Economics of Education Review 70 (2019) 173–191

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Promoting cognitive and soft skills acquisition in a disadvantaged public school system: Evidence from the Nurture thru Nature randomized experiment

Radha Jagannathan[a,*], Michael J. Camasso[b], Maia Delacalle[a]

[a] Edward J. Bloustein School of Planning & Public Policy, Rutgers University, USA
[b] School of Environmental & Biological Sciences, Rutgers University, USA

ARTICLE INFO

Keywords:
Cognitive skills
Non-cognitive skills
Randomized experiment
STEM enhancement program
Nurture thru Nature (NTN) program
Multi-level models

JEL Classification:
I260
I200
J490

ABSTRACT

It is widely acknowledged that our public schools have failed to produce sufficient levels of high quality STEM education. The mathematics and science performance of minority and disadvantaged students has been especially troubling with blacks and Hispanics substantially underrepresented in the STEM labor market. In this paper we examine the impacts of a STEM enhancement program called Nurture thru Nature (NtN) on the cognitive (academic grades) and soft skills development of 139 elementary school students who attended the program over an eight year period (2010–2017). Utilizing a randomized experimental design or RCT with a control group of 491 elementary school students, we find that NtN slows the deterioration in students' math and science grades relative to controls and improves soft skills such as conscientiousness, higher order thinking, empathy, and pro-social behavior.

1. Introduction

In a recent report issued by the Business Roundtable (2017) a troubling picture of the U.S. workforce is presented that depicts too few workers with specialized STEM skills and abilities. Especially alarming are the deficits found in math and science knowledge and skills, computer skills, mechanical skills and operations monitoring capabilities (Rothwell, 2013; Stewart, 2018). The same Business Roundtable report notes that an increasing number of job applicants lack fundamental employability skills as well, i.e., the ability to "communicate effectively, read technical manuals, work successfully in teams and participate in complex problem solving" (p.1). This set of skills along with clarity of oral expression, listening skills, the ability to motivate others and conscientiousness are believed by many labor economists and business leaders to comprise a panoply of "soft" or non-cognitive abilities, critical for successful labor force participation (Attanasio, 2015; Cunha & Heckman, 2008; Heckman, Stixrud, & Urzua, 2006; Ibarraran, Ripani, Tabooda, Villa, & Garcia, 2014; Stewart, 2018). Data published by the Manufacturing Institute (2016) and by the Bureau of Labor Statistics indicate that the scarcity of math, science and soft skills is most pronounced for African Americans and Hispanics (Committee on Highly Successful Schools or Programs, 2011; National Science Foundation, 2016), excluding many of these individuals from the higher wages and job security that accompany employment in STEM occupations.

While many gaps remain in our knowledge of human capital formation, two things have become increasingly clear: (1) skill and ability formation exhibit dynamic complementarities and interactions, i.e., abilities beget skills and more skills increase abilities (Attanasio, Meghir, & Nix, 2017; Heckman, 2000); and (2) efforts to build an individual's technical and soft skill portfolios yield their highest returns to investment when introduced at a young age, certainly well before entry to high school (Bagiati, Yoon, Evangelou, & Ngambeki, 2010; Cunha & Heckman, 2008; Maltese & Tai, 2011). Tai, Liu, Maltese, and Fan (2006), for example, find that an early interest in pursuing a science-related career increased a child's chances of actually completing a science or engineering degree by about three and a half times. And increasingly the provision of "extra-school" robust natural and environmental science teaching, introduced at the elementary school level, has been identified as a promising pathway to STEM and green careers (Clarke, 2012; Royal Horticulture Society, 2010; U.S. Department of Education, Green Ribbon Schools, 2018).

In this paper, we examine how a natural science focused STEM enhancement program titled Nurture thru Nature (NtN) can increase

* Corresponding author.
E-mail addresses: radha@rutgers.edu (R. Jagannathan), mcamasso@sebs.rutgers.edu (M.J. Camasso).

https://doi.org/10.1016/j.econedurev.2019.04.005
Received 9 July 2018; Received in revised form 12 April 2019; Accepted 12 April 2019
Available online 13 April 2019
0272-7757/

the technical and socio-emotional skills of disadvantaged black and Hispanic students from seven elementary schools in Central New Jersey, who were randomly assigned to treatment and control groups (RCT). Inspired by the active learning philosophy of John Dewey (1976, 1990) and its extension in the forms of the outdoor education movement (Ord & Leather, 2011; Quay & Seaman, 2012) and wonders of nature teaching model (Camasso & Jagannathan, 2017b; Jagannathan et al., 2018), NtN focuses on the concomitant improvement of STEM cognitive skills and a set of socio-emotional or 'soft skills' that help facilitate the acquisition of the former.

1.1. Insufficient STEM education and the extra-school program response

In a 2015 report to Congress, the Committee on Equal Opportunities in Science and Engineering (2015) identified poor elementary and high school education as one of the major reasons that STEM careers are ignored, dismissed, or abandoned. Experts gathered by the National Academies of Science (Committee on Highly Successful Schools or Programs for K-12 STEM Education, 2011) concluded that in too many public schools there is a lack of authentic learning activities in STEM subjects, little time for science in elementary school, inadequate teacher preparation in STEM content and insufficient collaboration between K-12 and higher education institutions to smooth student transitions from high school to college. The failure of public schools to prepare students for the millions of unfilled STEM jobs has also been acknowledged by Congress (U.S. Congress Joint Economic Committee Report, 2012), by the U.S. Department of Commerce (2011), and the U.S. Department of Education (2008).

