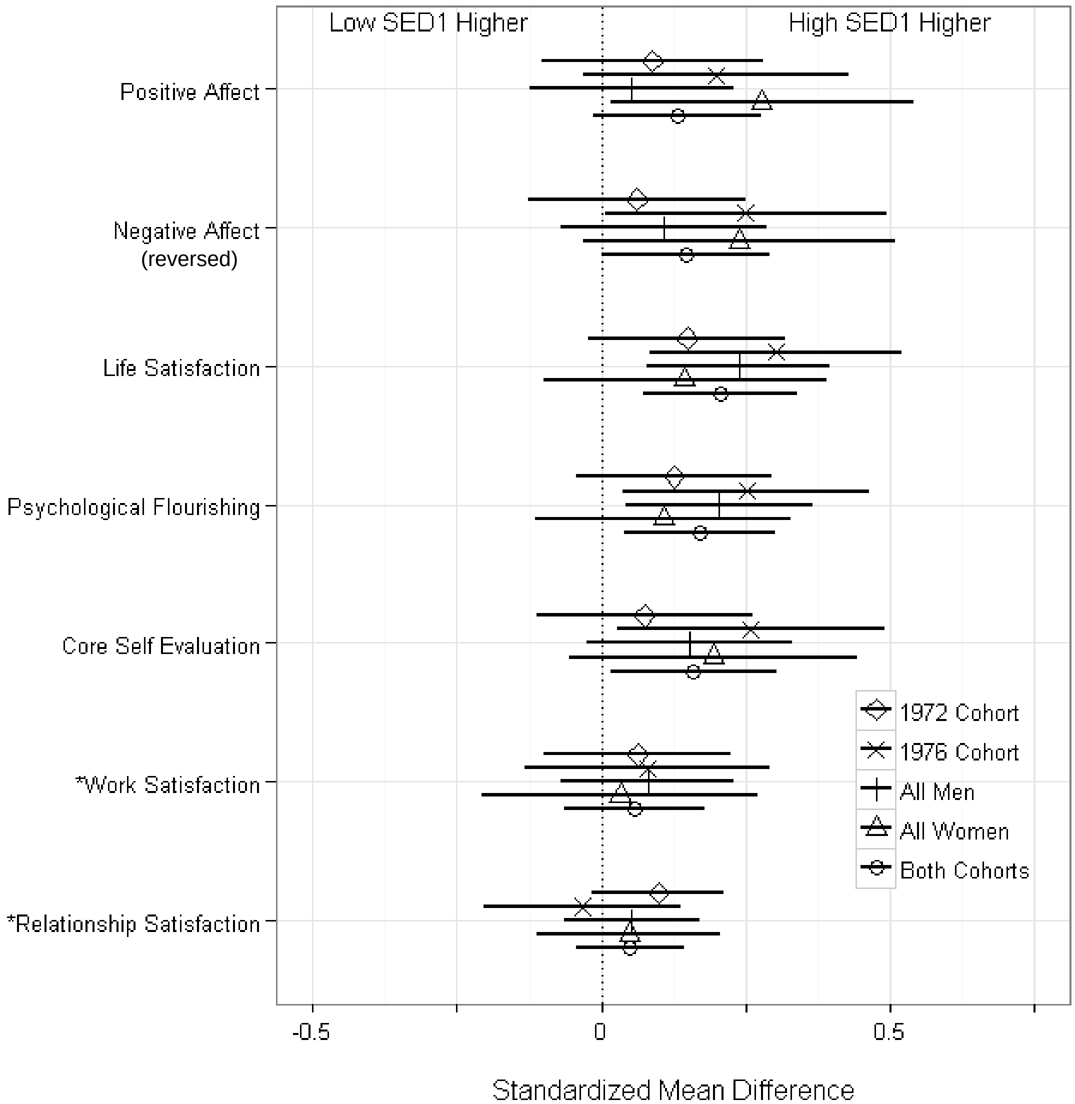

Figure 14: Adjusted Standardized Mean Differences Between High and Low STEM Educational Dose (SED1) Groups on Measures of Well-Being and Core Self-Evaluations

Point estimates represent regression coefficients for the dichotomous variable that indicates whether the participant received a high SED (receiving a score of 1) or a low SED (receiving a score of 0). For the 1972 Cohort and 1976 Cohort, all background covariates were included as predictors in the linear regression. Because background covariates were different for the 1972 and 1976 Cohorts, other analyses only included background covariates that both cohorts have in common. Horizontal lines represent a 95% confidence interval. Shapes represent different samples (see legend). Results were combined across the 10 imputed data sets. Data come from the 1972 and the 1976 Cohorts. * Outcomes with a star represent a single item. For these outcomes, regression residuals for these outcomes were not normally distributed, therefore results must be interpreted with care.

To conclude, no evidence was found for harmful effects of STEM educational dose on well-being at age 50. To the contrary, a higher STEM educational dose level seems to be associated with higher well-being at age 50, although the size of association was small.

Health

Table 19 and 20 give an overview of the prevalence of severe and common health issues among high and low STEM Educational Dose participants from Wai et al. (2010) who completed the age 50 follow-up. Overall, participants in this study tend to have less severe physical health issues than a nationally representative sample (see Table 28). However, they report more emotional, nervous, or psychiatric problems. Similar to study 1, these comparisons must be interpreted with care (see p. 41).

To explore consistent patterns between the 1972 and 1976 cohorts, the difference in prevalence (in percentages) between high and low STEM educational dose groups for each health issue are plotted in Figure 15. The x-axis represents the difference between high and low dose groups in the 1972 cohort, with higher values representing higher prevalence among the low STEM educational dose participants in both cohorts. The same is done for the 1976 on the y-axis. The color of the health issues represents their base rate prevalence, with grey indicating low prevalence and black indicating high prevalence. Most health issues cluster around zero on both axes, indicating that these health issues are equally common among High STEM Educational Dose and Low STEM Educational Dose participants. Emotional, nervous, and psychiatric problems are more prevalent among low dose than among high dose participants. Back problems and adverse or allergic reaction to any serum, drug or medicine seem to be more prevalent among high dose participants. For other health issues, differences were inconsistent or absent across cohorts. All differences were small. Without controlling for multiple comparisons, a tumor, growth, or cyst was more likely among high dose participants from the 1972 cohort and all female high dose participants combined. The 1976 high dose participants were more likely to report thyroid trouble or goiter. When controlling for multiple comparisons, none of these differences were significant.

A visual presentation of the adjusted risk ratios can be found in Figure 16. Without controlling for multiple comparisons, only one risk ratio was statistically significant: among all men, high STEM Educational Dose participants had more back problems. Emotional, nervous, and psychiatric problems, hypertension, and cholesterol were more common among low STEM Educational Dose participants. Arthritis and back problems were somewhat more common among high STEM Educational Dose participants (except for back problems for all women combined). But none of the differences were statistically significant.

To conclude, no evidence was found for harmful effects of high levels of STEM Educational Dose on severe and common health issues at age 50. To the contrary: most health issues seem equally likely or somewhat less likely among high STEM Educational Dose participants than among low STEM Educational Dose participants.

Table 19: Prevalence of Severe Health Issues among High and Low STEM Educational Dose Participants of Wai et al. (2010) who Completed the Age 50 Follow-Up

	All respondents	Both Cohorts		1972 Cohort		1976 Cohort		All Men		All Women	
Severe Health Issues		LD	HD	LD	HD	LD	HD	LD	HD	LD	HD
N	1474	409	438	271	246	138	192	250	318	159	120
A heart attack or myocardial infarction?	0.5	0.5	0.5	0.4	0.8	0.7	0.0	0.8	0.6	0.0	0.0
Angina or chest pains due to your heart?	1.2	1.2	0.7	1.9	1.2	0.0	0.0	1.6	1.0	0.6	0.0
Congestive heart failure?	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A stroke?	0.4	1.0	0.0	1.5	0.0	0.0	0.0	1.6	0.0	0.0	0.0
Not including asthma, but chronic lung disease such as chronic bronchitis or emphysema?	0.7	0.7	0.2	1.1	0.4	0.0	0.0	0.4	0.3	1.3	0.0
High blood pressure or hypertension?	18.1	19.6	16.7	20.5	19.3	17.8	13.2	19.9	17.5	19.1	14.4
Diabetes or high blood sugar?	5.7	6.2	5.8	6.3	6.6	5.9	4.8	5.7	6.1	7.0	5.1
Arthritis or rheumatism?	13.1	12.9	11.8	14.2	13.6	10.4	9.5	8.9	8.6	19.1	20.3
Emotional, nervous, or psychiatric problems?	15.3	17.9	13.9	16.0	11.9	21.5	16.4	13.4	12.1	24.8	18.6
Cancer or malignant tumor of any kind except skin cancer?	3.2	2.5	2.8	2.6	2.9	2.2	2.6	2.0	1.6	3.2	5.9
Other heart problems?	6.4	7.2	5.8	7.8	5.8	5.9	5.8	7.7	6.1	6.4	5.1

Note. HD = High STEM Educational Dose, LD = Low STEM Educational Dose.

Table 20: Common Health Issues among High and Low STEM Educational Dose Participants of Wai et al. (2010) Who Completed the Age 50 Follow-Up.

	All respondents	Both Cohorts		1972 Cohort		1976 Cohort		All Men		All Women	
Common Health Issues		LD	HD	LD	HD	LD	HD	LD	HD	LD	HD
N	1474	1474	409	438	271	246	138	192	250	318	159
Eye trouble, other than glasses or contacts?	8.1	6.5	7.6	7.1	7.4	5.2	7.9	4.5	8.0	9.6	6.8
Ulcer?	1.5	1.2	1.4	1.5	1.2	0.7	1.6	1.6	1.3	0.6	1.7
Severe tooth or gum trouble?	3.6	3.2	3.5	3.0	5.3	3.7	1.1	4.1	3.5	1.9	3.4
Epilepsy or fits?	0.6	0.7	0.2	1.1	0.4	0.0	0.0	1.2	0.3	0.0	0.0
Stomach or intestinal ulcers?	2.2	2.0	1.4	2.6	1.2	0.7	1.6	2.0	1.3	1.9	1.7
Lameness or paralysis (including polio)?	0.7	0.7	0.9	0.7	1.6	0.7	0.0	0.4	1.0	1.3	0.8
Frequent trouble sleeping?	15.8	14.9	15.5	14.2	16.9	16.3	13.8	12.6	15.0	18.5	16.9
Frequent or severe headaches, dizziness or fainting spells?	5.4	5.2	6.0	5.6	6.6	4.4	5.3	2.4	4.8	9.6	9.3
Pain or pressure in your chest, palpitation or pounding heart, or heart trouble?	4.3	4.2	2.5	5.2	2.5	2.2	2.6	2.4	2.5	7.0	2.5
Anemia?	4.3	4.5	3.5	4.5	4.5	4.4	2.1	0.8	1.0	10.2	10.2
Swollen or painful joints, frequent cramps in your legs or bursitis?	6.3	6.0	6.0	7.5	7.0	3.0	4.8	4.1	4.8	8.9	9.3
Problems with your feet and legs?	11.3	11.7	12.0	13.8	14.8	7.4	8.5	8.9	10.8	15.9	15.3
Neuritis?	1.1	0.5	1.6	0.4	1.6	0.7	1.6	0.4	1.3	0.6	2.5
Asthma? (Shortness of breath or chronic cough?)	7.0	5.7	6.5	4.9	6.2	7.4	6.9	4.5	5.7	7.6	8.5
Depression or excessive worry or nervous trouble of any kind?	17.3	19.1	17.4	16.8	16.0	23.7	19.0	14.2	15.6	26.8	22.0
Kidney or bladder problems?	3.5	4.0	4.6	5.6	6.6	0.7	2.1	2.4	5.1	6.4	3.4
Hardening of the arteries?	0.5	0.5	0.2	0.7	0.4	0.0	0.0	0.8	0.3	0.0	0.0
Frequent urinary tract infections? (other than kidney problems discussed earlier?)	1.1	1.0	0.9	1.5	1.6	0.0	0.0	0.4	0.3	1.9	2.5
Scarlet fever, rheumatic fever, tuberculosis, jaundice or hepatitis?	0.8	1.0	0.9	1.1	1.2	0.7	0.5	0.4	0.6	1.9	1.7
Problems with your back?	23.5	21.1	24.5	21.6	26.7	20.0	21.7	18.7	24.2	24.8	25.4
Osteoporosis?	1.0	1.5	0.7	1.5	0.8	1.5	0.5	0.8	0.0	2.5	2.5
Frequent indigestion, stomach, liver or intestinal trouble, gall bladder trouble or gallstones?	9.6	10.2	8.8	9.7	9.5	11.1	7.9	8.5	6.1	12.7	16.1
Painful or "trick" shoulder or elbow, "trick" or locked knee?	9.4	11.2	9.5	12.7	7.4	8.1	12.2	10.6	8.6	12.1	11.9
Ear, nose, or throat trouble?	6.9	6.5	8.3	6.3	9.1	6.7	7.4	7.7	8.0	4.5	9.3
Low blood pressure?	2.6	2.5	1.4	3.0	0.8	1.5	2.1	1.2	0.6	4.5	3.4
Skin disease?	5.7	5.7	5.8	5.2	6.2	6.7	5.3	4.9	4.1	7.0	10.2
Chronic or frequent colds, sinus problems, hay fever or allergies?	21.4	21.1	20.6	19.0	19.3	25.2	22.2	15.9	18.8	29.3	25.4
Adverse or allergic reaction to any serum, drug or medicine?	13.2	11.7	13.9	14.2	15.6	6.7	11.6	7.3	9.9	18.5	24.6
Bone, joint or other deformity?	2.6	3.0	2.5	3.7	2.9	1.5	2.1	2.4	2.2	3.8	3.4
High cholesterol?	21.7	24.3	22.5	25.7	25.1	21.5	19.0	28.9	26.8	17.2	11.0
Thyroid trouble or goiter?	6.5	6.9	6.7	10.1	7.0	0.7	6.3	1.2	3.5	15.9	15.3
Tumor, growth, or cyst?	5.6	3.7	6.5	3.0	7.4	5.2	5.3	3.7	4.5	3.8	11.9
Loss of finger or toe?	0.3	0.5	0.0	0.4	0.0	0.7	0.0	0.8	0.0	0.0	0.0

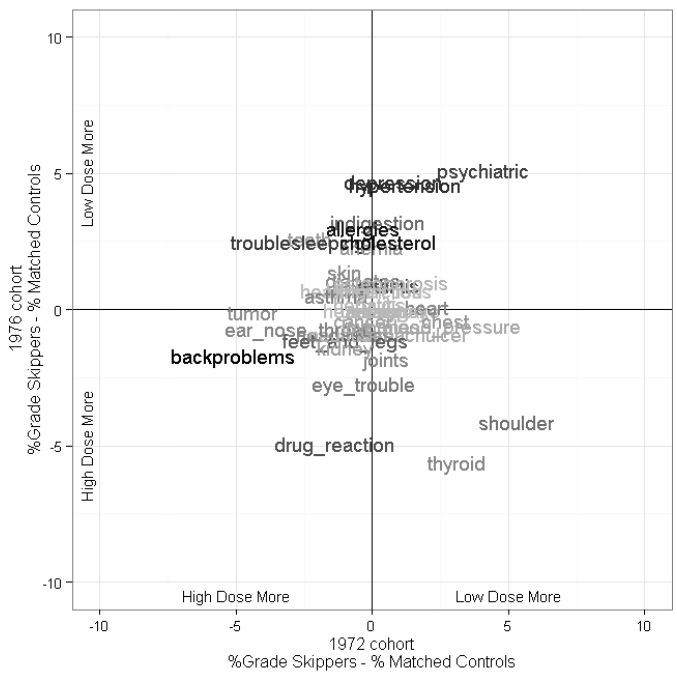

Figure 15: Differences in Percentage Endorsement of Health Items Between High and Low STEM Educational Dose Groups from Wai et al (2010) that Completed the Age 50 Survey for the 1972 and the 1976 Cohorts.

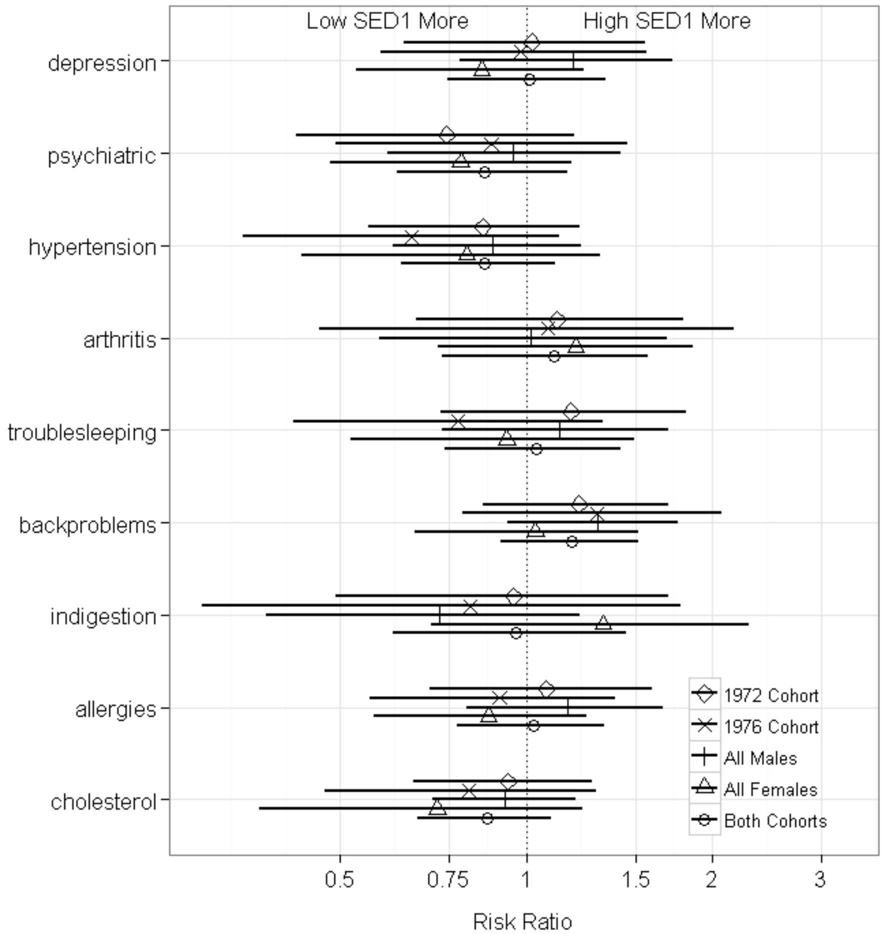

Figure 16: Adjusted Risk Ratios for High and Low STEM Educational Dose Participants from Wai et al. (2010) Who Completed the Age 50 Follow-Up

Estimated risk ratios for selected health issues comparing High and Low STEM Educational Dose (SED) groups. Point estimates represent regression coefficients for the dichotomous variable that indicates whether the participant was above the median in SED1 (receiving a score of 1) or below the median of SED1 (receiving a score of 0). For the 1972 Cohort and 1976 Cohort, all background covariates were included as predictors in the logistic regression. Because background covariates were different for the 1972 and 1976 Cohorts, other analyses only include background covariates that both cohorts have in common. Horizontal lines represent a 95% confidence interval. Shapes represent different samples (see legend). Results are averaged over the 10 imputed data sets. Data come from the 1972 and the 1976 Cohorts.