Poor preparation for careers in STEM occupations is especially acute in urban public schools. A recent Science and Engineering Indicators report released by the National Science Board (2016) concludes that students in disadvantaged school districts are most affected by deficiencies in STEM education. The Board concluded that "fully certified mathematics and science teachers were less prevalent in high minority and high poverty schools" (p.5) and that a lower proportion of math and science teachers held in-field degrees and had extensive teaching experience. A number of careful analyses of academic achievement differences in disadvantaged and more privileged schools, moreover, have yielded three important insights, viz., educational deficits are cumulative, they are accelerated in the summer, and these accumulating deficits are not limited to cognitive knowledge and skills. Fryer and Levitt (2004), Heckman (2013), and Heckman and Masterov (2007), among others, report that the already substantial differences in human capital investments between disadvantaged and more privileged students observed at entry to school increase with age. In their 25-year study of achievement in STEM education, Wai, Lubinski, Benbow, and Steiger (2010) conclude that disparities in science and math performance emerge early in elementary school and worsen over time. By tenth grade, black and Hispanic students are more likely than their white and Asian peers to filter into low education tracks and less likely to pursue STEM courses. Hill (2017) notes that growth curve analyses indicate learning increases more in elementary school than it does in middle school, and this deceleration is most pronounced in disadvantaged school districts. Because of this accumulating effect and the widening achievement gap it creates, labor economists and child development professionals have called for "pre-distribution" or early intervention strategies of investment. Heckman (2013) sums up this growing consensus when he asserts that "programs targeted toward the adolescent years [or later] of disadvantaged youth face an equity-efficiency tradeoff that programs targeted toward the earlier years of the lives of disadvantaged children avoid." (p.40).

It is also becoming more apparent that increases in the disadvantaged-privileged and white-minority academic achievement gaps are not steady state; rather, abrupt increases in the size of the gaps occur after each academic year and summer recess (Alexander, Entwisle, & Olsen, 2007; McCombs et al., 2011). Hanushek and Rivkin (2009) state that a consistent finding in the research literature is the phenomenon of "summer fallback" which suggests that while learning during the school year might, on average, be the same for white and minority students, the amount of learning in the summer months heavily favors white students (p.370).

Lastly, learning deficits between privileged and disadvantaged students have been shown to occur in socio-emotional as well as cognitive skills (Coleman, 1990; Heckman, 2000; Heckman & Kautz, 2012; Attanasio, 2015; Goldin, 2016; Ibarraran et al., 2014). Carneiro and Heckman (2003) maintain that a series of soft or "civic skills," e.g., perseverance, attentiveness, motivation, self-confidence, self-discipline, trustworthiness and dependability, are developed early in a child's life and are critical for success in school, the labor market, and life. What's more, these non-cognitive skills serve to promote the acquisition of cognitive skills early in a child's development; the relationship does not appear to be reciprocal, however (Cunha & Heckman, 2008; Cunha, Heckman, & Schennach, 2010).

At least since 1983, when the National Commission on Excellence in Education (1983) released its report *A Nation at Risk: The Imperative for Educational Reform*, a major response to the problem of insufficient math and science preparation in our public schools has been the extra-school program. These interventions have taken several forms, e.g., science and math suffused curriculum, after-school education and tutoring, and summer immersion; and have been implemented individually or in combination. Extra-school interventions also vary with respect to the point(s) of introduction into students' development stage with some programs like Nurture thru Nature in third-fourth grade and others, notably Career Academies (Steinberg, 1998), Early College High School Initiative (ECHSI) (Jobs for the Future, 2017), Mathematics, Engineering, and Science Achievement (MESA) (Alvarado & Muniz, 2018), and Bridge-to-Employment (BTE) (FHI-360, 2017) initiating programming in junior high school.

1.1.1. Science-suffused curriculum

One major effort to provide STEM opportunities through the distribution of science and technology curriculum is the work of the not-for-profit Project Lead the Way (PLTW). Over nearly three decades PLTW has provided hands-on STEM curriculum to over 10,000 elementary and middle schools from across the country. In Tai's (2012) review of 16 studies that have evaluated PLTW impact, the reviewer reports evidence of higher math or science scores in eleven of these evaluations. Lieberman and Hoody (1998) developed and tested the influence of a curriculum they call the Environment as an Integrating Context (EIC). Evaluations conducted in 14 schools by these educators indicate that elementary and middle school students exposed to EIC had higher math and science grades than students who were taught with a traditional curriculum. These students were also reported to have developed a variety of social skills (Lieberman & Hoody, 1998). Bagiati et al. (2010) examine the use of preK-12 engineering education materials that can be obtained on open websites. These researchers find that only in 4 percent of the cases they examined was this material used to enhance the curriculum of elementary school students – its utilization was more widespread in high schools where some positive effects on student math performance were found. Overall, the effectiveness of enhanced curriculum in improving STEM performance in disadvantaged schools is thin with RCT evaluations extremely rare.

1.1.2. After-school programs

The work of after-school tutoring and homework programs like Big Brothers/Big Sisters, and Boys and Girls Clubs, are widely applauded for their efforts to improve the math and science grades of disadvantaged and minority youth (Fashola, 1998; Heckman, 2000; Springer & Diffily, 2012). One example of a promising after-school program focusing on STEM is the "Science Club," a partnership between Boys and Girls Club of Chicago and Northwestern University.

Employing a RCT evaluation design, researchers report that low income middle school students taught by Northwestern graduate students were more likely than controls to express positive attitudes toward science and a STEM career, exhibited more pro-social behaviors, and were more likely to report an improvement in science skills (Krishnamurti, Ballard, & Noam, 2014). The nation-wide 4-H Tech Wizards after-school program, sponsored jointly by the National 4-H Council and the Office of Juvenile Justice and Delinquency Prevention, provides students from disadvantaged as well as privileged school districts with opportunities to master skills in website development, video and podcast productions, GIS and GPS technologies and LEGO robotics. There is some evidence that the program has generated interest in pursuing STEM careers but this evidence is largely anecdotal (Boscia, 2013; National 4-H Council, 2016).