Results Study 2 Phase 2

In a second phase, the association between the continuous predictors SED1 and SED2, and measures related to well-being and health at age 50 was evaluated using generalized linear regression models with generalized propensity score sub-classification.

Calculating the Generalized Propensity Score

STEM educational dose is a count variable: it can only take on non-negative discrete values (0, 1, 2, 3, etc.). For SED1, a regular Poisson regression was used to estimate the generalized propensity scores with SED1 as the outcome and all background covariates as the predictors. The SED2 exhibited an excess of zeros. Therefore Zero-Inflated Poisson regression was used to estimate the generalized propensity score for the SED2. The generalized propensity score is equal to the predicted level of either SED1 or SED2. Participants were divided in subclasses of equal sample size based on their generalized propensity scores. More classes will lead to more homogeneity within each class, but also reduced sample size and power (Imai et al., 2004). According to Rosenbaum (2002), five classes are sufficient for most data sets. This study presents results for 2, 3, and 5 subclasses. To obtain final estimates, a weighted average is taken across subclasses of regression coefficients and their standard errors.

Checking Balance

Balance was checked using standard normal quantile plots of the *t*-statistics that regress educational dose on each background covariate while controlling for the propensity score (see Figure 17; see also Figures 28, 29, & 30). Given the propensity score, educational dose and the background covariates should be independent. All figures show absence of a relation between educational dose and the background covariates after controlling for the propensity score, providing evidence for the validity of our propensity score model.

Well-Being

<u>SED1</u>

Table 21 show the resulting regression coefficient estimates for 1, 2, 3, and 5 subclasses for SED1. Subclasses were created to have approximately equal sample size across subclasses. Estimates are not affected much by the number of subclasses used, but the SE of the estimates increases with the number of subclasses. This is because the sample size per subclass decreases as the number of subclasses increases. The third column of Table 21 shows estimates and SE without subclasses. These are regular linear regression results. Most estimates are small, and the majority is positive. Estimates are consistently positive across samples for positive affect, life satisfaction, and relationship satisfaction. For females, the estimates are negative for psychological flourishing, core self-evaluations, and career satisfaction. Without controlling for multiple comparisons, only the regression coefficient for relationship satisfaction in the

1972 cohort is significantly different from zero, in a positive direction. Figure 18 gives a visual representation of these estimates without subclassification. Figure 20 shows estimates when using 3 subclasses. The pattern of the estimates across outcomes and samples is similar. Using the generalized propensity score subclassification does not meaningfully alter the estimates. This provides evidence that the assumptions of linearity of background covariates and well-being outcomes is met. There is no evidence that complex interactions between background covariates and the outcomes have biased the results.

SED2

Table 22 shows the linear regression coefficients and standard errors for SED2 without subclasses, and results averaged across 2, 3, and 5 subclasses. Subclasses were created to have approximately an equal sample size across subclasses. The results for a model without subclasses are visually presented in Figure 19. Compared to Figure 18, standard errors are smaller. Dichotomization in the calculation of the SED1 has decreased the variance compared to the SED2. All estimates are small, and almost all are positive. Only the coefficients for psychological flourishing and career satisfaction for all women are negative, but they are small and not significant. Without controlling for multiple comparisons, only the coefficient for SED2 in predicting positive affect for the 1976 cohort is significant. When averaging over 2, 3, or 5 subclasses, the pattern of results remains similar. The majority of coefficients are positive but small. Some coefficients estimated for all females become negative, but they are small and non-significant. Figure 21 visually presents the results averaged across 3 subclasses. The pattern of estimates remains similar, but the confidence intervals are wider. Most of the estimates are positive and none of them is significantly different from zero.

To conclude, no evidence was found for any harmful effects of STEM Educational Dose on Well-Being at age 50. To the contrary, most effect seem to be positive and favor participants that received a higher STEM Educational Dose. Effects were small and only few estimates were significantly different from zero.

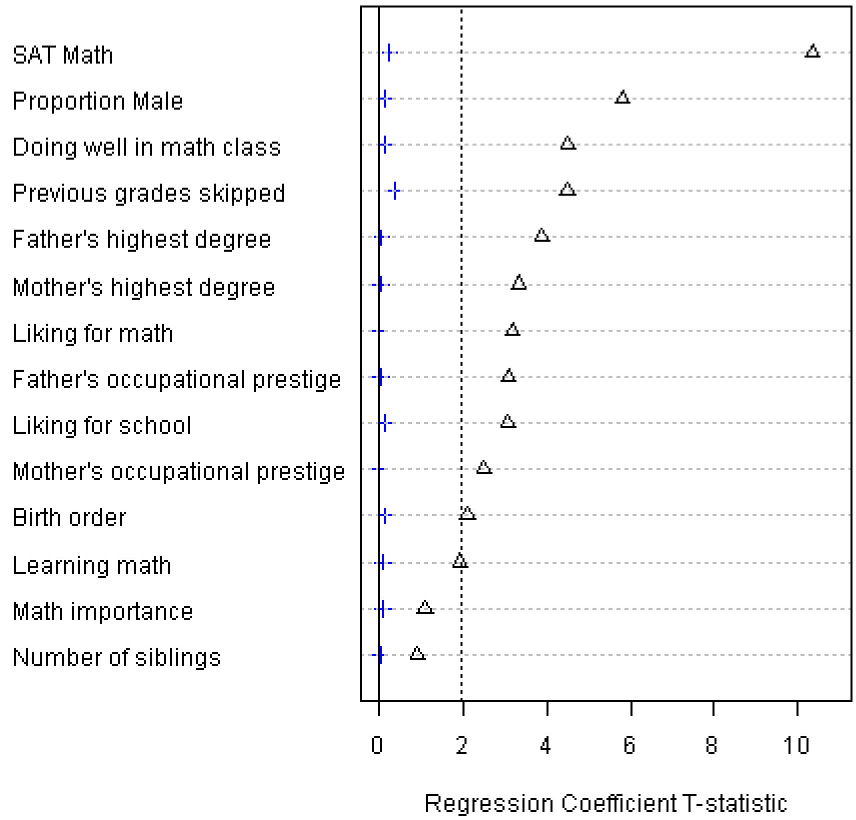

Figure 17: Evaluating Balance Using the Generalized Propensity Score for SED2 in the 1972 Cohort

Separate regression models with the background covariates as the outcomes were constructed. Triangles show the t-statistic for the regression coefficient of SED2 as the only predictor, crosses show the t-statistic for the regression coefficient of SED2 when the generalized propensity score is added as a predictor.

Table 21: Averaged Regression Coefficients and Standard Errors for SED1 for 1, 2, 3, and 5 Subclasses

		No Subclasses		2 Subclasses		3 Subclasses		5 Subclasses	
Sample	Outcome	Est.	SE	Est.	SE	Est.	SE	Est.	SE
1972 Cohort	Positive Affect	0.01	(0.052)	0.01	(0.075)	0.01	(0.094)	–0.01	(0.122)
	Negative Affect	–0.01	(0.050)	–0.01	(0.073)	–0.02	(0.091)	–0.03	(0.123)
	Life Satisfaction	0.04	(0.051)	0.04	(0.075)	0.04	(0.093)	0.02	(0.123)
	Psychological Flourishing	–0.01	(0.050)	0.00	(0.072)	–0.01	(0.091)	–0.03	(0.119)
	Core Self Evaluations	0.00	(0.050)	0.00	(0.072)	0.00	(0.092)	–0.02	(0.121)
	Career Satisfaction	–0.04	(0.051)	–0.04	(0.075)	–0.03	(0.094)	–0.05	(0.126)
	Relationship Satisfaction	0.12	(0.052)*	0.12	(0.074)	0.13	(0.092)	0.12	(0.122)
1976 Cohort	Positive Affect	0.06	(0.051)	0.05	(0.076)	0.06	(0.099)	0.02	(0.139)
	Negative Affect	0.06	(0.052)	0.05	(0.077)	0.07	(0.103)	0.04	(0.147)
	Life Satisfaction	0.08	(0.050)	0.08	(0.075)	0.10	(0.098)	0.05	(0.137)
	Psychological Flourishing	0.04	(0.051)	0.04	(0.074)	0.04	(0.098)	0.00	(0.139)
	Core Self Evaluations	0.04	(0.051)	0.04	(0.077)	0.06	(0.101)	0.03	(0.144)
	Career Satisfaction	–0.01	(0.052)	0.00	(0.078)	0.01	(0.104)	–0.01	(0.147)
	Relationship Satisfaction	0.00	(0.056)	–0.02	(0.079)	–0.02	(0.105)	–0.04	(0.145)
All Men	Positive Affect	0.00	(0.040)	0.00	(0.058)	0.00	(0.071)	–0.01	(0.094)
	Negative Affect	0.00	(0.039)	0.00	(0.056)	0.00	(0.070)	0.00	(0.092)
	Life Satisfaction	0.06	(0.039)	0.07	(0.058)	0.07	(0.071)	0.06	(0.093)
	Psychological Flourishing	0.03	(0.039)	0.04	(0.057)	0.04	(0.071)	0.03	(0.092)
	Core Self Evaluations	0.01	(0.039)	0.01	(0.057)	0.01	(0.070)	0.01	(0.092)
	Career Satisfaction	–0.01	(0.040)	–0.02	(0.059)	–0.02	(0.073)	–0.02	(0.095)
	Relationship Satisfaction	0.05	(0.041)	0.06	(0.059)	0.07	(0.072)	0.07	(0.093)
All Women	Positive Affect	0.09	(0.067)	0.07	(0.099)	0.06	(0.125)	0.04	(0.172)
	Negative Affect	0.06	(0.067)	0.05	(0.098)	0.04	(0.124)	0.02	(0.172)
	Life Satisfaction	0.01	(0.068)	0.00	(0.100)	0.00	(0.127)	–0.02	(0.173)
	Psychological Flourishing	–0.08	(0.067)	–0.08	(0.096)	–0.10	(0.124)	–0.12	(0.169)
	Core Self Evaluations	0.04	(0.066)	0.04	(0.097)	0.02	(0.123)	0.00	(0.169)
	Career Satisfaction	–0.06	(0.068)	–0.05	(0.098)	–0.07	(0.124)	–0.11	(0.170)
	Relationship Satisfaction	0.08	(0.069)	0.09	(0.100)	0.09	(0.126)	0.08	(0.177)
Both Cohorts	Positive Affect	0.03	(0.033)	0.02	(0.048)	0.03	(0.060)	0.01	(0.079)
	Negative Affect	0.01	(0.033)	0.01	(0.048)	0.02	(0.059)	0.01	(0.078)
	Life Satisfaction	0.05	(0.033)	0.05	(0.049)	0.05	(0.061)	0.05	(0.080)
	Psychological Flourishing	0.00	(0.033)	0.00	(0.049)	0.01	(0.059)	–0.01	(0.079)
	Core Self Evaluations	0.01	(0.033)	0.02	(0.048)	0.02	(0.059)	0.02	(0.079)
	Career Satisfaction	–0.03	(0.034)	–0.03	(0.050)	–0.03	(0.061)	–0.03	(0.081)
	Relationship Satisfaction	0.06	(0.035)	0.07	(0.050)	0.07	(0.062)	0.07	(0.081)

Table 22: Averaged Regression Coefficients and Standard Errors for SED2 for 1, 2, 3, and 5 Subclasses.

		No Subclasses		2 Subclasses		3 Subclasses		5 Subclasses	
Sample	Outcome	Est.	SE	Est.	SE	Est.	SE	Est.	SE
1972 Cohort	Positive Affect	0.01	(0.022)	0.01	(0.033)	0.01	(0.043)	0.01	(0.056)
	Negative Affect	0.01	(0.021)	0.02	(0.033)	0.02	(0.043)	0.03	(0.057)
	Life Satisfaction	0.01	(0.022)	0.02	(0.034)	0.01	(0.043)	0.01	(0.057)
	Psychological Flourishing	0.01	(0.022)	0.00	(0.032)	0.00	(0.041)	0.00	(0.055)
	Core Self Evaluations	0.02	(0.022)	0.03	(0.033)	0.02	(0.043)	0.02	(0.056)
	Career Satisfaction	0.00	(0.022)	0.00	(0.034)	0.00	(0.044)	0.00	(0.059)
	Relationship Satisfaction	0.02	(0.022)	0.03	(0.034)	0.03	(0.043)	0.04	(0.058)
1976 Cohort	Positive Affect	0.05	(0.021)*	0.05	(0.030)	0.05	(0.040)	0.04	(0.058)
	Negative Affect	0.04	(0.021)	0.04	(0.031)	0.04	(0.042)	0.02	(0.060)
	Life Satisfaction	0.04	(0.021)	0.04	(0.029)	0.04	(0.038)	0.02	(0.056)
	Psychological Flourishing	0.03	(0.021)	0.03	(0.030)	0.03	(0.040)	0.02	(0.056)
	Core Self Evaluations	0.04	(0.021)	0.04	(0.030)	0.05	(0.040)	0.03	(0.057)
	Career Satisfaction	0.01	(0.021)	0.01	(0.031)	0.02	(0.041)	0.00	(0.058)
	Relationship Satisfaction	0.00	(0.023)	0.00	(0.032)	–0.01	(0.042)	–0.02	(0.062)
All Men	Positive Affect	0.02	(0.016)	0.02	(0.023)	0.02	(0.029)	0.01	(0.037)
	Negative Affect	0.01	(0.016)	0.01	(0.023)	0.01	(0.029)	0.01	(0.038)
	Life Satisfaction	0.03	(0.016)	0.03	(0.023)	0.03	(0.029)	0.03	(0.038)
	Psychological Flourishing	0.03	(0.016)	0.02	(0.023)	0.02	(0.029)	0.02	(0.038)
	Core Self Evaluations	0.02	(0.016)	0.03	(0.023)	0.03	(0.029)	0.03	(0.037)
	Career Satisfaction	0.02	(0.016)	0.02	(0.024)	0.01	(0.029)	0.02	(0.038)
	Relationship Satisfaction	0.01	(0.017)	0.01	(0.024)	0.01	(0.029)	0.01	(0.038)
All Women	Positive Affect	0.04	(0.030)	0.02	(0.047)	0.02	(0.058)	0.00	(0.077)
	Negative Affect	0.03	(0.030)	0.01	(0.049)	0.02	(0.059)	0.00	(0.079)
	Life Satisfaction	0.01	(0.030)	–0.02	(0.047)	–0.01	(0.059)	–0.03	(0.078)
	Psychological Flourishing	–0.03	(0.030)	–0.06	(0.046)	–0.05	(0.058)	–0.07	(0.076)
	Core Self Evaluations	–0.01	(0.030)	–0.01	(0.047)	–0.01	(0.058)	–0.02	(0.078)
	Career Satisfaction	–0.03	(0.030)	–0.03	(0.047)	–0.03	(0.059)	–0.04	(0.078)
	Relationship Satisfaction	0.00	(0.031)	0.02	(0.048)	0.02	(0.060)	0.02	(0.079)
Both Cohorts	Positive Affect	0.03	(0.014)	0.02	(0.021)	0.01	(0.027)	0.01	(0.035)
	Negative Affect	0.01	(0.014)	0.02	(0.021)	0.01	(0.026)	0.01	(0.035)
	Life Satisfaction	0.02	(0.014)	0.03	(0.021)	0.02	(0.027)	0.01	(0.035)
	Psychological Flourishing	0.01	(0.014)	0.01	(0.021)	0.00	(0.026)	0.00	(0.034)
	Core Self Evaluations	0.01	(0.014)	0.02	(0.021)	0.01	(0.027)	0.01	(0.035)
	Career Satisfaction	0.00	(0.014)	0.01	(0.021)	–0.01	(0.027)	0.00	(0.035)
	Relationship Satisfaction	0.01	(0.014)	0.02	(0.021)	0.02	(0.027)	0.01	(0.035)

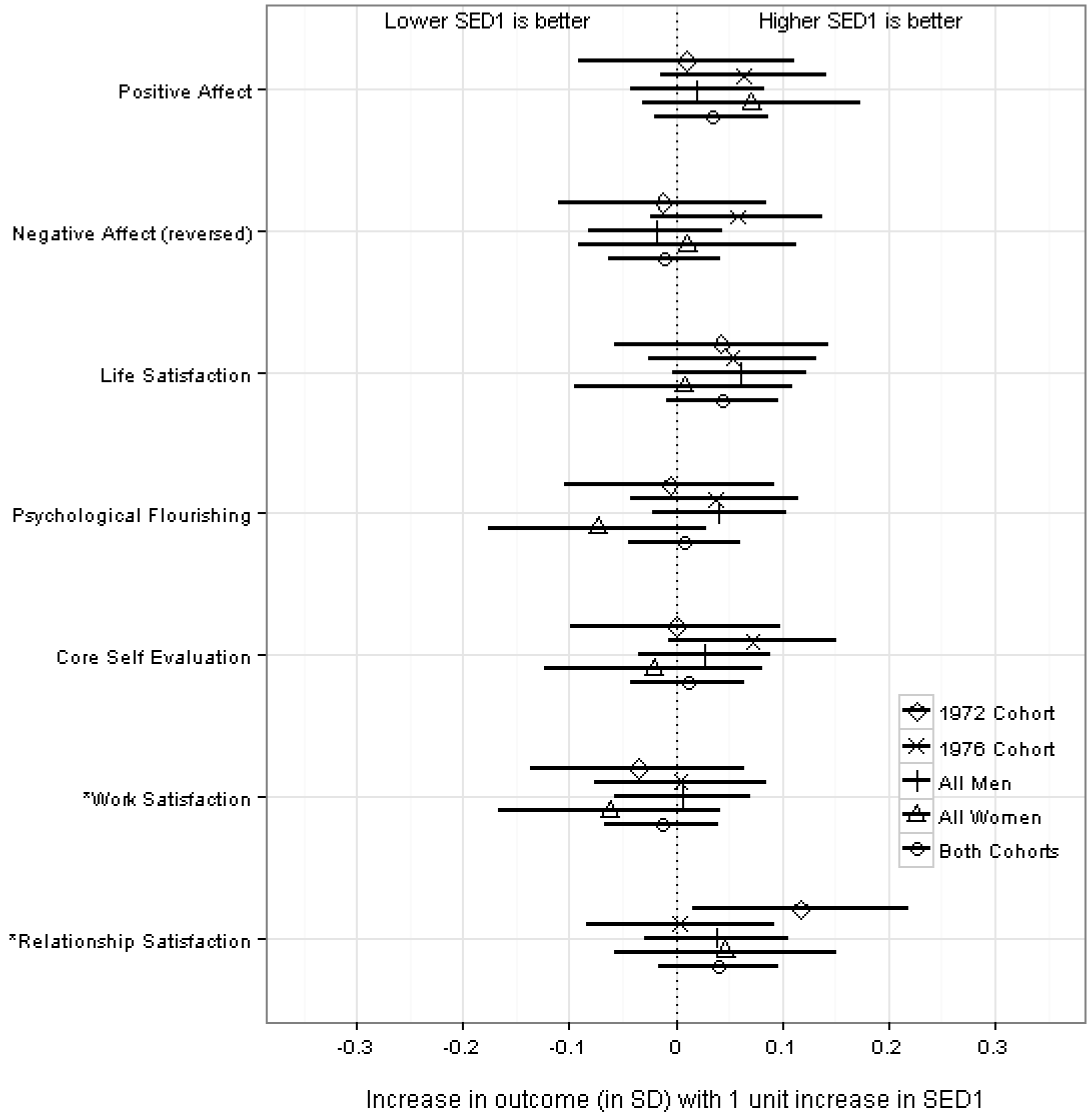

Figure 18: SED1 and Well-Being Outcomes After Controlling for Important Background Covariates without Subclassification

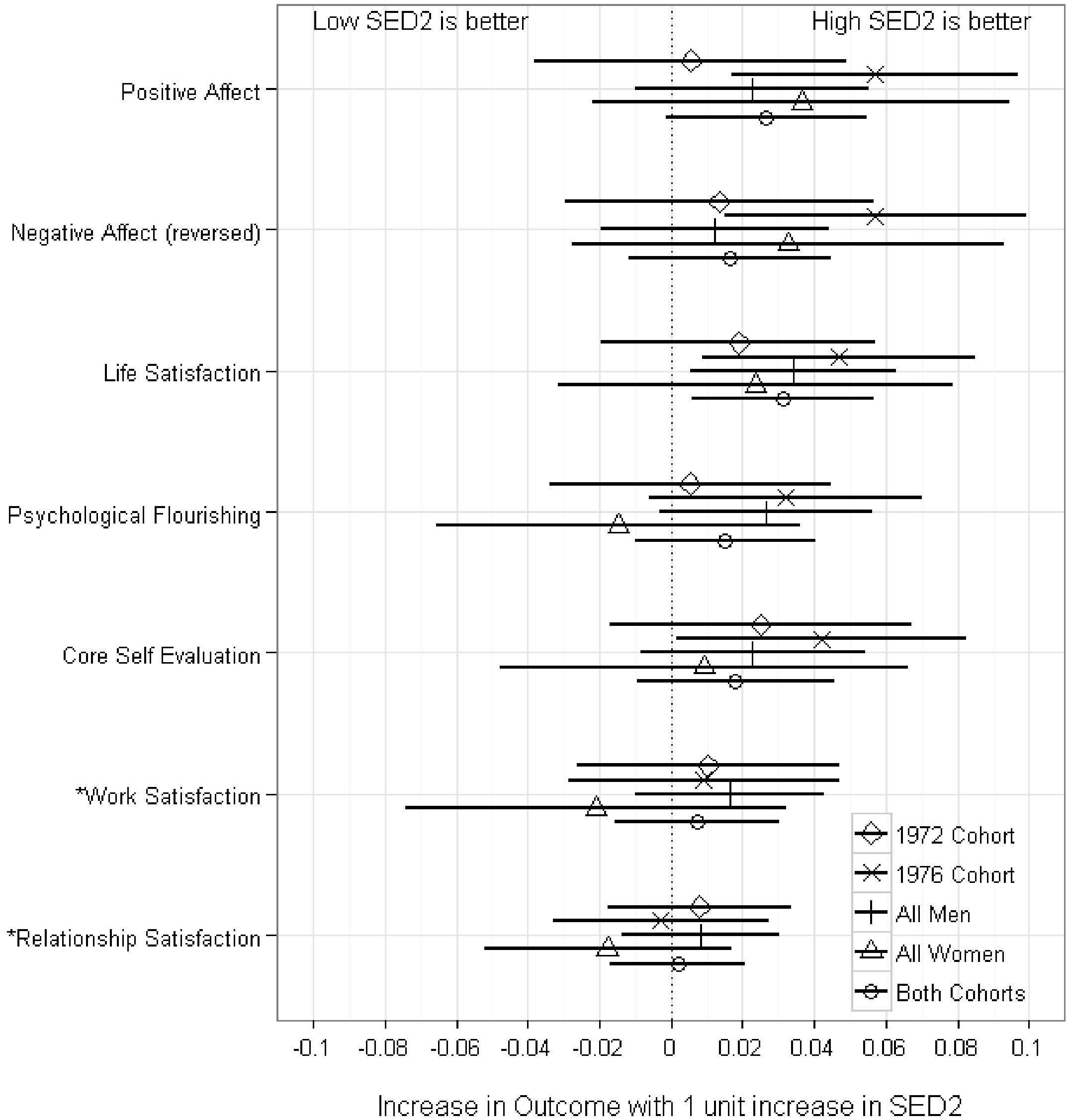

Figure 19: SED2 and Well-Being Outcomes after Controlling for Important Background Covariates without Subclassification

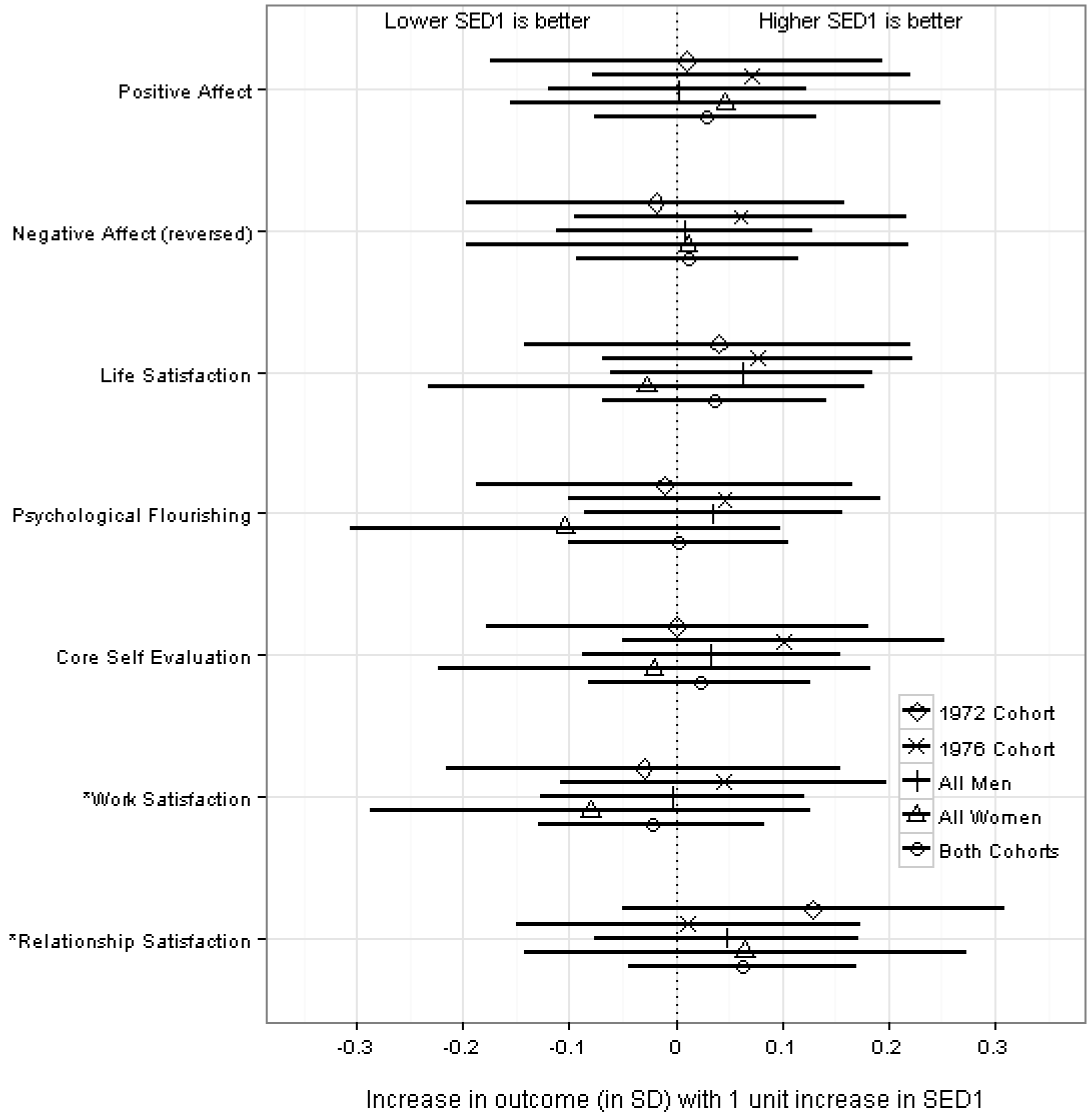

Figure 20: Plot of the Linear Regression Coefficients for SED1 Averaged Across 3 Subclasses

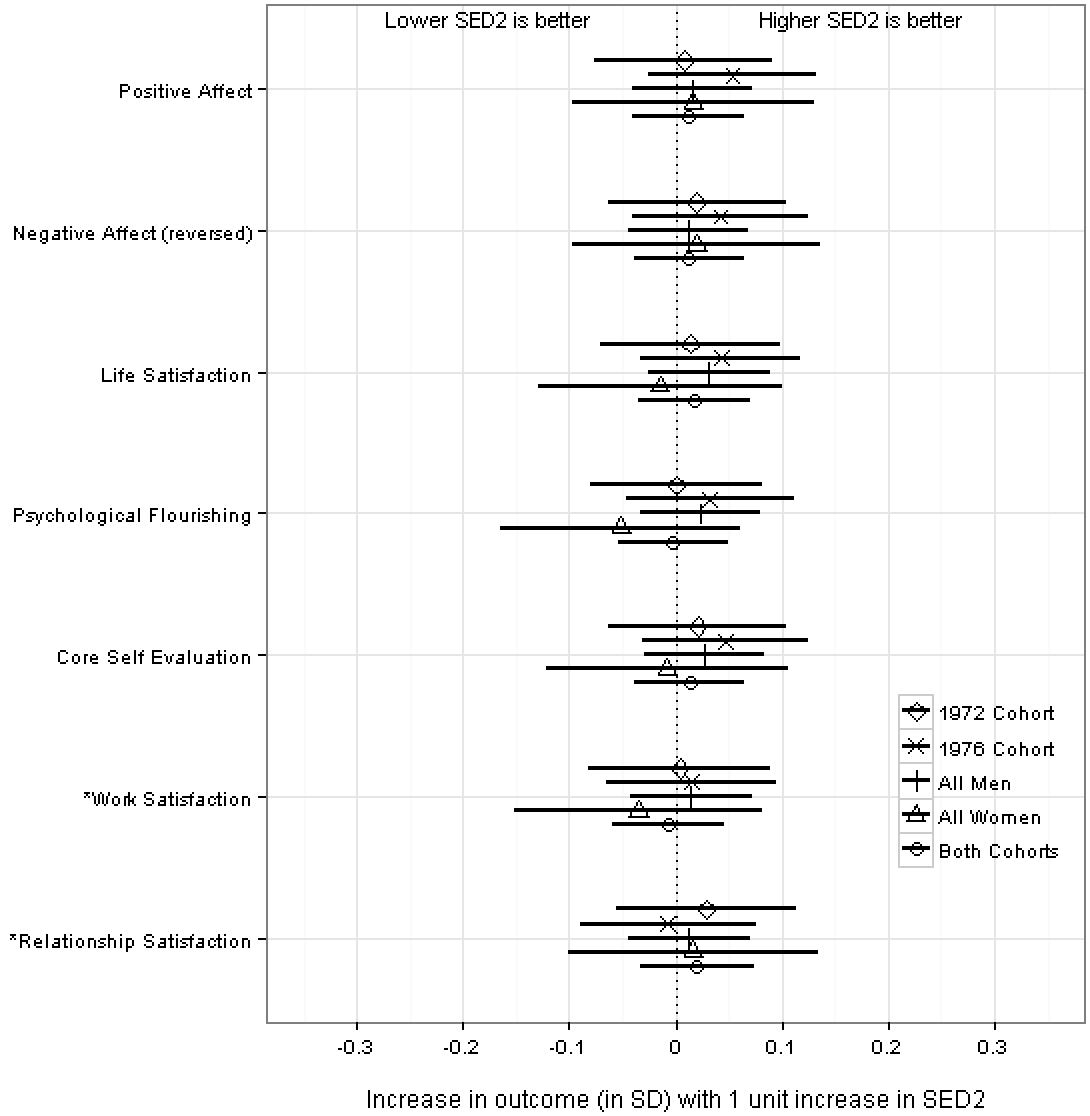

Figure 21: Plot of the Linear Regression Coefficients for SED2 Averaged Across 3 Subclasses

Health

Because most health issues were not high in prevalence, subclassification using more than 2 subclasses often resulted in problems estimating the logistic regression model. Therefore analyses were limited to a model without subclasses and a model with two subclasses.

Table 23 presents the regression coefficients for SED1 for a model without subclasses (regular logistic regression) and a model with 2 subclasses. Only 29 participants in the 1976 Cohort reported having arthritis. Creating two subclasses resulted in insufficient information in the data to estimate the regression models for subclasses separately. Therefore these results are omitted from the table and figure. Both analyses show consistent results. None of the regression coefficients is significantly different from zero. For outcomes hypertension and cholesterol all regression coefficients are negative, favoring higher levels of SED1. Results without subclassification and with two subclasses are visually presented in Figure 22 and Figure 23, respectively. Notice the increase in the standard error of the estimates when using subclassification.

In contrast with previous analyses, odds ratios were used to visualize results to better align the tables of coefficients and the figures. Odds ratios have a direct relationship with the regression coefficients (they are obtained by exponentiating the regression coefficients) and therefore can be calculated without information about a participant's score on the other background covariates. That is not the case for risk ratios. For small values, which is the case in the present study, odds ratios and risk ratios can be interpreted similarly.

Table 22 and Figures 24 & 26 show the result of similar analyses for SED2. The same pattern appears, with cholesterol and hypertension being consistently less prevalent among participants with higher levels of SED2. Without controlling for multiple comparisons, these regression coefficients were significantly different from zero for the 1976 cohort, all men, and both cohorts combined. In a model with two subclasses, these coefficients were no longer significant except hypertension for both cohorts combined, again favoring participants that received a higher SED2.

In conclusion, no evidence was found for a harmful effect of STEM Educational Dose on health at age 50. Most effects are small and non-significant. When patterns and significant effects were found, they always favored participants with higher levels of SED2.