When enhanced STEM curriculum and after-school programs (with a focus on math and/or science) are carefully examined for their efficacy in improving specific STEM knowledge and abilities using experimental or strong quasi-experimental evaluation designs, the issue remains far from being settled. Hollister (2003), for example, concludes that we do not know very much about such efforts, a viewpoint echoed by Lauer et al. (2006) and Levine and Zimmerman (2010). Perhaps the most comprehensive evaluation of a program designed to improve the academic performance of elementary school students through a program of before and after school teaching and tutoring ($n = 2308$) is the impact assessment of the 21st Century Community Learning Centers (CCLC) Program conducted by Mathematic Policy Research from 2000 to 2003 (James-Burdumy, Dynarski, Moor, Deke, & Mansfield, 2005). Employing a RCT these researchers found no effects on math or reading test scores or on science, math, or reading grades. Since this report, however, several states including Texas, New Jersey, and Washington have reported the 21st CCLC has increased state assessment scores in math and reading (American Institutes for Research, 2018). A study conducted by Public/Private Ventures of ten Big Brothers/Big Sisters agencies found that positive effects on science classwork and homework largely disappeared at a one year follow-up. Similarly, a Mathematica evaluation of the Quantum Opportunities Program (QOP), employing a RCT, reports disappointing results with no significant differences between QOP and control groups after a follow-up on either achievement tests or grades (Levine & Zimmerman, 2010) despite earlier reports of impact on grade point averages (GPA) and graduation rates (Carneiro & Heckman, 2003). RCTs of Career Academies (Kemple & Willner, 2008) and ECHSI (Berger, Turk-Bicakci, & Garett, 2013) also report little impact of these programs on math or science grades, but report some evidence on soft skill development at the junior and senior high school levels. In sum, the evidence for after school teaching, tutoring and/or mentoring on STEM cognitive skills is rather thin with the evidence of soft skill development somewhat stronger.

1.1.3. Summer programs

Intensive science and math programs offered to elementary and middle school students in July and August, are increasingly being used to reinforce STEM technical skill development in disadvantaged school districts. Especially popular are summer gardening and nature initiatives such as Education Outside (San Francisco), the Boston Schoolyard Initiative and the OSSE School Garden Program (Washington, DC) (Hirschi, 2005). Research by Klemmer, Waliczek, and Zajicek (2005) and Smith and Matsenbocker (2005) reports that hands-on instruction around school gardens can increase science achievement test scores of elementary school children. Evaluations of the Building Educational Leaders for Life (BELL) conducted by the Urban Institute find some "suggestive" evidence that this summer program can improve student math and reading scores (Somers, Welbeck, Grossman, & Gooden, 2015). In their review of programs designed to reduce "summer learning loss" McCombs et al. (2011), however, could find only four programs that were evaluated using an experimental design and these demonstrated an average effect size of 0.14 standard deviation units.

In summary, while enhanced STEM curriculum, after-school programs and summer interventions have in some cases shown promise (Berger et al., 2013; Carneiro & Heckman, 2003; Heckman, 2000; Kemple & Willner, 2008; Krishnamurti et al., 2014; Somers et al., 2015), the impacts of these programs on science and math performance are often temporary and fleeting (Lauer et al., 2006; Levine & Zimmerman, 2010). The effect of these interventions, especially after-school programs, is frequently confined to the enhancement of interests in science, avoidance of risky behavior, and to a variety of other soft skills (Boscia, 2013; Heckman, 2000; Lauer et al., 2006). A number of evaluators (McCombs et al., 2011; Somers et al., 2015; Camasso & Jagannathan, 2017a) have attributed this dearth of positive effects to the problem of "underpowered" treatment, i.e., low dosages of program inputs due to student recruitment/retention issues, weak research designs, short treatment periods, and low student attendance and its concomitant, low levels of parental involvement. In designing the Nurture thru Nature (NtN) program its creators identified two additional reasons for the weak performance of many STEM focused, extra school programs, viz., (1) a failure to exploit young students' innate inquisitiveness in nature and the natural sciences and to use this as a pathway to teaching math and science as an extension of ecological intelligence; and (2) an indifference to the impact that a student's personal interests can have on math and science learning. In short NtN attempts to remedy the under appreciation for the powerful role that active learning (Dewey, 1990) and the 'wonders of nature' can play in the creation of STEM identities in young students.

1.2. Theoretical framework for the evaluation

Our conceptual departure point resides in the many papers in labor economics that characterize the process of human capital formation as a production function for child growth and development (Attanasio, 2015; Cunha & Heckman, 2008; Goldin, 2016; Heckman et al., 2006; Todd & Wolpin, 2003). This production function can be formalized, as do Attanasio (2015) and Cunha et al. (2010), with this notation:

$$HC_{i,t+1} = g_t(HC_{it}, X_{it}, Z_{it}, e_{it}{}^{HC}) \tag{1}$$

where: HC_{it} is the human capital of a child aged t, raised in household i. HC_{it} comprises cognitive skills, non-cognitive skills and health;

X_{it} represents investments in human capital including schooling, parental engagement, extra-school investments, etc.;

Z_{it} are home background (fixed or time varying) factors including resources availability, mother and father education, etc.; and $e_{it}{}^{HC}$ reflect inputs that are not observed or measured.

As Attanasio (2015) correctly notes, parents are assumed to maximize their utility

$$\underset{(C_{it}, X_{it})}{\text{Max}} \quad U = (C_{it}, H_{i,t+1}) \tag{2}$$

subject to

$$\begin{aligned} &C_{it} + P_t{}^x X_{it} = Y_{it} \\ &\text{and} \quad H_{i,t+1} = g_t(H_{it}, X_{it+1,} Z_{it}, e_{it}) \end{aligned} \tag{3}$$

where C_{it} is consumption; $P_t{}^x$ is a vector of prices for investment X_{it}; and Y_{it} are available resources

For the parents of disadvantaged children the investment function is characterized by limited choices (see Heckman, 2000) and by limited family resources, both of which increase their dependence on public schools for the human capital formation in their children. In the context of the NtN experiment, we consider portions of X_{it} (extra-school investment in the form of the NtN program) as exogenous, and the Z_{it} randomly distributed between the experimental and control group children.