Table 23: Regression Coefficients for SED1 and Selected Health Issues

		No Subclasses		2 Subclasses	
Sample	Outcome	Est.	SE	Est.	SE
1972 Cohort	cholesterol	–0.04	(0.121)	–0.06	(0.182)
	allergies	0.17	(0.138)	0.12	(0.206)
	indigestion	–0.05	(0.183)	–0.11	(0.281)
	back problems	0.07	(0.123)	0.04	(0.183)
	trouble sleeping	0.12	(0.147)	0.11	(0.224)
	arthritis	–0.04	(0.159)	–0.05	(0.246)
	hypertension	–0.09	(0.131)	–0.13	(0.200)
	psychiatric	–0.08	(0.162)	–0.09	(0.240)
	depression	–0.02	(0.148)	–0.05	(0.222)
1976 Cohort	cholesterol	–0.13	(0.109)	–0.16	(0.171)
	allergies	–0.02	(0.104)	–0.04	(0.166)
	indigestion	–0.07	(0.160)	–0.35	(0.313)
	back problems	0.03	(0.106)	0.07	(0.170)
	trouble sleeping	–0.14	(0.117)	–0.13	(0.184)
	arthritis	0.02	(0.150)	NA	NA
	hypertension	–0.13	(0.120)	–0.16	(0.201)
	psychiatric	0.02	(0.118)	0.04	(0.186)
	depression	0.00	(0.111)	–0.01	(0.169)
All Men	cholesterol	–0.08	(0.073)	–0.08	(0.113)
	allergies	0.18	(0.084)	0.24	(0.133)
	indigestion	–0.13	(0.129)	–0.14	(0.198)
	back problems	0.01	(0.078)	0.07	(0.121)
	trouble sleeping	0.06	(0.093)	0.04	(0.148)
	arthritis	–0.03	(0.115)	–0.04	(0.179)
	hypertension	–0.09	(0.085)	–0.09	(0.130)
	psychiatric	0.06	(0.096)	–0.12	(0.167)
	depression	0.06	(0.090)	0.03	(0.141)
All Women	cholesterol	–0.27	(0.159)	–0.18	(0.257)
	allergies	–0.07	(0.118)	–0.17	(0.197)
	indigestion	0.09	(0.149)	0.05	(0.256)
	back problems	–0.05	(0.122)	–0.08	(0.202)
	trouble sleeping	–0.13	(0.139)	–0.09	(0.224)
	arthritis	–0.03	(0.134)	–0.03	(0.214)
	hypertension	–0.01	(0.142)	–0.10	(0.236)
	psychiatric	0.01	(0.126)	–0.07	(0.211)
	depression	0.02	(0.121)	–0.05	(0.208)
Both Cohorts	cholesterol	–0.12	(0.066)	–0.11	(0.105)
	allergies	0.10	(0.068)	0.08	(0.108)
	indigestion	–0.02	(0.096)	–0.05	(0.150)
	back problems	0.00	(0.066)	0.00	(0.105)
	trouble sleeping	0.00	(0.077)	0.00	(0.121)
	arthritis	–0.03	(0.087)	–0.04	(0.134)
	hypertension	–0.08	(0.072)	–0.08	(0.114)
	psychiatric	0.05	(0.077)	–0.04	(0.123)
	depression	0.05	(0.072)	0.01	(0.114)

Note. NA: could not be estimated, Est. = Estimate, SE = Standard Error of the Estimate

Table 24: Regression Coefficients for SED2 and Selected Health Outcomes

		No Subclasses		2 Subclasses	
Sample	Outcome	Est.	SE	Est.	SE
1972 Cohort	cholesterol	–0.03	(0.051)	–0.05	(0.085)
	allergies	0.08	(0.059)	0.04	(0.094)
	indigestion	0.02	(0.078)	–0.02	(0.127)
	back problems	–0.04	(0.053)	–0.05	(0.084)
	trouble sleeping	0.02	(0.064)	0.00	(0.104)
	arthritis	0.05	(0.067)	0.05	(0.103)
	hypertension	–0.09	(0.060)	–0.12	(0.096)
	psychiatric	–0.02	(0.071)	–0.06	(0.113)
	depression	–0.12	(0.072)	–0.16	(0.117)
1976 Cohort	cholesterol	–0.12	(0.057)*	–0.13	(0.092)
	allergies	–0.02	(0.054)	–0.03	(0.084)
	indigestion	0.04	(0.080)	0.00	(0.139)
	back problems	–0.01	(0.056)	0.03	(0.090)
	trouble sleeping	–0.04	(0.061)	–0.06	(0.098)
	arthritis	0.00	(0.082)	0.06	(0.157)
	hypertension	–0.14	(0.066)*	–0.20	(0.108)
	psychiatric	–0.02	(0.062)	–0.03	(0.100)
	depression	–0.01	(0.058)	–0.03	(0.088)
All Men	cholesterol	–0.09	(0.038)*	–0.09	(0.056)
	allergies	0.05	(0.042)	0.05	(0.061)
	indigestion	–0.03	(0.064)	–0.03	(0.092)
	back problems	–0.03	(0.040)	–0.05	(0.061)
	trouble sleeping	–0.02	(0.047)	–0.05	(0.077)
	arthritis	–0.02	(0.059)	–0.04	(0.090)
	hypertension	–0.10	(0.045)*	–0.12	(0.068)
	psychiatric	0.00	(0.050)	–0.02	(0.076)
	depression	–0.02	(0.047)	–0.03	(0.069)
All Women	cholesterol	–0.05	(0.090)	–0.03	(0.145)
	allergies	–0.03	(0.067)	–0.09	(0.117)
	indigestion	0.11	(0.082)	0.12	(0.133)
	back problems	–0.03	(0.070)	0.00	(0.112)
	trouble sleeping	–0.03	(0.078)	–0.01	(0.124)
	arthritis	0.03	(0.077)	0.05	(0.119)
	hypertension	–0.11	(0.089)	–0.17	(0.147)
	psychiatric	0.04	(0.071)	0.00	(0.119)
	depression	–0.01	(0.069)	–0.06	(0.118)
Both Cohorts	cholesterol	–0.08	(0.035)*	–0.09	(0.054)
	allergies	0.04	(0.035)	0.00	(0.056)
	indigestion	0.04	(0.049)	0.02	(0.075)
	back problems	–0.02	(0.034)	–0.04	(0.053)
	trouble sleeping	–0.01	(0.040)	–0.04	(0.064)
	arthritis	0.00	(0.046)	–0.01	(0.069)
	hypertension	–0.10	(0.040)*	–0.13	(0.063)*
	psychiatric	0.02	(0.041)	0.00	(0.063)
	depression	–0.01	(0.039)	–0.02	(0.058)

Note. NA: could not be estimated, Est. = Estimate, SE = Standard Error of the Estimate, * p<.05

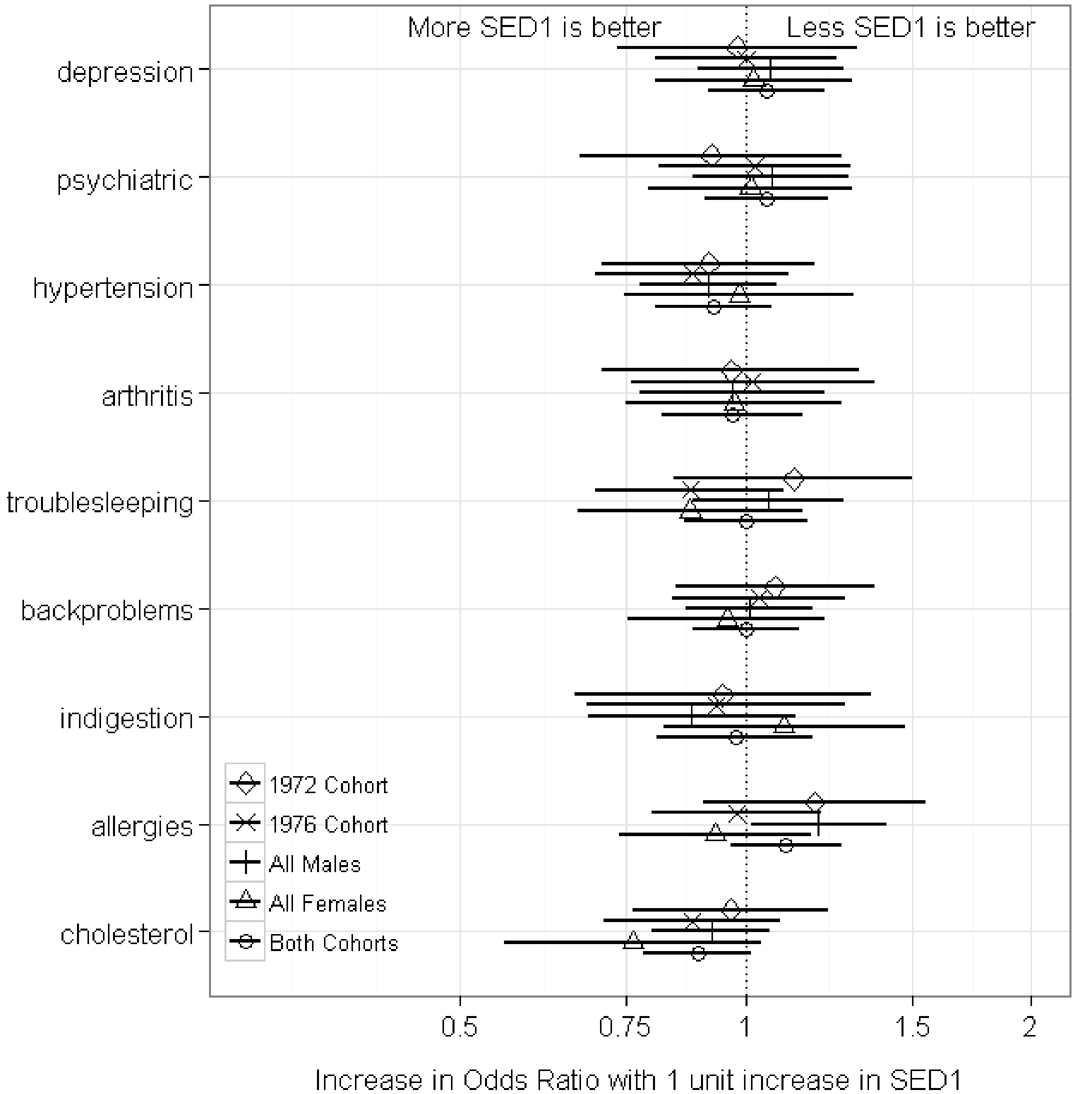

Figure 22: Association between SED1 and Selected Health Issues without Subclassification

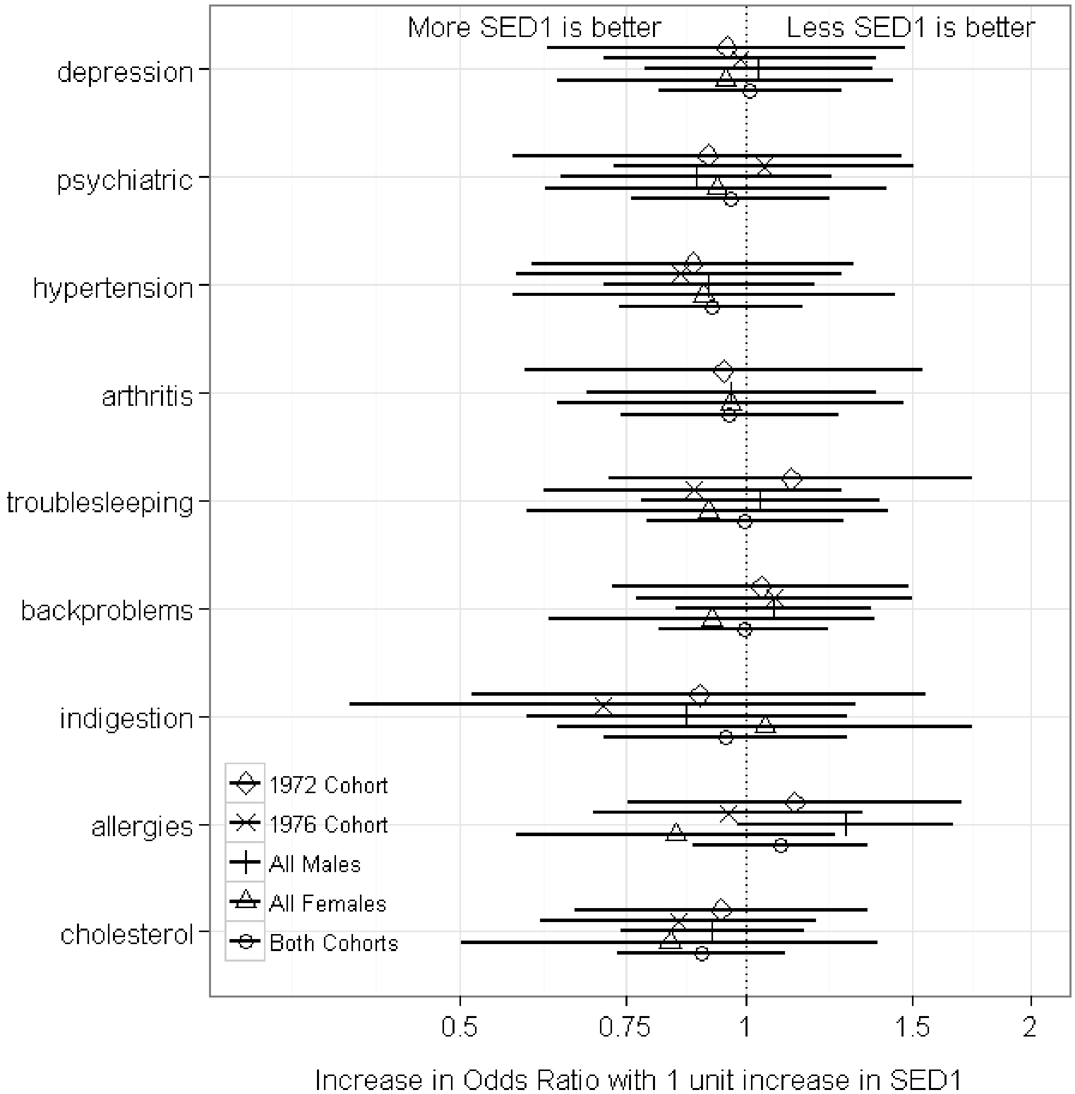

Figure 23: Association between SED1 and Selected Health Issues Averaged Across 2 Subclasses

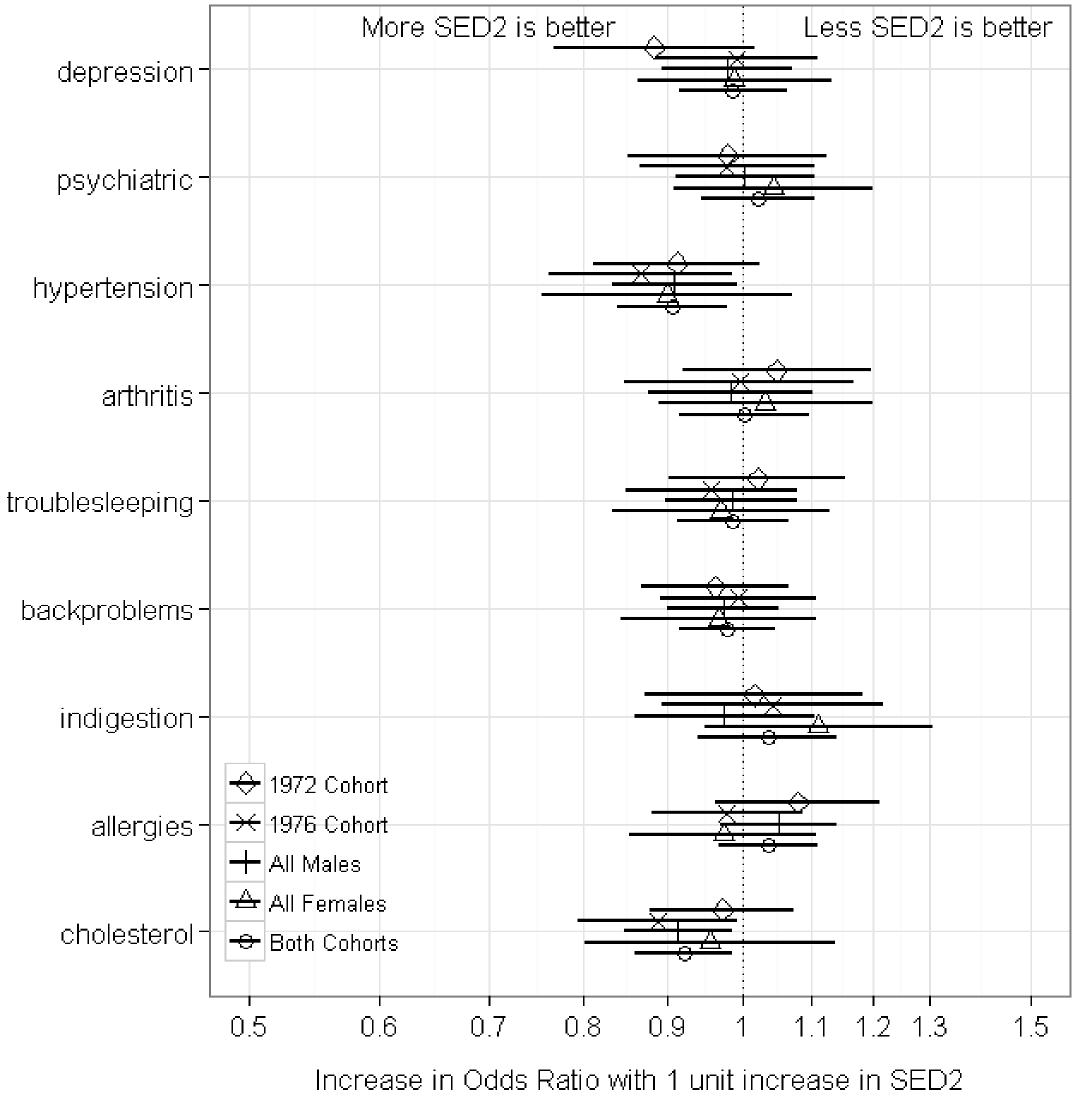

Figure 24: Association between SED2 and Selected Health Issues without Subclassification

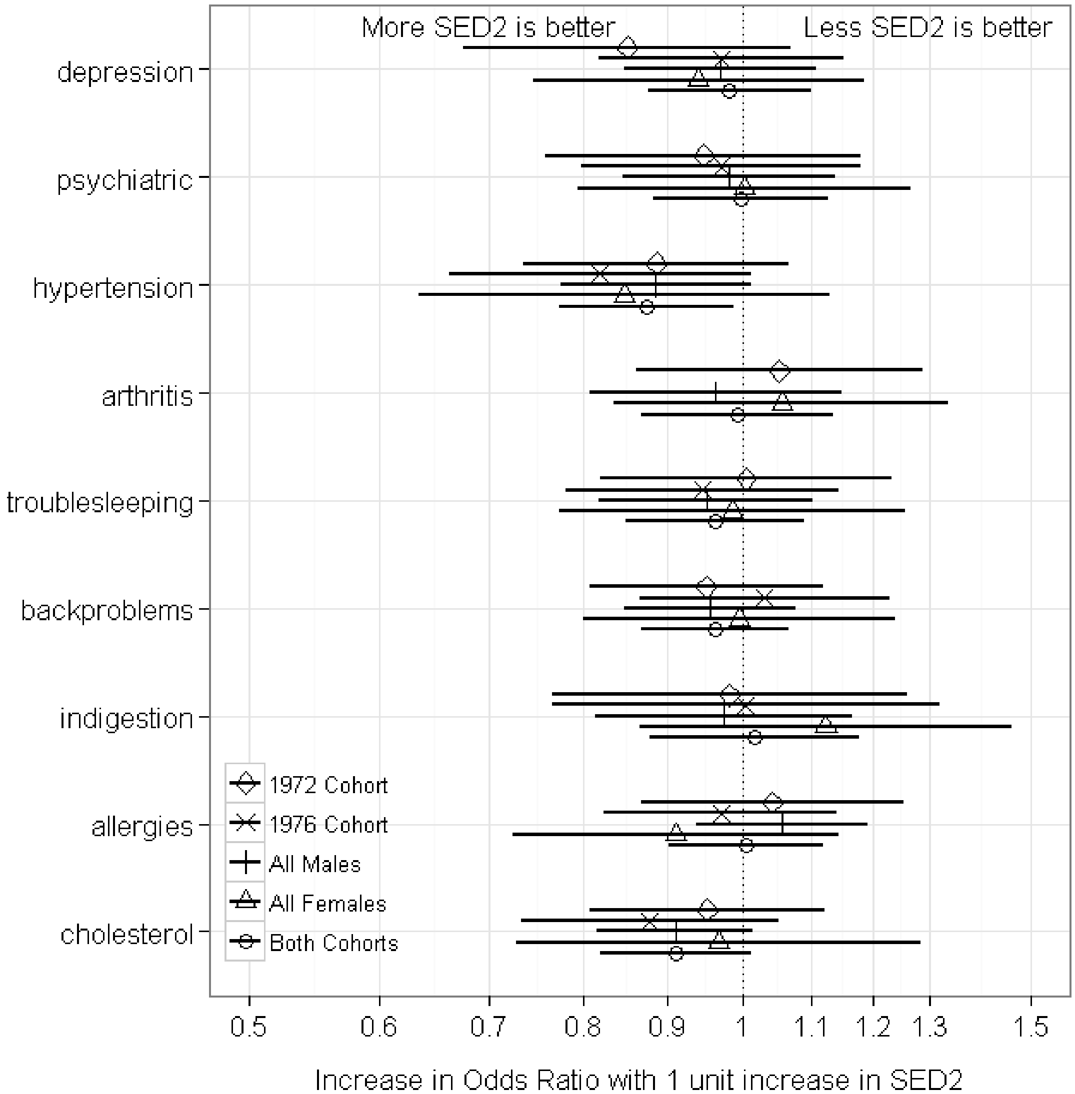

Figure 25: Association between SED2 and Selected Health Issues Averaged Across 2 Subclasses

Discussion and limitations

Study 2 evaluated the association between STEM educational dose and well-being 35 years later in the top 1 in 200 in mathematical ability. It tested the hypothesis that STEM educational dose among intellectually talented youth results in psychological maladjustment or distress at age 50.

Similar to study 1, participants in this study report experiencing mostly positive emotions and being satisfied with their lives, themselves, their work and their relationships. Compared to a nationally representative sample, they have less severe health issues. Only one item (emotional, nervous, or psychiatric problems) was endorsed more often in the current sample.

No evidence was found for a harmful effect of STEM educational dose on well-being or health at age 50. When there were differences between participants that received different levels of STEM Educational Dose, they favored those who received a higher STEM educational dose. Nevertheless, readers need to be aware of the following limitations.

First, the validity of the findings depends on the assumption that all confounding variables were included in the observed background covariates. For example measures for spatial ability or personality traits were not included in the sub-classification procedure.

Second, the high amount of attrition at the age 50 follow-up may have introduced bias. However, a comparison of the high and low dose groups before and after age 50 attrition on important background covariates did not provide evidence for differential attrition among groups.

Third, the operationalization of educational dose used in this study is rather crude. For example, participation in a competition or taking an AP course are both worth one Educational Dose Unit (EDU). The survey yielded insufficient information to judge if the intellectual challenge of participating in a competition is comparable in intensity with taking a whole semester of AP.

Despite these limitations, this was the first time to evaluate the effects of subject-based acceleration on social and emotional adjustment of the top 1 in 200 in mathematical ability 35 years later. In addition, it is the first time that this topic is investigated using data that are pre-processed using generalized propensity scores.

Implications of this study will be further discussed in the general discussion and conclusion.

CHAPTER VI

GENERAL DISCUSSION AND CONCLUSION

Concerns about the social-emotional development have plagued acceleration and its use. Since its inception, administrators, counselors, teachers, and parents have been concerned that acceleration may harm social and emotional adjustment, causing a decrease in the academic self-concept, social isolation, stress, burnout, psychological disorders, and elitism. Study after study has been conducted failing to provide evidence for these concerns (e.g., Makel, Lee, Olszewski-Kubilius, & Putallaz, 2012; Fund for the Advancement of Education, 1957; Terman & Oden, 1947). Meta-analyses after meta-analysis yields the same result: there are no harmful effects of acceleration on social and emotional development (Kulik & Kulik, 1984; Rogers, 1991; Kent, 1992; Kulik, 2004; Steenbergen-Hu & Moon, 2011). Still, parents and educators remain skeptical.

The current study shows that there are no harmful effects of acceleration on social and emotional adjustment at midlife. Using rigorous methods to control for confounding background covariates including abilities, SES, and motivation for school, the current study found no differences in well-being and health at midlife between accelerants and non-accelerants. Both groups were equally well-adjusted and report feeling happy, enjoying life, and being satisfied with their career, relationship, themselves, and their lives. In addition, no sex differences were found. Both male and female accelerants are equally well-adjusted as their non-accelerated counterparts.

Based on a review of previous evidence and the empirical evidence reported here, it is time to put these concerns aside. Educational policy and decision making should not be based on individual bias and preference, but on the vast amount of research that shows the educational efficacy of acceleration for intellectually talented students without harming their social and emotional development. Through educational legislation, policy makers appear justified in establishing acceleration as the preferred intervention for carefully selected, highly able, and motivated students. Acceleration is a highly effective intervention, with no additional costs, and no need for additional curriculum or additional teacher training. When making the decision to accelerate a student, guidelines from research and experts in the field should be followed. For example, skipping a whole grade should not solely be based on academic ability and achievement, but also on non-cognitive factors, including attitudes and support of school administrators and teachers, emotional maturity, and social skills (for more details, see Colangelo et al., 2004).

Optimal talent development requires appropriate developmental placement (Benbow & Stanley, 1996; Bleske-Rechek, Lubinski, & Benbow, 2004; Lubinski & Benbow, 2000). Highly able students will not excel without appropriate nourishment. It is the combination of high aptitude and extended periods of deliberate practice that will help them develop into eminent achievers (Ericsson, Krampe, & Tesch-Römer, 1993; Subotnik, Olszewski-Kubilius, & Worrell, 2011). Deliberate practice requires a good match between the aptitude of the student and the educational environment. Academic acceleration provides intellectually able learners with such a match. It helps maximize the amount of time they spend being intellectually challenged and deliberately working on improving their knowledge and skills. This is the kind of learning that is necessary to attain exceptional levels of achievement (Ericsson, Krampe, & Tesch-Römer, 1993; Subotnik, Olszewski-Kubilius, & Worrell, 2011).

Our society needs more, not less high achievement. In a world that is becoming increasingly more complex and dominated by a knowledge based economy, a nation's future prosperity will largely depend on how well it succeeds in cultivating its intellectual capital, especially in those individuals that have the potential to become exceptional achievers (Friedman, 2005; Domestic Policy Council, 2006; National Academy of Sciences, 2010; COMPETES, 2012). A large body of research supports the positive relationship between educational acceleration and academic, occupational, and creative accomplishments in the short and long term (Kulik, 2004; Kulik & Kulik, 1984; National Mathematics Advisory Panel, 2008; Rogers, 1991; Steenbergen-Hu & Moon, 2011; Park, Lubinski, & Benbow, 2012; Wai, Lubinski, Benbow, Steiger, 2010). Therefore, prejudices about acceleration should be set aside and carefully selected, able, and motivated students should be allowed to receive easy access to educational acceleration to facilitate the attainment of happy and productive lives. The empirical evidence suggests that acceleration will help these individuals excel without harming their social and emotional development.

APPENDIX

Table 25: Cross-Match Test Results for the 1972 Cohort

Imputation	Park et al. (2013)					Follow-up of Park et al. (2013)					
	Observed CM	Expected CM	Variance CM	Deviation	Appr. p-value	Observed CM	Expected CM	Variance CM	Deviation	Exact p-value	Appr. p-value
1	113	119.44	53.09	–0.88	0.1882	55	60.31	26.40	–1.03	0.1992	0.1506
2	113	119.44	53.09	–0.88	0.1882	49	60.31	26.40	–2.20	0.0229	0.0139
3	114	119.11	52.79	–0.70	0.2409	52	61.32	25.96	–1.83	0.0519	0.0336
4	121	119.44	53.09	0.21	0.5845	54	60.75	26.02	–1.32	0.1292	0.0927
5	118	119.11	52.79	–0.15	0.4393	69	61.13	26.35	1.53	0.9607	0.9374
6	123	119.44	53.09	0.49	0.6872	60	60.16	26.07	–0.03	0.5621	0.4875
7	108	119.11	52.79	–1.53	0.0631	55	61.33	26.33	–1.23	0.1489	0.1089
8	115	119.44	53.09	–0.61	0.2709	49	60.31	26.40	–2.20	0.0229	0.0139
9	106	119.11	52.79	–1.80	0.0356	53	60.93	26.36	–1.54	0.0887	0.0613
10	111	119.44	53.09	–1.16	0.1232	49	60.93	26.36	–2.32	0.0172	0.0101

Note. CM = number of cross-matches, Deviation = deviation of the observed number of cross-matches from the expected number of cross-matches in SD units. The Dunn-Šidák correction for multiple comparisons requires a p-value < .0051 to keep the family wise error rate at .05.

Table 26: Cross-Match Test Results for the 1976 Cohort

Imputation	Park et al. (2013)					Follow-up of Park et al. (2013)					
	Observed CM	Expected CM	Variance CM	Deviation	Appr. p-value	Observed CM	Expected CM	Variance CM	Deviation	Exact p-value	Appr. p-value
1	69	77.00	34.12	–1.37	0.0854	44	46.20	20.18	–0.49	0.3909	0.3121
2	75	77.00	34.12	–0.34	0.3660	46	46.61	20.15	–0.14	0.5307	0.4459
3	81	77.00	34.12	0.68	0.7533	54	46.20	20.18	1.74	0.9777	0.9587
4	79	77.00	34.12	0.34	0.6340	50	47.20	20.09	0.63	0.8020	0.7342
5	81	77.00	34.12	0.68	0.7533	52	45.78	20.19	1.39	0.9490	0.9170
6	67	77.00	34.12	–1.71	0.0435	42	45.78	20.19	–0.84	0.2658	0.2004
7	81	77.00	34.12	0.68	0.7533	40	46.61	20.15	–1.47	0.1054	0.0704
8	81	77.00	34.12	0.68	0.7533	48	46.61	20.15	0.31	0.7008	0.6216
9	79	77.00	34.12	0.34	0.6340	53	45.00	19.90	1.79	0.9807	0.9635
10	83	77.00	34.12	1.03	0.8478	47	46.04	19.85	0.21	0.6672	0.5851

Note. CM = number of cross-matches, Deviation = deviation of the observed number of cross-matches from the expected number of cross-matches in SD units. The Dunn-Šidák correction for multiple comparisons requires a p-value < .0051 to keep the family wise error rate at .05.

Table 27: Comparing Average Scores Across Studies with Normative Samples

	Study 1 Means [95% CI]	Study 2 Means [95% CI]	Standardization Sample
N	490	847	
Positive Affect	3.70 [3.64–3.76]	3.71 [3.66–3.75]	3.68[1]
Negative Affect	4.65 [4.56–4.75]	4.69 [4.62–4.75]	
Life Satisfaction	5.08 [4.97–5.19]	5.14 [5.05–5.22]	4.70[2]
Psychological Flourishing	5.81 [5.75–5.87]	5.83 [5.78–5.88]	5.62[1]
Core Self-Evaluations	3.75 [3.70–3.80]	3.76 [3.73–3.80]	3.87[3]
Career Satisfaction	5.38 [5.24–5.52]	5.45 [5.34–5.54]	
Relationship Satisfaction	6.20 [6.07–6.32]	6.16 [6.06–6.24]	

Note.Averages for Study 1 (grade skippers and matched controls combined), Study 2 (high and low SED groups combined), and the test construction.and validation sample.
[1] Diener et al. (2010)
[2] Diener, Emmons, Larsen, & Griffin (1985). Average score of the study 1 sample.
[3] Judge, Erez, Bono, & Thoresen (2003). Average score of all samples.

Table 28: Comparing the Prevalence of Severe Health Issues among the Study 1 and Study 2 Samples with the Prevalence among the NLSY79 Nationally Representative Sample.

	Study 1 Means [95% CI]		Study 2 Means [95% CI]		NLSY79 Age 40 no weights	NLSY79 Age 40 weighted[1]	NLSY79 Age 50[2] no weights	NLSY79 Age 50[2] weighted[1]
Congestive heart failure?	0.2	[0.0–0.7]	0.0	[0.0-NA]	0.3	0.2	1.7	1.8
A stroke?	0.2	[0.0–0.7]	0.5	[0.1–0.9]	1.0	1.0	2.7	2.6
Not including asthma, but chronic lung disease such as chronic bronchitis or emphysema?	0.5	[0.0–1.1]	0.5	[0.1–0.9]	2.9	2.9	3.4	3.4
High blood pressure or hypertension?	18.1	[14.7–21.6]	18.1	[15.5–20.7]	17.3	15.5	24.9	22.0
Diabetes or high blood sugar?	6.3	[4.2–8.5]	6.0	[4.4–7.6]	5.3	4.4	9.9	9.6
Arthritis or rheumatism?	13.7	[10.6–16.8]	12.3	[10.1–14.6]	11.7	12.3	16.7	16.5
Emotional, nervous, or psychiatric problems?	16.3	[13.0–19.6]	15.8	[13.3–18.3]	7.4	8.3	3.6	3.4
Cancer or malignant tumor of any kind except skin cancer?	3.3	[1.7–4.9]	2.6	[1.6–3.7]	2.1	2.4	3.2	3.5
Heart problems? [3]	6.6	[4.4–8.8]	7.2	[5.4–8.9]	2.9	2.9	6.3	6.6

Note. NA = not applicable (because none of the participants endorsed the item, a confidence interval could not be calculated), NLSY79 = National Longitudinal Survey of Youth 1979, a longitudinal survey study following a nationally representative sample of 12,686 individuals since 1979. At the time of the 2010 interviews, respondents were between 45 to 53 years old. Age 40 data come from the 40+ Health Module. NLSY79 Age 50 data come from the 50+ Health Module. These Health Modules were presented to the NLSY79 participants in the survey year after they turned 40 or 50, respectively. Not all participants received the modules in the same survey year. Therefore both unweighted and weighted results must be interpreted with care.