1.3. The Nurture thru Nature (NtN) experiment

The NtN program was initiated in 2010 as the community partnership of Rutgers University, the Johnson & Johnson (J & J) pharmaceutical company headquartered in New Brunswick, New Jersey, and the New Brunswick Public School (NBPS) district to enhance the STEM knowledge and skills of disadvantaged minority students in the district. The program, which is currently in its 9th year of operation, is designed as a classical experiment with random assignment of students in their 3rd grade into the NtN and control groups. The program provides STEM enrichment activities to students selected into the program group until they graduate high school. Students typically meet 2–3 times a week for 3 h during the school year and 3 days a week for 7.5 h per day in July and August.

NtN uses a focus on natural and environmental sciences to build elementary, middle, and high school children's knowledge and interest in STEM subject and careers. The program traces its conceptual roots from John Dewey who introduced an occupational approach to early-year education and emphasized (1) the need to connect a student's prior knowledge and experience to learning experiences, and (2) the importance of situating learning in the “here and now,” providing opportunities to apply mathematics and science to everyday situations (Dewey, 1976; 1990). Like other STEM enrichment programs, NtN programming makes the assumption that participation will increase STEM exposure and skills; this, in turn, will translate into stronger orientations which will help students develop technical and soft skill competencies. What distinguishes NtN is a “naturalist approach” to gaining STEM “identities” and competencies through the intensive use of school naturescape/gardens, interactions with live plants and animals, outdoor and lab experimentation, and observation and taxonomic learning methods.

NtN has five core components, viz., (1) a grade-specific STEM-centered curriculum aligned with the curriculum taught by public school science and math teachers; (2) after-school and summer components that reinforce school curriculum; (3) math, science and language arts tutoring; (4) the use of garden/naturescape and indoor lab assets that extend indoor classroom; and (5) a commitment to keep parents aware and involved in their child(ren)’s math and science education. The NtN curriculum for the last two years of high school also incorporates school-to-college and school-to-career activities such as SAT prep classes, college visits, exposure to STEM careers through guest lectures from Johnson & Johnson professionals as well as scientists from Rutgers University, and internship opportunities at various Rutgers professional schools and local non-profits. An outline of NtN's STEM curriculum content areas and the context of where content is implemented appear in Fig. 1. A more detailed description of these curricular components and after-school and summer operations can be found in Camasso and Jagannathan (2017b) and Jagannathan et al. (2018).

The measures that NtN employs to determine if the program is, indeed, raising cognitive and soft skill competencies derive from two sources. The principal measure of technical performance is the academic year-end grades in math, science, and language arts reported by the school district to the New Jersey Department of Education. As pointed out by Borghans, Golsteyn, Heckman, and Humphries (2016), we acknowledge that it is also possible that these grades may capture some of the soft or non-cognitive skills. These grades are collected for each NtN and control group student at the end of 3rd grade (baseline) and at the end of each subsequent academic year until the student graduates, moves out of district, drops out of school, or otherwise attrits.

While grades produced for the purposes of student report cards and administrative reporting have obvious advantages over student self-reports, their utility in determining competency change over time can be seriously limited if curriculum and teacher grading is not subject to guidelines and standards that are rooted in vertical scaling and developmental appropriateness (Briggs & Peck, 2015; Iowa Testing Programs, 2018). Student grading in New Jersey public schools has been guided since the early 1990s by the Core Curriculum Content Standards in nine subject areas including math, English language arts, and science. In 2010 Common Core standards replaced the previous standards in math and language arts (State of New Jersey, 2018). Both sets of standards are based on the “learning progression hypothesis” that is used to direct curriculum development and teacher evaluations of student performance (State of New Jersey, 2018).

The determination of NtN's impact on soft skill acquisition is derived from student self-reports on the NtN Knowledge, Skills and Abilities Inventory (NtN-KSAI). As in the case of academic grades, information on the NtN-KSAI is collected at the end of 3rd grade (baseline) and at the end of each subsequent school year for both NtN and control group students. One question, Question 4 (which is provided in Appendix A) asks students to rate how good they are at performing 20 skills. This skills list was informed by previous work conducted by Garcia (2014); Gutman and Schoon, 2013; Judge, Higgins, Thoresen, & Barrick, 1999; Platt (2008); and Marsh, Richards, & Barnes, 1986).

As a self-rating instrument the KSAI is subject to the same reference and social desirability biases that can affect other questionnaires which rely on self-reports. KSAI attempts to minimize these sources of measurement error by focusing items on actual observable behaviors rather than on the endorsement of statements (Center for the Economics of Human Development, 2015; Judge et al., 1999; Soares, Babb, Diener, Gates, & Ignatowski, 2017). In their review of 244 instruments that have been used to measure soft skills within the domains workforce success, violence prevention and sexual health, Galloway, Lippman, Burke, Diener, and Gates (2017) find that 9 of the 10 instruments with the highest levels of reported reliability and validity employed self-rating or reporting. These same researchers, in addition to others, report that the soft skills with the strongest empirical links to career and workforce success are social skills like teamwork and communication, empathy, dependability and responsibility, complex problem solving, goal orientation, and self-control (Galloway et al., 2017; Garcia, 2014; Judge et al., 1999; Marsh et al., 1986).

The 20 item KSAI embeds 6 technical skills (Reading and understanding written text/instructions, Writing reports, Testing ideas about science, Solving math problems, Using computers, and Conducting science labs/experiments) with 14 items designed to measure soft skills found to have strong connections to educational and career success, i.e., social skills, higher-order thinking skills, conscientiousness, communication and teamwork, and positive attitude (Garcia, 2014; Judge et al., 1999; Marsh et al., 1986; Platt, 2008). Principal components factor analysis (with orthogonal rotation and Eigen value cutoff of 1.0) when applied to the KSAI reveals a four factor solution. The first dimension comprises the 6 technical skills identified above. We do not include this technical skills factor in our subsequent analyses of impact inasmuch as we have a more direct measure of these competencies in the form of academic grades. Factor 2 is dominated by loadings of 0.45 or higher on the following items: Working on your own, Being sensitive to others’ feelings, Not giving up on a task that is too hard to finish, Always doing what you said you were going to do, and Being on time; and we have termed this Factor a measure of conscientiousness or dependability. Factor 3 is dominated by high loadings on items Listening to others, Talking to others, Working with others, Being sensitive to others’ feelings, and Asking questions and gathering information to solve problems – all measures of pro-social behavior (communication/teamwork, empathy). Finally Factor 4 contains high loadings on items Solving problems, Thinking creatively and coming up with new ideas, and Making presentations – indicators of higher-order thinking and problem solving. Once we identify these subscales from the factor analysis, we follow the convention of summating the scores on these items to create composite measures (Kerlinger, 1986; Kline, 1998; Likert, 1932; Oppenheim, 1992) for (a) overall soft skills, (b) pro-social behavior, (c) higher order thinking, and (d) conscientiousness. A summary of the factor analysis results appears in Appendix B along with the

Grade Level	School Curriculum Focus	NtN Instruction	
		After School Program	Summer Program
Fourth Grade	Basic Astronomy, The Human Body, Rocks and Minerals	Rocks and Minerals, The Human Body STEM- Make your own Slime, Sugar Water Density, anti-gravity water, rock candy, skewed balloons, clouds activity	Basic Astronomy, Birds, Insects, Reptiles and Amphibians
Fifth Grade	Mixtures and Solutions, Elements and Atoms, Principles of Anatomy	Principles of Anatomy, Mixtures and Solutions STEM- EGG Bungee Jump, Touchdown Activity, Bouncy Ball Activity, Slime Activity, Make your own fidget spinners, balloon powered cars	Fish, Pond Life, Basic Horticulture, Flowers
Sixth Grade	Forces and Motion, Electricity, and Basic Physics, Heat Transfer, The Human Brain	The Human Brain, Newton's Laws, The Scientific Method,Experimentation STEM- LED Projector, lava lamp, balloon hovercraft, foosball table, kinetic sculpture, electrolysis of water, lemon battery, Oobleck	Naturalistic Methods, Evolution and Ecologic niches,Photosynthesis
Seventh Grade	Acids and Bases, Mitosis and Meiosis, Osmosis and Diffusion, Basic Cell Biology	Cell Biology, Lab Reports, Principles of Micro-Biology,Eukaryotic and Prokaryotic Cells STEM – Re-growing cabbage, make a kite, oobleck, static electricity, gliders	Microscopic Pond Life, Healthy Foods and Nutrition, Gardening Techniques / Naturescape Development
Eighth Grade	Processes of Science, Forces and Motion, Energy Transformations, Light, Heat, Solar Energy and Weather, Rocks and the Rock Cycle, Geological time	Atoms and Elements, Evolution, The Processes of Science, Substances, Molecules STEM- rubber band helicopter, LEGO flashlight, propeller powered car, mini robot, fidget spinners, cork launcher	Tree Identification and Uses, Solar Energy and Weather
Ninth Grade	Biology - Evolution, Genetics	Saturday Program- Evolution, Genetics STEM- holiday circuits, levee, rube Goldberg machine, paper bridge	Critical Thinking, Statistics, Pig Dissection, Mentoring
Tenth Grade	Chemistry - Element creation and stars, atoms, fission, fusion, big bang, radiation, chemical reactions, balancing chemical equations, the carbon cycle, global warming	Saturday program- Element creation and stars, atoms, fission, fusion, big bang, radiation, chemical reactions, balancing chemical equations, the carbon cycle, global warming STEM- bioenergy farm game, carbon footprint, duct tape wallets/accessories	SAT Prep, Resume Building, Financial aid resourcing, New Brunswick Farmers Market Volunteering, Mentoring
Eleventh Grade	Physics-Waves, sound, motion	Saturday Program –Waves, sound, motion, SAT prep, college tours, guest speakers STEM- drone discovery activity, bio-degradable fork, hydraulics with syringes	Summer internships/career shadowing, Mentoring
Twelfth Grade	AP Science Courses- Physics, Biology, Chemistry	STEM mini projects (TBD), Applying to colleges, college tours, financial aid resourcing	Summer mentoring with younger NtN students

Fig. 1. NtN augmentation and extension of the Public School Science Curriculum.

Cronbach alphas used to assess subscale reliability.

2. Data and methods

2.1. Sample Characteristics, study variables and data sources

Data used to test NtN impact are generated using a classical experiment, with students randomly assigned to the NtN group and a control group (RCT) in seven New Brunswick, NJ elementary schools.[1] Random assignment helps (a) alleviate concerns about selection bias that is very common in observational studies, (b) generate differences in outcomes between groups that are otherwise statistically similar, and (c) obviate the necessity of using schoolteachers/ counselors as judges of student fitness for NtN, thereby minimizing the outcry of the parents of unselected students around issues of "favoritism." In addition to using the random assignment procedure to select students into NtN and control groups, we also randomly drew students to populate a 'Waiting List' from which students could be selected to become a part of the program if students who were already selected needed to withdraw for some reason or dropped out of the program at a later date. For each of the 8 cohorts, sample selection was conducted through a lottery when students were in the end of 3rd grade or beginning of 4th grade after stratifying by classroom.[2] The lottery numbers were drawn by the School Principal and the 'winners' were assigned to the NtN group, and the remaining students comprised the control group. The first 20 numbers drawn identified NtN students and the last two numbers drawn within each classroom and gender group identified students who were placed into the waiting list.