[1] Weights used come from survey year 2010

[2] In 2010, only 24% of the NLSY79 sample had reached the age of 50 and completed the 50+ health module

[3] Also includes congestive heart failures

Table 29: Means, Standardized Mean Differences (SMD), and SMD's Adjusted for Background Covariates for Grade Skippers and Matched Controls

Sample	Outcome	Matched Controls	Grade Skippers	SMD	ASMD	Lower 95% CI ASMD	Upper 95% CI ASMD
1972 Cohort	Positive Affect	22.24	22.59	0.09	0.15	–0.13	0.43
	Negative Affect (reversed)	46.98	47.01	0.00	0.03	–0.27	0.33
	Life Satisfaction	25.52	24.70	–0.12	–0.11	–0.37	0.15
	Psychological Flourishing	70.14	70.03	–0.01	–0.02	–0.28	0.25
	Core Self Evaluations	44.93	45.03	0.01	–0.01	–0.29	0.28
	*Career Satisfaction	5.35	5.32	–0.02	–0.07	–0.33	0.18
	*Relationship Satisfaction	6.14	6.19	0.03	0.05	–0.13	0.22
1976 Cohort	Positive Affect	21.99	22.04	0.01	–0.01	–0.32	0.30
	Negative Affect (reversed)	45.75	46.31	0.05	0.09	–0.24	0.42
	Life Satisfaction	25.44	25.90	0.07	0.03	–0.27	0.34
	Psychological Flourishing	69.61	68.74	–0.09	–0.14	–0.45	0.16
	Core Self Evaluations	44.61	45.75	0.17	0.13	–0.19	0.44
	*Career Satisfaction	5.49	5.32	–0.12	–0.10	–0.40	0.21
	*Relationship Satisfaction	6.31	6.11	–0.17	0.02	–0.23	0.27
All Men	Positive Affect	21.81	22.03	0.05	0.04	–0.22	0.30
	Negative Affect (reversed)	47.39	46.95	–0.04	–0.02	–0.29	0.26
	Life Satisfaction	25.31	24.90	–0.06	–0.11	–0.37	0.14
	Psychological Flourishing	69.12	68.10	–0.11	–0.16	–0.42	0.11
	Core Self Evaluations	45.28	45.14	–0.02	–0.05	–0.32	0.23
	*Career Satisfaction	5.42	5.24	–0.12	–0.15	–0.39	0.08
	*Relationship Satisfaction	6.19	6.08	–0.08	0.03	–0.14	0.21
All Women	Positive Affect	22.61	22.81	0.05	0.12	–0.25	0.50
	Negative Affect (reversed)	45.04	46.36	0.12	0.16	–0.19	0.52
	Life Satisfaction	25.75	25.68	–0.01	0.07	–0.26	0.40
	Psychological Flourishing	71.10	71.41	0.04	0.08	–0.23	0.39
	Core Self Evaluations	44.08	45.63	0.22	0.24	–0.08	0.56
	*Career Satisfaction	5.39	5.44	0.03	0.06	–0.27	0.39
	*Relationship Satisfaction	6.27	6.27	0.00	0.02	–0.21	0.25
Both Cohorts	Positive Affect	22.13	22.35	0.06	0.07	–0.14	0.27
	Negative Affect (reversed)	46.45	46.71	0.02	0.05	–0.18	0.27
	Life Satisfaction	25.48	25.22	–0.04	–0.06	–0.26	0.14
	Psychological Flourishing	69.91	69.46	–0.05	–0.06	–0.27	0.14
	Core Self Evaluations	44.80	45.34	0.08	0.05	–0.16	0.25
	*Career Satisfaction	5.41	5.32	–0.06	–0.07	–0.26	0.11
	*Relationship Satisfaction	6.22	6.16	–0.05	0.03	–0.11	0.16

Note. SMD = Standardized Mean difference or Cohen's d, ASMD = Adjusted Standardized Mean Difference, or the robust regression coefficient for the dichotomous variable indicating whether a participant skipped one or more grades, adjusted for background covariates, CI = Confidence Interval (not adjusted for multiple comparisons). * Outcomes with a star represent single items. Regression residuals for these outcomes were not normally distributed, therefore results must be interpreted with care.

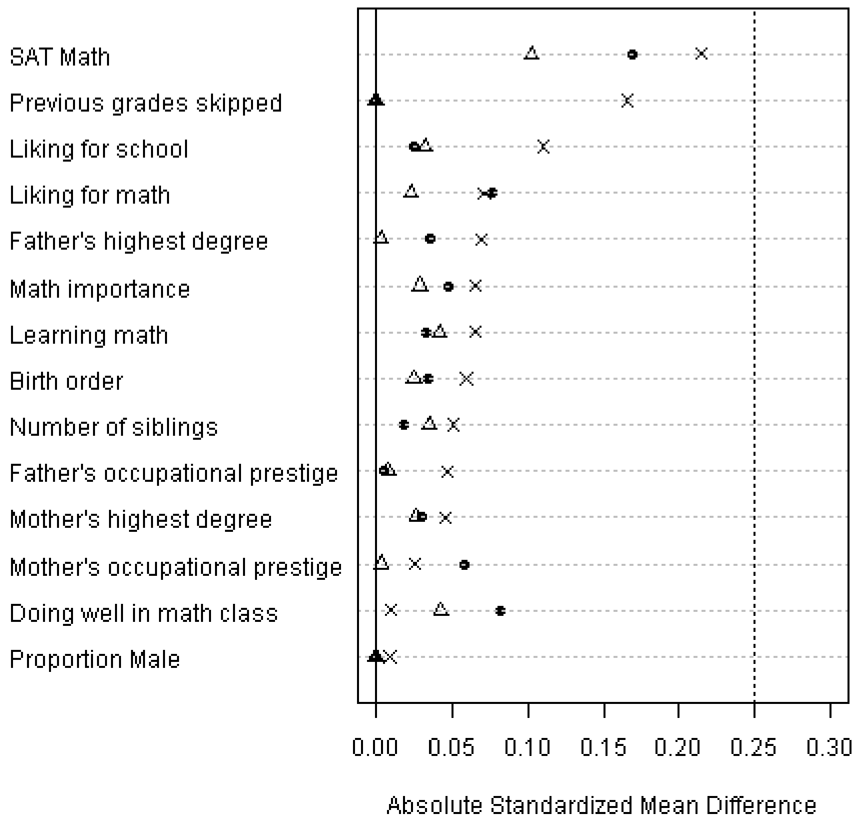

Figure 26: Absolute Standardized Mean Differences between Grade Skippers and Newly Matched Controls for the 1972 Cohort

The absolute standardized mean differences between the grade skippers (*n* = 116) and the matched controls (*n* = 231) of Park et al. (2013; triangles), for grade skippers (*n* = 68) and matched controls (*n* = 143) who completed the age 50 survey (x's) on important background covariates, and for grade skippers and newly matched controls (black dots). The dotted vertical line indicates the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (2009). Data come from the 1976 Cohort.

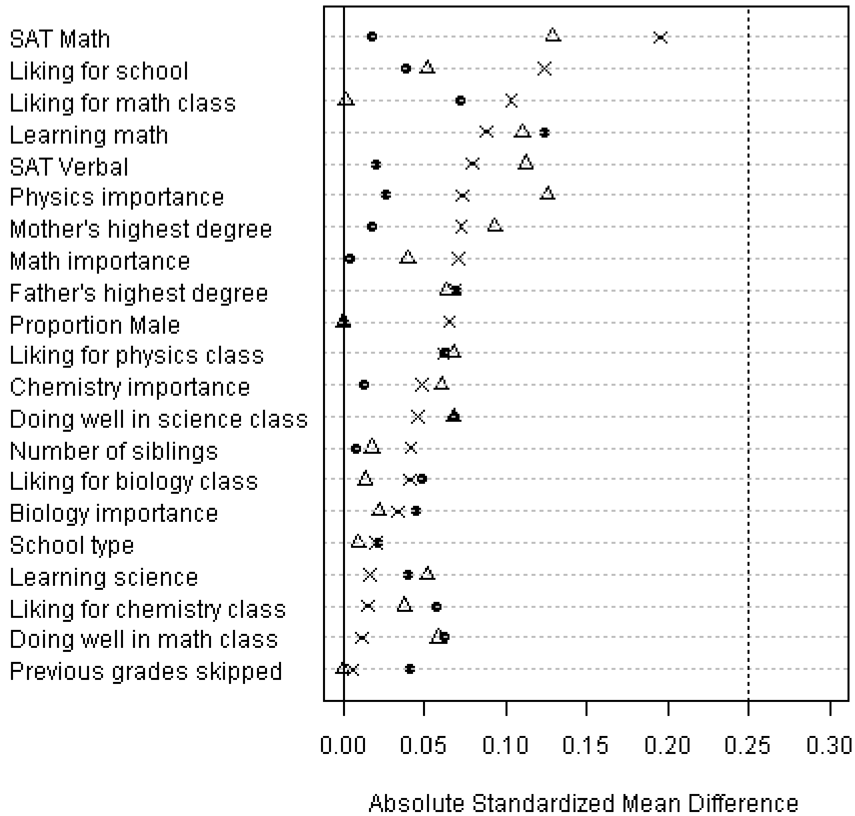

Figure 27: Aboslute Standardized Mean Differences between Grade Skippers and Newly Matched Controls for the 1976 Cohort

The absolute standardized mean differences between the grade skippers (*n* = 179) and the matched controls (*n* = 358) of Park et al. (2013; triangles), for grade skippers (*n* = 89) and matched controls (*n* = 191) who completed the age 50 survey (x's), and for grade skippers and newly matched controls (black dots) on important background covariates. The dotted vertical line indicates the maximum allowed difference between equivalent groups in a quasi-experimental design according to What Works Clearinghouse (2009). Data come from the 1972 Cohort.

Table 30: Cross-Match Test Results for the 1972 Cohort (Newly Matched Groups)

Imputation	Follow-up of Park et al. (2013)						Newly Matched Groups					
	Observed CM	Expected CM	Variance CM	Deviation	Exact p-value	Appr. p-value	Observed CM	Expected CM	Variance CM	Deviation	Exact p-value	Appr. p-value
1	55	60.31	26.40	–1.03	0.1992	0.1506	62	59.11	26.12	0.57	0.7765	0.7142
2	49	60.31	26.40	–2.20	0.0229	0.0139	60	59.11	26.12	0.17	0.6422	0.5692
3	52	61.32	25.96	–1.83	0.0519	0.0336	48	59.11	26.12	–2.17	0.0244	0.0149
4	54	60.75	26.02	–1.32	0.1292	0.0927	56	59.11	26.12	–0.61	0.3372	0.2715
5	69	61.13	26.35	1.53	0.9607	0.9374	55	59.45	26.42	–0.86	0.2494	0.1936
6	60	60.16	26.07	–0.03	0.5621	0.4875	50	59.11	26.12	–1.78	0.0566	0.0373
7	55	61.33	26.33	–1.23	0.1489	0.1089	53	59.45	26.42	–1.25	0.1440	0.1049
8	49	60.31	26.40	–2.20	0.0229	0.0139	52	59.11	26.12	–1.39	0.1156	0.0821
9	53	60.93	26.36	–1.54	0.0887	0.0613	57	59.45	26.42	–0.48	0.3864	0.3171
10	49	60.93	26.36	–2.32	0.0172	0.0101	50	59.11	26.12	–1.78	0.0566	0.0373

Note. CM = number of cross-matches, Deviation = deviation of the observed number of cross-matches from the expected number of cross-matches in SD units. The Dunn-Šidák correction for multiple comparisons requires a p-value < .0051 to keep the family wise error rate at .05.

Table 31: Cross-Match Test results for the 1976 Cohort (Newly Matched Groups)

Imputation	Follow-up of Park et al. (2013)						Newly Matched Groups					
	Observed CM	Expected CM	Variance CM	Deviation	Exact p-value	Appr. p-value	Observed CM	Expected CM	Variance CM	Deviation	Exact p-value	Appr. p-value
1	44	46.20	20.18	–0.49	0.3909	0.3121	54	45.56	20.20	1.88	0.9845	0.9699
2	46	46.61	20.15	–0.14	0.5307	0.4459	46	45.56	20.20	0.10	0.6234	0.5393
3	54	46.20	20.18	1.74	0.9777	0.9587	54	45.56	20.20	1.88	0.9845	0.9699
4	50	47.20	20.09	0.63	0.8020	0.7342	46	45.56	20.20	0.10	0.6234	0.5393
5	52	45.78	20.19	1.39	0.9490	0.9170	54	45.56	20.20	1.88	0.9845	0.9699
6	42	45.78	20.19	–0.84	0.2658	0.2004	50	45.56	20.20	0.99	0.8892	0.8386
7	40	46.61	20.15	–1.47	0.1054	0.0704	48	45.56	20.20	0.54	0.7781	0.7067
8	48	46.61	20.15	0.31	0.7008	0.6216	48	45.56	20.20	0.54	0.7781	0.7067
9	53	45.00	19.90	1.79	0.9807	0.9635	46	45.56	20.20	0.10	0.6234	0.5393
10	47	46.04	19.85	0.21	0.6672	0.5851	52	45.56	20.20	1.43	0.9540	0.9242

Note. CM = number of cross-matches, Deviation = deviation of the observed number of cross-matches from the expected number of cross-matches in SD units. The Dunn-Šidák correction for multiple comparisons requires a p-value < .0051 to keep the family wise error rate at .05.

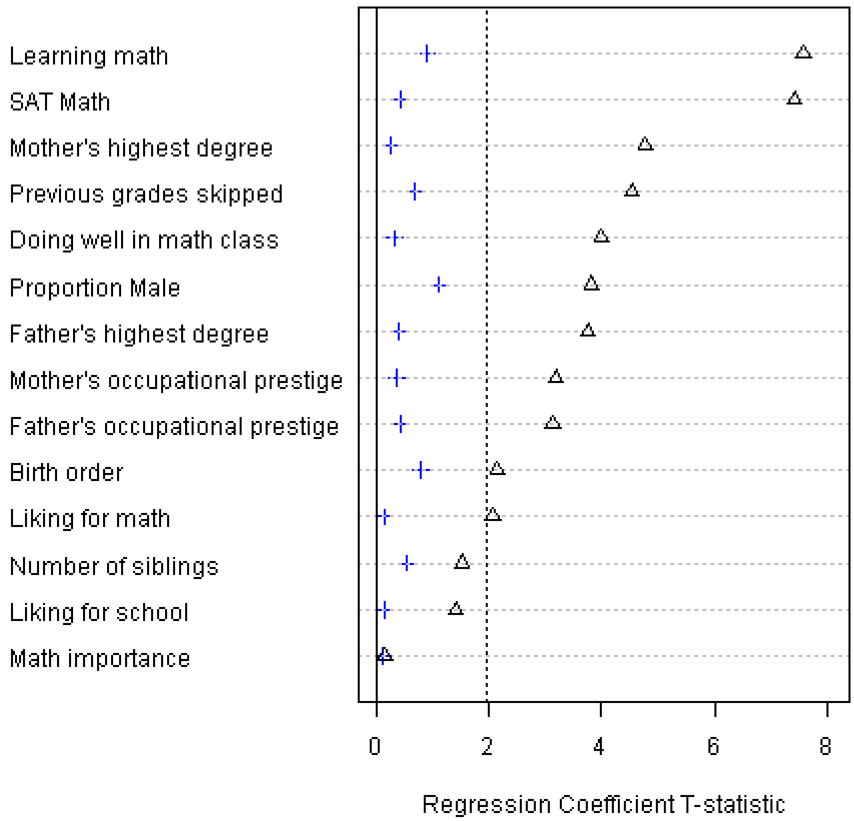

Figure 28: Evaluating Balance Using the Generalized Propensity Score for SED1 in the 1972 Cohort

Separate regression models with the background covariates as the outcomes were constructed. Triangles show the t-statistic for the regression coefficient of SED1 as the only predictor, crosses show the t-statistic for the regression coefficient of SED1 when the generalized propensity score is added as a predictor.

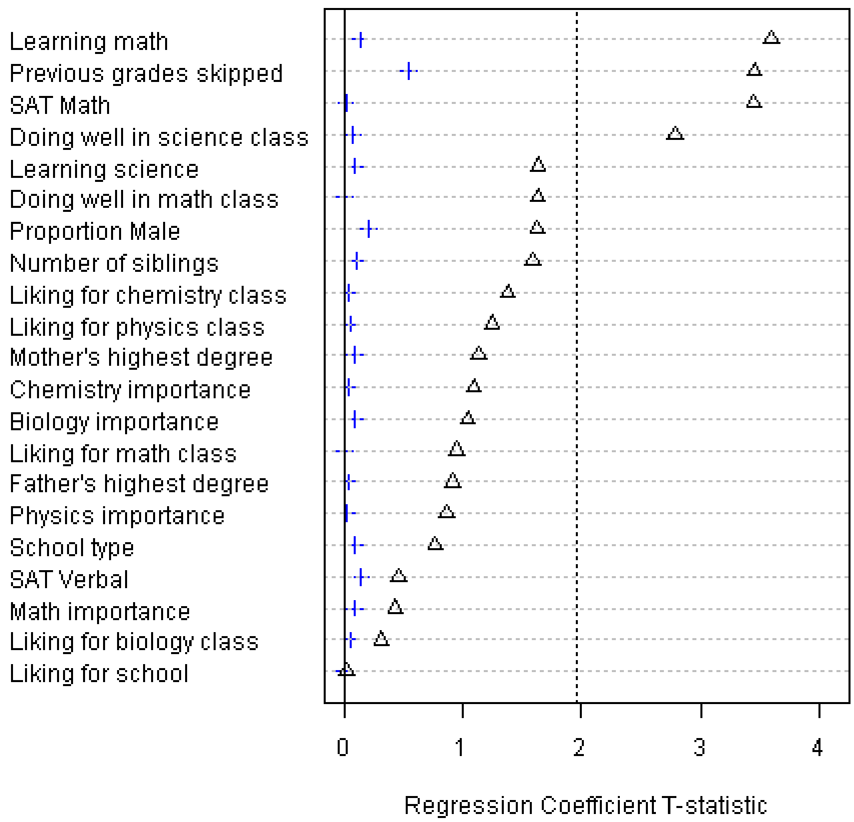

Figure 29: Evaluating Balance for the Generalized Propensity Score Using SED1 for the 1976 Cohort

Separate regression models with the background covariates as the outcomes were constructed. Triangles show the t-statistic for the regression coefficient of SED1 as the only predictor, crosses show the t-statistic for the regression coefficient of SED1 when the generalized propensity score is added as a predictor.

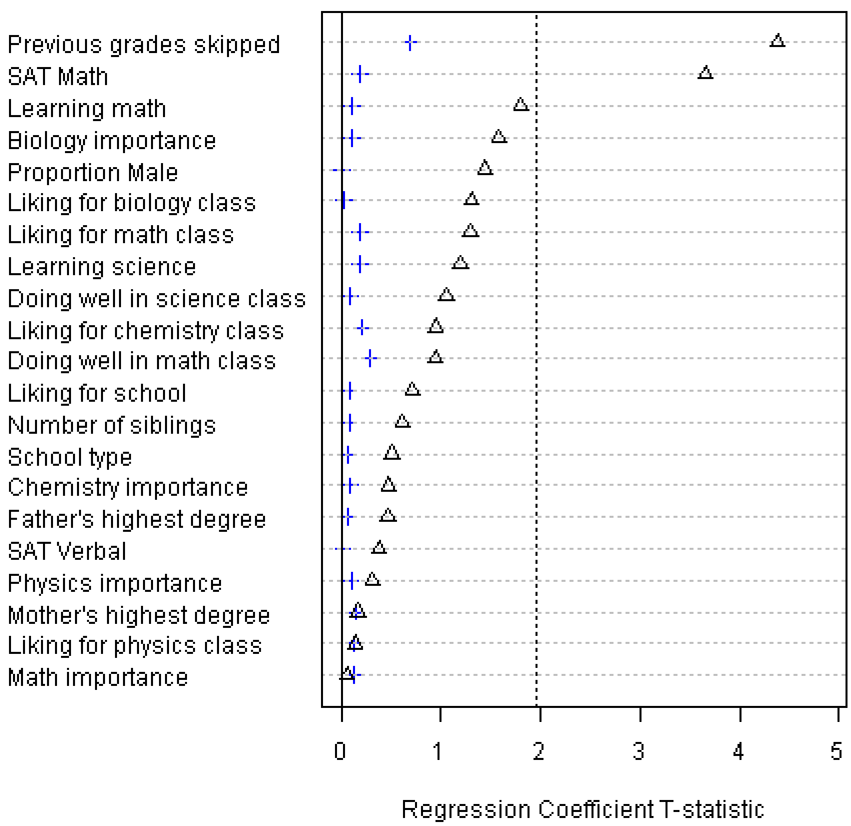

Figure 30: Evaluating Balance Using the Generalized Propensity Score for SED2 in the 1976 Cohort

Separate regression models with the background covariates as the outcomes were constructed. Triangles show the t-statistic for the regression coefficient of SED2 as the only predictor, crosses show the t-statistic for the regression coefficient of SED2 when the generalized propensity score is added as a predictor.

REFERENCES

Asher, S. R., & Gottman, J. M. (Eds.). (1981). *The development of children's friendships*. New York: Cambridge University Press.

Bachman, J. G., Staff, J., O'Malley, P. M., Schulenberg, J. E., & Freedman-Doan, P. (2011). Twelfth-grade student work intensity linked to later educational attainment and substance use: New longitudinal evidence. *Developmental Psychology*, *47*(2), 344–363. doi: 10.1037/a0021027

Bai, H. (2011). Using propensity score analysis for making causal claims in research articles. *Educational Psychology Review*, *23*(2), 273–278. doi: http://dx.doi.org/10.1007/s10648–011–9164–9

Barrett, L. F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*(4), 967–984. doi: http://dx.doi.org/10.1037/0022–3514.74.4.967

Becker, M., Luedtke, O., Trautwein, U., Koeller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology*, *104*(3), 682–699. doi: 10.1037/a0027608

Becker, S. O., Egger, P. H., & von Ehrlich, M. (2012). Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review*, *56(*4), 648–668. doi: 10.1016/j.euroecorev.2012.03.001

Benbow, C. P. (1983). Adolescence of the mathematically precocious : A five year longitudinal study. In C. P. Benbow & J. C. Stanley, (Eds.), *Academic precocity: Aspects of its development* (pp. 9–29). Maryland: The Johns Hopkins University Press.

Benbow, C. P., & Stanley, J. C. (1996). Inequity in equity: How "equity" can lead to inequity for high-potential students. *Psychology, Public Policy, & Law*, *2*(2), 249–292.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability: Their status 20 years later. *Psychological Science*, *11,* 474–480.

Bleske-Rechek, A., Lubinski, D., & Benbow, C. P. (2004). Meeting the educational needs of special populations: Advanced Placement's role in developing exceptional human capital. *Psychological Science, 15,* 217–224.

Borland, J. (1989). *Planning and implementing programs for the gifted*. New York: Teachers College Press.

Bowman, A.W., & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford University Press, Oxford.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statisistical Society*, *Series B.*, *26*, 211–246.

Breusch, T.S. & Pagan, A.R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*, 1287–1294.

Brody, L. E., & Benbow, C. P. (1987). Accelerative strategies: How effective are they for the gifted? *Gifted Child Quarterly*, *31*, 105–110.

Campbell, D. T., & Boruch, R. F. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. Bennett & A. Lumsdaine (Eds.), *Evaluation and experiment: Some* *critical issues in assessing social programs* (pp. 195–296). San Diego, CA: Academic Press.

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life*. New York: Russell Sage Foundation.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasiexperimental designs for research*. Chicago: Rand McNally.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*(2), 319–333. doi: http://dx.doi.org/10.1037/0022–3514.67.2.319

Cattell, R. B., & Butcher, H. J. (1968). *The prediction of achievement and creativity*. Indianapolis: Bobbs-Merrill.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colangelo, N., Assouline, S. G., & Gross, M. U. M. (Eds.). (2004). *A nation deceived: How schools hold back America's brightest students*. Iowa City: University of Iowa.

Colangelo, N., Assouline, S. G., & Lupkowski-Shoplik, A. E. (2004). Whole-grade acceleration. In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived: How schools hold back America's brightest students* (pp. 77–86). Iowa City: University of Iowa Press.

Compas, B. E., Connor-Smith, J., Saltzman, H., Thomsen, A. H., & Wadsworth, M. E. (2001). Coping with stress during childhood and adolescence: Problems, progress, and potential in theory and research. *Psychological Bulletin*, *127*(1), 87–127. doi: http://dx.doi.org/10.1037/0033–2909.127.1.87

Cornell, D. G., Callahan, C. M., & Loyd, B. H. (1991a). Socio-emotional adjustment of adolescent girls enrolled in a residential acceleration program. *Gifted Child Quarterly*, *35*, 58–66.

Cornell, D. G., Callahan, C. M., & Loyd, B. H. (1991b). Research on early college entrance: A few more adjustments are needed. *Gifted Child Quarterly*, *35*, 71–72.

Cornell, D. G., Callahan, C. M., Bassin, L. E., & Ramsay, S. G. (1991). Affective development in accelerated students. In W. T. Southern & E. D. Jones (Eds.), *The academic acceleration of gifted children* (pp. 74–101). New York: Teachers College.

Crick, N., & Dodge, K. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York: HarperCollins Publishers.

Dai, D. Y. (2004). How universal is the big-fish-little-pond effect? *American Psychologist*, *59*, 267–268.

Dark, V. J., & Benbow, C. (1990). Mathematically talented students show enhanced problem translation and enhanced short-term memory for digit and spatial information. *Journal of Educational Psychology*, *82*, 420–429.

Dark, V. J., & Benbow, C. P. (1991). Differential enhancement of working memory with mathematical and vertical precocity. *Journal of Educational Psychology*, *83*, 48–60.

Dark, V. J., & Benbow, C. P. (1994). Type of stimulus mediates the relationship between performance and type of precocity. *Intelligence*, *19*, 337–357.

Daurio, S. P. (1979). Educational enrichment versus acceleration: A review of the literature. In W. C. George, S. J. Cohn, and J. C. Stanley (Eds.). *Educating the gifted: Acceleration and enrichment* (pp. 13–63). Baltimore: Johns Hopkins University Press.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. doi: http://dx.doi.org/10.1207/S15327965PLI1104_01

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, *124*(2), 197–229. doi: http://dx.doi.org/10.1037/0033–2909.124.2.197

Depue, R. A., Luciana, M., Arbisi, P., Collins, P., & Leon, A. (1994). Dopamine and the structure of personality: Relation of agonist-induced dopamine activity to positive emotionality. *Journal of Personality and Social Psychology, 67*(3), 485–498. doi: http://dx.doi.org/10.1037/0022–3514.67.3.485

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. doi: http://dx.doi.org/10.1037/0033–2909.95.3.542

Diener, E., & Biswas-Diener, R. (2008) *Happiness: unlocking the mysteries of psychological wealth*. Malden, MA: Wiley-Blackwell.

Diener, E., Emmons, R.A., Larsen, R.J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.

Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology, 54*, 403–425. doi: http://dx.doi.org/10.1146/annurev.psych.54.101601.145056

Diener, E., Sandvik, E., Seidlitz, L., & Diener, M. (1993). The relationship between income and subjective well-being: Relative or absolute? *Social Indicators Research, 28*(3), 195–223. doi: http://dx.doi.org/10.1007/BF01079018

Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology, 69*(1), 130–141. doi: http://dx.doi.org/10.1037/0022–3514.69.1.130

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. doi: http://dx.doi.org/10.1037/0033–2909.125.2.276

Domestic Policy Council. (2006). *American competitiveness initiative: Leading the world in innovation* (Tech. Rep.). Washington, DC: Domestic Policy Council Office of Science and Technology.

Duncan, O. D. (1961). A socioeconomic index for all occupations. In J.Reiss, Jr. (Ed.), Occupations and social status (pp. 109–138). New York, NY: Free Press of Glencoe.

Engle, T. L. (1938). A study of the effects of school acceleration upon the personality and social adjustments of high-school and university students. *Journal of Educational Psychology*, *29*, 523–539.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100(3), 363–406. doi:http://dx.doi.org/10.1037/0033–

295X.100.3.363

Feng, P., Zhou, X.-H., Zou, Q.-M., Fan, M.-Y., & Li, X.-S. (2012). Generalized propensity score for estimating the average treatment effect of multiple treatments. *Statistics in Medicine*, *31*(7), 681–697. doi: 10.1002/sim.4168

Ferriman-Robertson, K., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science*, *19*, 346–351.

Fitts, W. H. (1965). *Tennessee Self-Concept Scale Manual*. Nashville, TN: Counselor Recordings and Tests.

Flores, C. A., Flores-Lagunes, A., Gonzalez, A., & Neumann, T. C. (2012). Estimating the effects of length of exposure to instruction in a training program: the case of job corps. *Review of Economics and Statistics*, *94*(1), 153–171. doi: 10.1162/REST_a_00177

Fowles, D. C. (1987). Application of a behavioral theory of motivation to the concepts of anxiety and impulsivity. *Journal of Research in Personality, 21*(4), 417–435. doi: http://dx.doi.org/10.1016/0092–6566(87)90030–4

Fowles, D. C. (1994). A motivational theory of psychopathology (pp. 181–238). Lincoln, NE, US: University of Nebraska Press, Lincoln, NE.

Fox, L. H., Benbow, C. P., & Perkins, S. (1983). An accelerated mathematics program for girls: A longitudinal evaluation (pp. 113–138). In C. P. Benbow & J. P. Stanley (Eds.), *Academic precocity: Aspects of its development*. Baltimore: Johns Hopkins University Press.

Friedman, T. (2005). *The world is flat*. New York: Farrar, Straus & Giroux.

Fund for the Advancement of Education of the Ford Foundation. (1957). *They went to college early*. New York, NY: Research Division of the Fund.

Gagné, F., & Gagnier, N. (2004). The socio-affective and academic impact of early entrance to school. *Roeper Review*, *26*, 128–139.

Gallagher, J. J. (1985). Peer acceptance of highly gifted children in elementary school. *The Elementary School Journal*, *58*(8), pp. 465–470.

Gallagher, J. J. (2004). No Child Left Behind and Gifted Education. *Roeper Review*, *26*(3), 121–123.

Gallagher, J., Harradine, C. C., Coleman, M. R. (1997). Challenge or boredom? Gifted students' views on their schooling, *Roeper Review*, *19*(3), 132–141.

Gibbons, F. X., Benbow, C. P., & Gerrard, M. (1994). From top dog to bottom half: Social comparison strategies in response to poor performance. *Journal of Personality and Social Psychology*, *67*, 638–652.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42.

Gough, H. G. (1969). *Manual for the California Psychological Inventory* (rev. ed.). Palo Alto, CA: Consulting Psychologists Press.

Gough, H.G., & Heilbrun, A.B. (1980). *The adjective checklist manual*. Palo Alto, CA: Consulting Psychologists Press, Inc.

Gross, M. U. M. (1989). The pursuit of excellence or the search for intimacy? The forced-choice dilemma of gifted youth. *Roeper Review*, *11*(4), 189–194.

Gross, M. U. M. (1993a). *Exceptionally gifted children*. London: Routledge.

Gross, M. U. M. (1993b). Nurturing the talents of exceptionally gifted individuals. In K. A. Heller, F. J. Monks, R. J. Sternberg, & R. F. Subotnik (Eds). *International handbook of giftedness and talent* (pp. 473–490). Oxford: Pergamon Press.

Gross, M. U. M. (2003). *Exceptionally gifted children* (2nd ed.). London: Routledge.

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity, 3*(1), 53–68.

Haas, L., Stargardt, T., & Schreyoegg, J. (2012). Cost-effectiveness of open versus laparoscopic appendectomy: a multilevel approach with propensity score matching. *European Journal of Health Economics*, *13*(5), 549–560. doi: 10.1007/s10198–011–0355–6

Han, W.-J., Miller, D. P., & Waldfogel, J. (2010). Parental Work Schedules and Adolescent Risky Behaviors. *Developmental Psychology*, *46*(5), 1245–1267. doi: 10.1037/a0020178

Hartup, W. W. (1983). Peer relations. In E. M. Hetherington (Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 103–198). New York: Wiley.

Hathaway, S. R., & McKinley, J. C. (1940). A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, *10*, 249–254.

Haugaard, J. J., & Tilly, C. (1988). Characteristics predicting children's responses to sexual encounters with other children. *Child Abuse and Neglect, 12*(2), 209–218. doi: http://dx.doi.org/10.1016/0145–2134(88)90029–4

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. doi: 10.1093/pan/mpl013

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. doi: 10.2307/2289064

Hollingworth, L. S. (1931). The child of very superior intelligence as a special problem in social adjustment. *Mental Hygiene*, *15*(1), 3–16.

Hollingworth, L. S. (1942). *Children above IQ 180: Their origin and development*. New York: World Books.

Honaker, J., King, G., & Blackwell, M. Amelia II: A Program for Missing Data. *Journal of Statistical Software*, *45*(7), 1–47.

Horton, N. J., & Kleinman, K. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *American Statistician*, *61*, 79 –90. doi:10.1198/ 000313007X172556

Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley and Sons.

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.

Imai, K., & Ratkovic, M. (in review). Covariate Balancing Propensity Score. *Journal of the Royal Statistical*

Society, Series B (Statistical Methodology).

Imai, K., & van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*(467), 854–866. doi: 10.1198/016214504000001187

Isaacs, M. L., & Duffus, L. R. (1995). Scholars' club: A culture of achievement among minority students. *The School Counselor*, *42*, 204–210.

Jackson, D. N. (1976). *Jackson Personality Inventory Manual*. Goshen, NY: Research Psychologists Press.

Jackson, N. E., Famiglietti, J., & Robinson, H. B. (1981). Kindergarten and first grade teachers' attitudes toward early entrants, intellectually advanced students, and average students. *Journal for the Education of the Gifted*, *4*, 132–142.

Janos, P. M. (1987). A fifty-year follow-up of Terman's youngest college students and IQ-matched comparison subjects. *Gifted Child Quarterly*, *31*, 55–58.

Janos, P. M., Robinson, N. M., & Lunneborg, C. E. (1989). Markedly early entrance to college: A multi-year comparative study of academic performance and psychological adjustment. *Journal of Higher Education*, *60*, 496–518.

Janos, P. M., Robinson, N. M., Carter, C., Chapel, A., Cufley, R., Curland, M., et al. (1988). Social relations of students who enter college early. *Gifted Child Quarterly*, *32*, 210–215.

Jin, S.-U., & Moon, S. M. (2006). A study of well-being and school satisfaction among academically talented students attending a science high school in Korea. *Gifted Child Quarterly*, *50*, 169 - 184.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The Core Self-Evaluations Scale (CSES): Development of a measure. *Personnel Psychology*, *56*, 303–331.

Jung, C. G. (1954). The gifted child. In H. Read, M. Fordham, & G. Adler (Eds.), *The collected works of C. G. Jung: Vol. 17. The development of personality* (pp. 133–148). Princeton, NJ: Princeton University Press.

Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussions). *Statistical Science*, *22*(4), 523–539.

Kasser, T., & Ryan, R. M. (1993). A dark side of the American dream: Correlates of financial success as a central life aspiration. *Journal of Personality and Social Psychology*, *65*, 410–422.

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin*, *22*, 80–87.

Kasser, V., & Ryan, R. M. (1999). The relation of psychological needs for autonomy and relatedness to vitality, well-being, and mortality in a nursing home. *Journal of Applied Social Psychology*, *29*, 935–454.

Kent, S. D. (1992). The effects of acceleration on the social and emotional development of gifted elementary students: A meta-analysis. *Dissertation Abstracts International*, *54*, 419-A. (University Microfilms No. AAT93–16362)

Kincaid, D. (1969). A study of highly gifted elementary pupils. *Gifted Child Quarterly*, *13*(4), 264–267.

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*(1), 49–69.

Klausmeier, H. J. (1963). Effects of accelerating bright older elementary pupils: A follow up. *Journal of Educational Psychology*, *54*, 165–171.

Klausmeister, H. J., & Ripple, R. E. (1962). Effects of accelerating bright older pupils from second to fourth grade. *Journal of Educational Psychology*, *53*(2), 93–100.

Kluve, J., Schneider, H., Uhlendorff, A., & Zhao, Z. (2012). Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society Series A-Statistics in Society*, *175*, 587–617. doi: 10.1111/j.1467–985X.2011.01000.x

Kolmogorov, A. N. (1933). *Sulla Determinazione Empirica di una Legge di Distribuzione Giornale dell'Istituto Italiano degli Attuari*,*4*, pp. 83–91.