While our random assignment of students to treatment and control groups adjusts for group differences in cognitive and non-cognitive achievement at baseline, it is not blind to the individuals who are providing the academic outcome data. Classroom teachers could, in principle, assign grades that are systematically higher (or lower) to NtN participants based on their favorable (or unfavorable) judgments of program value. We are confident that this type of teacher effect or classroom bias is not a factor in this experiment for two reasons: (1) Using its designation as a Title I and Supplemental Education Programs school district, New Brunswick Public School administrators have put into place a set of broad guidelines that seek to promote equitable learning opportunities for all students. Among these guidelines is a policy that discourages school principals and teachers from giving preferential treatment to students engaged in special programs, sports and other extracurricular activities. The policy extends to class release time, minimum hours of class instruction, and activities that could interfere with the delivery of core curriculum; and (2) School principals assign students in each grade, i.e., elementary through middle school, to classroom teachers through a process they term equal opportunity assignment. This method works as follows: (a) all students entering a grade are arrayed in alphabetical or reverse alphabetical order; (b) all teachers who teach that grade are arrayed in alphabetical or reverse alphabetical order; (c) students are assigned to teachers using a count-off rule. For example, in the school from which the first cohort was selected, there were 66 students in 3rd grade who were assigned to the 3 teachers who teach 3rd grade using the rule that the 1st, 4th, 7th, etc. student was assigned to teacher 1; the 2nd, 5th, 8th, etc. to teacher 2 and so on, resulting in class sizes of 22 students. This process was again used with the 4th grade students and teachers, and in subsequent grades through middle school. In practical terms, this meant that the approximately 6 NtN students who were selected from each 3rd grade classroom by lottery would in all likelihood find themselves assigned to different teachers in the 4th grade, diminishing the prospects of any ongoing lottery (classroom) effect. Of course, parents of NtN students

[1] Two cohorts were selected from one elementary school two years apart.

[2] As a sensitivity check to determine if these lottery strata (i.e., initial classroom assignment) have any bearing on our treatment effects, we re-estimated our regressions including dummy variables for the lottery strata. Incorporation of the lottery effects into our models made little difference with respect to treatment effect estimates (either coefficients or effect sizes). These results are available from the authors.

Table 1
Sample characteristics at baseline by program status and cohort.

Characteristic	Cohort 1		Cohort 2		Cohort 3		Cohort 4
	NtN	Control	NtN	Control	NtN	Control	NtN
	(*n* = 19)	**(*n* = 47)**	**(*n* = 21)**	**(*n* = 76)**	**(*n* = 17)**	**(*n* = 90)**	**(*n* = 18)**
Demographic (%)							
Male	42.0	55.3	47.6	46.1	41.2	55.6	66.7
Hispanic	78.9	82.9	76.2	76.3	94.1	84.4	100.0
English Language	33.3	21.3	42.9	36.5	17.6	23.3	5.6
Special Education	10.5	10.6	4.8	5.3	23.5	18.9	11.1
Academic Outcomes							
Mean (Std.dev.)							
Math grade	80.2 (11.4)	80.4 (10.6)	82.6 (9.8)	79.8 (8.2)	79.1 (7.7)	81.1 (7.3)	82.1 (10.2)
Language Arts grade	79.6 (9.4)	77.0 (9.7)	78.6 (9.8)	77.5 (8.0)	77.5 (9.2)	78.4 (8.2)	80.5 (11.3)
Science grade	81.7 (8.1)	79.2 (9.8)	83.5 (7.5)	82.6 (6.5)	80.9 (4.7)	82.6 (6.3)	78.4 (9.7)
Soft skills							
Mean (Std.dev.)							
Overall Soft skills	14.6* (22.3)	27.1 (21.2)	28.7 (24.0)	22.6 (22.8)	23.6 (23.5)	26.9 (22.5)	33.7* (19.4)
Pro-social behavior	5.4* (8.2)	10.0 (7.9)	9.9 (8.3)	8.3 (8.5)	8.5 (8.5)	9.9 (8.3)	12.3* (7.3)
Higher order thinking	3.2 (4.9)	5.4 (4.3)	6.2 (5.2)	4.3 (4.7)	5.0 (5.2)	5.4 (4.8)	6.8* (4.2)
Conscientiousness	4.8* (7.6)	9.7 (7.7)	10.1 (8.5)	8.4 (8.5)	8.3 (8.3)	9.5 (8.3)	12.4* (7.0)

Characteristic	Cohort 4	Cohort 5		Cohort 6		Cohort 7		Cohort 8	
	Control	NtN	Control	NtN	Control	NtN	Control	NtN	Control
	(*n* = 113)	**(*n* = 14)**	**(*n* = 70)**	**(*n* = 19)**	**(*n* = 19)**	**(*n* = 18)**	**(*n* = 33)**	**(*n* = 13)**	**(*n* = 43)**
Demographic (%)									
Male	52.2	57.1	61.4	47.4	52.6	33.5	36.4	38.5*	69.8
Hispanic	94.7	93.0	88.6	68.4	52.6	55.6	60.6	92.3	88.4
English Language	18.6	15.4	17.6	68.4	50.0	66.7	57.6	23.1	20.9
Special Education	19.5	7.1	28.6	31.6	21.1	33.3	30.3	7.7	18.6
Academic Outcomes									
Mean (Std.dev.)									
Math grade	83.5 (7.8)	80.9 (7.9)	79.3 (6.2)	77.2 (12.5)	76.5 (6.8)	72.8 (14.0)	75.5 (10.4)	82.2 (6.6)	81.4 (6.7)
Language Arts grade	80.4 (8.6)	84.2 (5.9)	81.0 (6.8)	78.4 (11.3)	78.6 (5.5)	76.3 (14.0)	76.7 (8.9)	82.8 (5.6)	82.7 (6.4)
Science grade	79.4 (8.0)	88.6 (5.3)	84.9 (5.3)	81.9 (10.8)	82.7 (5.1)	82.5 (7.2)	82.3 (6.2)	89.9 (3.1)	90.0 (5.0)
Soft skills									
Mean (Std.dev.)									
Overall Soft skills	22.1 (21.5)	37.7 (16.8)	32.7 (21.4)	35.9 (16.8)	37.5 (17.7)	30.3 (20.0)	31.8 (20.2)	34.0 (16.6)	34.1 (19.3)
Pro-social behavior	8.1 (8.0)	14.4 (6.5)	11.9 (7.8)	13.2 (6.0)	13.5 (6.4)	10.4 (7.1)	11.7 (7.5)	11.9 (6.0)	12.3 (7.1)
Higher order thinking	4.5 (4.5)	4.5 (4.5)	6.6 (4.7)	7.7 (3.9)	7.8 (4.3)	6.4 (4.5)	6.5 (4.4)	7.2 (4.0)	7.0 (4.2)
Conscientiousness	8.0 (8.0)	13.6 (6.3)	11.7 (7.8)	12.9 (6.4)	12.8 (6.3)	11.1 (7.4)	11.3 (7.4)	12.8 (6.3)	12.3 (7.2)

* indicates significant group differences at baseline.

could request that their child be placed with a specific teacher or with other NtN students, but this was an uncommon event in the district. The two factors combined to limit teacher knowledge of who was in NtN after the original assignment, and to discourage subsequent teachers who became aware of a student's NtN status from using this knowledge to affect their assessment of student performance.