Kulik, J. A. (2004). Meta-analytic studies of acceleration. In N. Colangelo, S. G. Assouline & M. U. M. Gross (Eds.), *A nation deceived* (pp. 13–22). Iowa City: University of Iowa.

Kulik, J. A., & Kulik, C. C. (1984). Effects of acceleration on students. *Review of Educational Research*, *54*, 409–425.

Learned, W. S., & Wood, B. D. (1938). The student and his knowledge. A report to the Carnegie Foundation on the results of the high school and college examinations of 1928, 1930, and 1932. New York: Carnegie Foundation for the Advancement of Teaching.

Lessinger, L. M., & Martinson, R. A. (1961). The use of the CPI with gifted pupils. *Personnel and Guidance Journal*, *39*, 572–75.

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, *55*, 137–150.

Lubinski, D., & Benbow, C. P. (2006). Study of Mathematically Precocious Youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives in Psychological Science*, *1*, 316–345.

Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology*, *86*, 718–729.

Lupkowski, A. E., Whitmore, M., & Ramsay, A. (1992). The impact of early entrance to college on self-esteem: A preliminary study. *Gifted Child Quarterly, 36*(2), 87–90.

Lynch, S. J. (1992). Fast-paced high school science for the academically talented: A six-year perspective. *Gifted Child Quarterly*, *36*, 147–154.

Ma, X. (2002). Early acceleration of mathematics students and its effect on growth in self-esteem: A longitudinal study. *International Review of Education*, *48*(6), 443–468.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. doi:http://dx.doi.org/10.1037/1082–989X.7.1.19

Makel, M. C., Lee, S., Olszewki-Kubilius, P., & Putallaz, M. (2012). Changing the pond, not the fish: Following high-ability students across different educational environments. *Journal of Educational Psychology, 104*(3), 778–792. doi: http://dx.doi.org/10.1037/a0027558

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*, 280–295.

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal*, *28*(2), 445–480. doi: http://dx.doi.org/10.2307/1162948

Marsh, H. W., & Hau, K-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*, 364–376.

Marsh, H. W., Hau, K., & Craven, R. (2004). The big-fish-little-pond effect stands up to scrutiny. *American Psychologist, 59*(4), 269–271. doi: http://dx.doi.org/10.1037/0003–066X.59.4.269

Marsh, H. W., Chessor, D., Craven, R., Roche, L. (1995). The Effects of Gifted and Talented Programs on Academic Self-Concept: The Big Fish Strikes Again. *American Education Research Journal*, *32*(2), 285–319.

Marsh, H. W., & Rowe, K. J. (1996). The negative effect of school-average ability on academic self-concept: An application of multilevel modeling. *Australian Journal of Education*, *40*, 65–87.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416.

Meehl, P. E. (1971). High school yearbooks: A reply to Schwartz. *Journal of Abnormal Psychology*, *77*(2), 143–148. doi: 10.1037/h0030750

Miller, R. V. (1956). Social status and socio-empathic differences among mentally superior, mentally typical, and mentally retarded children. *Exceptional Child*, *23*, 104–19.

Miller, G. E., Chen, E., & Parker, K. J. (2011). Psychological stress in childhood and susceptibility to the chronic diseases of aging: Moving toward a model of behavioral and biological mechanisms. *Psychological Bulletin, 137*(6), 959–997. doi: http://dx.doi.org/10.1037/a0024768

Miquelon, P., & Vallerand, R. J. (2008). Goal motives, well-being, and physical health: An integrative model. *Canadian Psychology/Psychologie Canadienne, 49*(3), 241–249. doi: http://dx.doi.org/10.1037/a0012759

Moodie, E. E. M., & Stephens, D. A. (2012). Estimation of dose-response functions for longitudinal data using the generalised propensity score. *Statistical Methods in Medical Research*, *21*(2), 149–166. doi: 10.1177/0962280209340213

National Academy of Sciences. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5* (Tech. Rep.). National Academies Press.

National Mathematics Advisory Panel. (2008). *Foundations for success: The national report of the national mathematics advisory panel.* (Tech. Rep.). Washington, D.C.: U.S. Department of Education.

Neihart, M. (2007). The socioaffective impact of acceleration and ability grouping: Recommendations for best practice. *Gifted Child Quarterly, 51*(4), 330–341. doi: http://dx.doi.org/10.1177/0016986207306319

Newcomb, A. F., & Bagwell, C. L. (1995). Children's friendship relations: A meta-analytic review. *Psychological Bulletin, 117*(2), 306–347. doi: http://dx.doi.org/10.1037/0033–2909.117.2.306

Neyman, J. (1990). On the application of probability theory to agricultural experiments: Essays on principles, Section 9. *Statistical Science*, *5*, 465–480. (Original work published in 1923)

Noble, K. D., Robinson, N. M., & Gunderson, S. (1993). All rivers lead to the sea: A follow-up study of young adults. *Roeper Review*, *15*, 124–129.

Noeth, R. J. (1978). *Washington Pre-College Counselor's Guide*. Seattle, Washington: Washington Pre-College Testing Program.

O'Leary, A. (1990). Stress, emotion, and human immune function. *Psychological Bulletin, 108*(3), 363–382. doi: http://dx.doi.org/10.1037/0033–2909.108.3.363

Okun, M. A., & George, L. K. (1984). Physician- and self-ratings of health, neuroticism and subjective well-being among men and women. *Personality and Individual Differences, 5*(5), 533–539. doi: http://dx.doi.org/10.1016/0191–8869(84)90027–8

Olszewski-Kubilus, P. (Ed.).(2007). Best Practices in Gifted Education (Special Issue). *Gifted Child Quarterly*, *51*(4).

Painter, F: (1976). *Gifted children: A research study*. Knebworth, England: Pullen Publications.

Park, G., Lubinski, D., & Benbow, C. P. (2013). When less is more: Effects of grade skipping on adult STEM productivity among mathematically precocious adolescents. *Journal of Educational Psychology*, *105*(1), 176–198.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin, 102*(3), 357–389. doi: http://dx.doi.org/10.1037/0033–2909.102.3.357

Peel, M. J., & Makepeace, G. H. Differential Audit Quality, Propensity Score Matching and Rosenbaum Bounds for Confounding Variables. *Journal of Business Finance and Accounting*, *39*(5–6), 606–648. doi: 10.1111/j.1468–5957.2012.02287.x

Plucker, J. A., Robinson, N. M., Greenspon, T. S., Feldhusen, J. F., McCoach, D. B., & Subotnik, R. R. (2004). It’s not how the pond makes you feel, but rather how high you can jump. *American Psychologist*, *59*, 268–269.

Pollins, L. D. (1983). The effects of acceleration on the social and emotional development of gifted students. In C. P. Benbow & J. C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 160–178). Baltimore, MD: Johns Hopkins University Press.

Pressey, S. L. (1946). Age of college graduation and success in adult life. *Journal of Applied Psychology*, *30*, 226–233.

Pressey, S. L. (1949). *Educational acceleration: appraisals and basic problems*. Columbus: Ohio State University Press.

Pressey, S. L. (1967). "Fordling" accelerates ten years after. *Journal of Counseling Psychology*, *14*, 73–80.

Rambo, K. E., & McCoach, D. B. (2012). Teacher attitudes toward subject-specific acceleration: Instrument development and validation. *Journal for the Education of the Gifted, 35*(2), 129–152. doi: http://dx.doi.org/10.1177/0162353212440591

Redd, A. D., Mullis, C. E., Serwadda, D., Kong, X., Martens, C., Ricklefs, S. M., et al. (2012). The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *Journal of Infectious Diseases*, *206*(2), 267–274. doi: 10.1093/infdis/jis325

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*, *26*, 419–435.

Reis, S. M., & McCoach, D. B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly*, *44*(3), 152–170.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. doi: http://dx.doi.org/10.1037/a0026838

Richardson, T. M., & Benbow, C. P. (1990). Long-term effects of acceleration on the social-emotional adjustment of mathematically precocious youths. *Journal of Educational Psychology*, *82*, 464–470.

Rimm, S. (1986). *The underachievement syndrome: Causes and cures*. Watertown, WI: Apple Publishing Company.

Rimm, S. (1995). Why bright kids get poor grades and what you can do about it. New York, NY: Crown Publishing Group.

Robinson, N. M., & Janos, P. M. (1986). Psychological adjustment in a college-level program of marked academic acceleration. *Journal of Youth and Adolescence*, *15*(1), 51–60.

Robinson, N. M., & Noble, K. D. (1991). Social-emotional development and adjustment of gifted children. In M.C. Wang, M.C. Reynolds, & H.J.Walberg (Eds.) *Handbook of special education: Research and Practice, Volume 4: Emerging programs* (pp. 57–76). New York: Pergamon Press.

Rogers, K. B. (1991). A best-evidence synthesis of research on types of accelerative programs for gifted students. *Dissertation Abstracts International*, *52*, 796-A. (University Microfi lms No. AAT91–22206).

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

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B*, *67*, 515–530.

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

Rowan, C. G., Brunelli, S. M., Munson, J., Flory, J., Reese, P. P., Hennessy, S., et al. (2012). Clinical importance of the drug interaction between statins and CYP3A4 inhibitors: a retrospective cohort study in The Health Improvement Network. *Pharmacoepidemiology and Drug Safety*, *21*(5), 494–506. doi: 10.1002/pds.3199

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. B. (1980). Discussion of "Randomization Analysis of Experimental Data in the Fisher Randomization Test," by D. Basu. *Journal of the American Statistical Association*, *75*, 591–593.

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

Rubin, D. B. (2004). Teaching statistical inference for causal effects in experimental and observational studies. *Journal of Educational and Behavioral Statistics*, *29*, 343–367.

Rubin, D. B. (2010). Reflections Stimulated by the Comments of Shadish (2010) and West and Thoemmes (2010). *Psychological Methods*, *15*(1), 38–46.

Rubin, D.B. (1996). Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association*, *91*, 473–489. doi:10.2307/2291635

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*, 141–166. doi: http://dx.doi.org/10.1146/annurev.psych.52.1.141

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719–727. doi: http://dx.doi.org/10.1037/0022–3514.69.4.719

Ryff, C. D., & Singer, B. H. (2006). Best news yet on the six-factor model of well-being. *Social Science Research, 35*(4), 1103–1119. doi: http://dx.doi.org/10.1016/j.ssresearch.2006.01.002

Sayler, M. F., & Brookshire, W. K. (1993). Social, emotional, and behavioral adjustment of accelerated students, students in gifted classes, and regular students in eighth grade. *Gifted Child Quarterly*, *37*, 150–154.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London, England: Chapman & Hall.

Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research*, *33*(4), 545–571.

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*, 403–419. doi:10.1037/a0013838

Sekhon, J.S. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization. *Journal of Statistical Software*, *42*(7), 1–52.

Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods*, *15*(1), 3–17. doi: 10.1037/a0015916

Sheldon, K. M., Ryan, R. M., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*, 1270–1279.

Šidàk, Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*, 626–633.

Silverman, L. K. (1997). The contruct of asynchronous development. *Peabody Journal of Education*, *72*(3/4), 36–58.

Smirnov, N. V. (1948). Tables for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, *19*, 279.

Smith, E. (1984). Giftedness on demand in every classroom. *Gifted Education International*, *2*, 142–144.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1–2), 305–353.

Snodgrass, G. M., Blokland, A. A. J., Haviland, A., Nieuwbeerta, P., & Nagin, D. S. (2011). Does the time cause the crime? An examination of the relationship between time served and reoffending in the Netherlands. *Criminology*, *49*(4), 1149–1194. doi: 10.1111/j.1745–9125.2011.00254.x

Southern, W. T., & Jones, E. D. (Eds.) (1991). *The academic acceleration of gifted children*. New York, NY: Teachers College Press.

Southern, W. T., & Jones, E. D. (2004). Types of acceleration: Dimensions and issues. In N. Colangelo, S. Assouline, & M. U. M. Gross (Eds.), *A nation deceived: How schools hold back America's brightest students* (Vol. 2, pp. 5–12). Iowa City, IA: The University of Iowa, The Connie Belin & Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Southern, W. T., Jones, E. D., & Fiscus, E. D. (1989). Practitioner objections to the academic acceleration of gifted children. *Gifted Child Quarterly*, *33*(1), 29–35.

Springer, K. W., & Hauser, R. M. (2006). An assessment of the construct validity of Ryff's scales of psychological well-being: Method, mode, and measurement effects. *Social Science Research, 35*(4), 1080–1102. doi: http://dx.doi.org/10.1016/j.ssresearch.2005.07.004

Springer, K. W., Hauser, R. M., & Freese, J. (2006). Bad news indeed for Ryff's six-factor model of well-being. *Social Science Research, 35*(4), 1120–1131. doi: http://dx.doi.org/10.1016/j.ssresearch.2006.01.003

Stanley, J. C. (1973). Accelerating the educational progress of intellectually gifted youths. *Educational Psychologist*, *10*, 133–146.

Stanley, J. C. (1991). Critique of "Socioemotional adjustment of adolescent girls enrolled in a residential acceleration program. *Gifted Child Quarterly*, *35*, 67–70.

Stanley, J. C. (2000). Helping students learn only what they don't already know. *Psychology, Public Policy, and Law*, *6*(1), 216–222.

Stanley, J. C, & Stanley, B. S. K. (1986). High-school biology, chemistry, and physics learned well in three weeks. *Journal of Research in Science Teaching*, *23*, 237–250.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, *134*(1), 138–161. doi: http://dx.doi.org/10.1037/0033–2909.134.1.138

Steenbergen-Hu, S., & Moon, S. M. (2011). The Effects of Acceleration on High-Ability Learners: A Meta-Analysis. *Gifted Child Quarterly*, *55*, 39–53.

Stigler, J. W., & Miller, K. F. (1993). A good match is hard to find: Comment on Mayer, Tajika, and Stanley (1991). *Journal of Educational Psychology*, *85*(3), 554–559. doi: http://dx.doi.org/10.1037/0022–0663.85.3.554

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54.

Swiatek, M. A., & Benbow, C. P. (1991a). Ten-year longitudinal follow-up of ability-matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology*, *83*, 528–538.

Swiatek, M.A., & Benbow, C. P. (1991b). A ten-year longitudinal follow-up of participants in a fast-paced Mathematics course. *Journal for Research Mathematics Education*, *22*, 138–150.

Trautwein, U., Lu¨dtke, O., Marsh, H. W., & Nagy, G. (2009). Within school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, *101*, 853–866. doi:10.1037/a0016306

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect.

Psychological Science, 10(4), 297–303. doi: http://dx.doi.org/10.1111/1467–9280.00157

Terman, L. M. (1926). Mental and physical traits of a thousand gifted children. Vol. 1. Genetic studies of genius (2nd ed.). Stanford, CA: Stanford University Press.

Terman, L. M., & Oden, M. H. (1947). *The gifted child grows up. Vol 4. Genetic studies of genius*. Stanford, CA: Stanford University Press.

Thorndike, R. L. (1940). Performance of gifted children on a test of developmental age. *Journal of Psychology*, *9*, 337–343.

Tomarken, A. J., & Keener, A. D. (1998). Frontal brain asymmetry and depression: A self-regulatory perspective. *Cognition and Emotion, 12*(3), 387–420. doi: http://dx.doi.org/10.1080/026999398379655

U.S. Department of Commerce. (2012). *The competitiveness and innovative capacity of the United States*. Retrieved from http://www.commerce.gov/sites/default/files/documents/2012/january/competes_010511_0.pdf

VanTassel-Baska, J. (1983). Illinois' state-wide replication of the Johns Hopkins Study of Mathematically Precocious Youth. In C. P. Benbow & J. C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 179–191). Baltimore: Johns Hopkins University Press.

Vialle, W., Ashton, T., Carlton, G., & Rankin, F. (2001). Acceleration: A coat of many colours. *Roeper Review*, *24*, 14–19.

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*, *102*, 860–871.

Watson, D., Gamez, W., & Simms, L. J. (2005). Basic dimensions of temperament and their relation to anxiety and depression: A symptom-based perspective. *Journal of Research in Personality, 39*(1), 46–66. doi: http://dx.doi.org/10.1016/j.jrp.2004.09.006

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219–235. doi: http://dx.doi.org/10.1037/0033–2909.98.2.219

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*(5), 820–838. doi: http://dx.doi.org/10.1037/0022–3514.76.5.820

What Works Clearinghouse. (2009). *What Works Clearinghouse procedures and standards handbook* (Version 2.0). Washington, DC: U.S.Department of Education, Institute of Education Sciences.

Wilson, W. (1967). Correlates of avowed happiness. *Psychological Bulletin*, *67*, 294–306.

Wilson, W. R. (1960). An attempt to determine some correlates and dimensions of hedonic tone (Doctoral dissertation, Northwestern University, 1960). *Dissertation Abstracts*, *22*, 2814. (University Microfilms No. 60–6588)

Witter, R. A., Okun, M. A., Stock, W. A., & Haring, M. J. (1984). Education and subjective well-being: A meta-analysis. *Educational Evaluation and Policy Analysis, 6*(2), 165–173. doi: http://dx.doi.org/10.2307/1163911

Wood, S., Portman, T. A. A., Cigrand, D. L., & Colangelo, N. (2010). School counselors' perceptions and experience with acceleration as a program option for gifted and talented students. *Gifted Child Quarterly*, *54*(3), 168–178. doi: http://dx.doi.org/10.1177/0016986210367940

Yuan, L. L. (2010). Estimating the effects of excess nutrients on stream invertebrates from observational data. *Ecological Applications*, *20*(1), 110–125. doi: 10.1890/08–1750.1

Zeidner, M., & Schleyer, E. (1999). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology*, *24*, 305–329.