NtN began with the first cohort of 19 students in the summer of 2010 when these students had completed 3rd grade. Subsequent cohorts were drawn in 2012 (2 cohorts), 2013 (1 cohort), 2014 (3 cohorts) and 2015 (1 cohort) as more funding became available. Because of the phased-in way in which the sample was accumulated, NtN participants had differing years of program exposure. For example, as of 2017, the first cohort had 8 full years of NtN participation, while the last cohort had only 2 full years of NtN exposure. The sample contains a total of 630 students, with 139 students in the NtN group and the remaining 491 students in the control group.

In Table 1, we provide descriptive data on the NtN and control group students for the 8 cohorts at baseline. For each cohort, we provide comparative information on demographic, academic, and soft skill measures for the NtN and control group students. All demographic and academic data (that includes subject grades) were obtained from the centralized data system maintained by the New Brunswick school district for purposes of producing student report cards and/or reporting student-level information to the state department of education. Data on soft skills come from the NtN-KSAI described above.

Table 1 shows baseline equivalence between the experimental and control groups on nearly all the variables listed, validating the random assignment. The two groups show no significant differences with respect to any of the academic outcomes we consider here – student grades in math, language arts, and science. However, NtN students in cohort 4 score significantly higher in many of the soft skills measures than their control counterparts, while those in cohort 1 score significantly lower. One cohort (cohort 8) also exhibits a statistically significant difference in its gender composition (38.5 vs. 70 percent males). We control for all baseline characteristics in our multivariate analyses so that the effect of any ‘unhappy randomization’ may be minimized. In fact, once all baseline characteristics are controlled for, our multivariate analyses find no statistically significant baseline differences between NtN and control groups in any of the outcomes we have considered here, once again lending credence to the integrity of random assignment.

Table 2 provides the reader with an overall summary of student demographics, academics, and soft skill measures employing data pooled from all 8 cohorts at baseline. With two exceptions, this Table also indicates that random assignment was successful. We see a statistically significant difference between the NtN and control groups in the percent of homes where English is spoken (35.8 in the NtN group vs. 26.2 percent in the control group). In addition, NtN students score slightly higher in the overall soft skills measure (20.2 vs. 16.3).

Table 2
Sample characteristics at baseline by program status - All Cohorts.

Characteristic	NtN (n = 139)	Control (n = 491)
Demographic%		
Male	46.8	53.9
Hispanic	81.3	83.5
English Language	35.8	26.2
Special Education	16.5	18.3
Academic [Mean (Std. Dev.)]		
Math grade	79.5 (10.7)	80.4 (8.1)
Language Arts grade	79.5 (9.5)	79.3 (8.1)
Science grade	83.1 (8.2)	82.8 (7.4)
Soft Skills [Mean (Std. Dev.)]		
Overall Soft skills*	20.2 (21.7)	16.3 (21.5)
Pro-social behavior	7.3 (7.9)	5.9 (7.9)
Higher order thinking	4.2 (4.7)	3.3 (4.5)
Conscientiousness*	7.4 (8.0)	5.8 (7.8)

1Maximum N shown - it may vary from variable to variable within each group.
* indicates significant group differences at baseline.

Table 3
Distribution of study variables across student-year observations.[1]

Characteristic	n	Mean	Standard Deviation	Minimum	Maximum
Demographic%					
Male	2930	51.3	–	0	1
Hispanic	2936	84.4	–	0	1
English Language	2873	27.0	–	0	1
Special Education	2933	16.8	–	0	1
Academic [Mean (Std. Dev.)]					
Math grade	2545	77.6	10.5	35	100
Language Arts grade	2539	77.9	9.0	43	100
Science grade	2502	81.4	8.9	37	100
Soft Skills [Mean (Std. Dev.)]					
Overall Soft skills*	2937	19.6	20.7	0	52
Pro-social behavior	2937	7.7	8.2	0	20
Higher order thinking	2937	4.3	4.7	0	12
Conscientiousness	2937	7.6	8.1	0	20

[1] Contribution from each cohort: Cohort 1 16%; Cohort 2 17.2%; Cohort 3 19.4%; Cohort 4 21.1%; Cohort 5 11.2%; Cohort 6 4.8%; Cohort 7 6.5%; Cohort 8 3.8%.

We do not limit our analyses to difference-in-difference estimates since this would not allow us to exploit the rich, multi-year structure of the NtN intervention. Instead we examine student trajectories across all 8 cohorts yielding a maximum of 2937 student-year observations.[3] Table 3 shows the distribution of all study variables in the student-year unit. We see that a little over half the student-year sample is male, over 80 percent is Hispanic, about 17 percent are classified as special education students, and about a quarter of the homes where English is spoken as the main language. Average student grade is a $C+$ in math and language arts, and a low B in science.

2.2. Analytic approach

Longitudinal data permit the systematic assessment of stability and change over time and can provide valuable insights into the course and causes of many social behaviors. Despite many advantages, such data also bring with them an array of new challenges, especially with respect to data analysis and meeting the critical assumption of error independence (Cheslock & Rios-Aguilar, 2011; Clark, Crawford, Steele, & Vignoles, 2015). Traditionally, longitudinal data have been analyzed using a member of the General Linear Model (GLM) family, which includes the repeated measures t-test, analysis of variance and covariance, multivariate analysis of variance and covariance and multiple regression models (see Dwyer, 1983; Menard, 2002, and Crowder and Hand, 1996 for a good review of these methods). A common thread that ties all these methods together is that they tend to be considered *fixed effects* models, where systematic relationships are estimated by pooling all cross-sectional units over all time periods and the only source of random variation comes from the traditional regression error term. A particular disadvantage with these techniques is their inability to model in a parsimonious fashion, an underlying trajectory of change in the dependent variable that may unfold over time and that may differ across units. There has been an explosion of statistical methods and software in the last two decades that facilitate analysis of over time data that explicitly account for the error dependence and permit estimation

[3] Because we examine individual differences over time, our multivariate analyses exclude students from both the experimental and control groups who appear in the data set only once (n = 9).

of random variation in the levels and rate of change in the dependent variable. One such method for investigating change is termed multilevel model, often known as hierarchical linear model (HLM). These models permit straightforward examination of both *intra*-unit (within unit) change over time and *inter*-unit (between units) variability in intra-unit change. The HLM is designed to explicitly recognize nested data structures as in the case of individuals nested within organizations, children nested within classes and classes nested within schools, or neighborhoods nested within cities. Bryk and Raudenbush (1987) extended the use of HLM to estimate trajectory models and demonstrated that nesting could take the form of repeated measures nested within units. That is, when units are followed over time, the measurement occasions (micro level) for any particular unit (macro level) form a group, the same way as students are grouped within a class. HLM is referred to as latent trajectory or growth curve models in psychology Curran & Hussong, 2003) multilevel linear models in sociology (Goldstein, 2010), mixed-effects and random-effects models in biometrics (Singer, 1998), random coefficients regression models in the econometrics literature (Cameron & Trivedi, 2005; Cheslock & Rios-Aguilar, 2011; Clark et al., 2015; Longford, 1993; Wooldridge, 2002) and as covariance components models in the statistical literature (Snijders & Bosker, 1999; Longford, 1993). Hence, models that produce identical results tend to vary substantially in their nomenclature across disciplines (Cheslock & Rios-Aguilar, 2011).

In this paper, we employ the HLM tradition of writing the model equations that is common in education research, followed by the corresponding reduced form econometric equation in order to appeal to an interdisciplinary audience. In econometric terms, the HLM is equivalent to the random effects panel regression model that is generalized to include a randomly varying slope coefficient(s). In other words, the econometric random effects panel regression model which allows for a random intercept is a special case of the more general random coefficient model we estimate here. Randomization of students into experimental and control groups enables us to meet the basic econometric assumption of orthogonality between unobserved omitted variables and the treatment assignment, and to use our analyses to draw causal conclusions or bolster internal validity (Cheslock & Rios-Aguilar, 2011). Our decision to fit random effects models is based on two factors: (1) Given random assignment, we can assume that the individual fixed effects are uncorrelated with the treatment and other independent variables. Satisfying this exogeneity assumption makes random effects estimators more efficient; and (2) The random coefficients model also allows us to estimate the variability around the average slope, i.e., the variability in individual student's development overtime. We believe this aspect of the model is important in light of our reliance on a Dewian philosophy that underpins the NtN program and stresses the importance of recognizing individual growth.[4]

Two issues regarding the data structure that may influence our impact estimates merit mention here. First, since data are longitudinal there is the strong possibility of student attrition overtime. While attrition itself does not threaten the internal validity of our treatment estimates in the context of a randomized experiment, *differential* attrition between the experimental and control groups could pose a problem. In order to determine whether or not differential attrition is a problem in this experiment, we estimated a logistic regression of student attrition as a function of treatment status, baseline characteristics, and the interaction of these characteristics with the treatment status. We show the results from this analysis in Appendix C. The table provides no evidence of differential attrition – that is, any student attrition is independent of treatment status and other baseline characteristics. A second issue concerns the varying number of observations each student contributes to the dataset since students from the earlier cohorts have longer program exposure, and therefore the potential for unduly influencing impact estimates. Our estimation procedure allows each student his/her own slope over time, assigning proper weights to observations that take into account within-student correlation and the corresponding amount of new information brought to bear on the estimation process, thereby only *duly* influencing the estimates (Aitkin & Longford, 1986; Clark et al., 2015; Goldstein, 1997).

To get intention-to-treat estimates of NtN program impact on STEM and soft skills, we estimate our multilevel models by the method of maximum likelihood using Stata's xtmixed command (Version 15). Here, we estimate two-level models, where the first level investigates *within* student changes over time in their academic and soft skills outcomes (i.e., trajectories), and the second level explores if these individual trajectories are different for NtN participants and non-participants.

We estimate five different specifications, starting with a simple *unconditional means* only model (Model 1), followed by an *unconditional growth* model (Model 2) – these two models provide a useful baseline for comparison with our subsequent models (Models 3–5) that incorporate experimental group status, demographic factors, and cohort fixed effects. These unconditional models decompose the outcome variability into (a) across students irrespective of time and (b) across both students *and* time, and help establish whether there is predictable variability in the outcome that warrants an investigation and if so, whether this variability exists within or between students (Singer & Willett, 2003). The unconditional models are systematically augmented with predictors, with Model 3 introducing experimental status; Model 4 examining experimental (NtN) impact while controlling for student gender, race, special education status, and language spoken at home; and finally Model 5 that looks at NtN impact while also controlling for cohort-specific, time invariant differences.

Model 1 is specified as follows, with a Level 1 equation that models the observed outcome as a function of the individual-specific true mean and its deviation at time *t,* while Level 2 examines how this individual-specific mean varies from the grand mean:

$$\text{Level 1:} \quad Y_{it} = \pi_{0i} + \varepsilon_{it} \tag{4.1}$$

$$\text{Level 2:} \quad \pi_{0i} = \gamma_{00} + \zeta_{0i} \tag{4.2}$$

where

Y_{it} represents a particular academic or soft skills outcome (viz., school grade in math, language arts, science, overall soft skills, prosocial behavior, higher order thinking, and conscientiousness) for student *i* at time *t,*

π_{0i} is the individual-specific mean outcome,

ε_{it} is the deviation of the observed outcome from the individual-specific mean,

Υ_{00} is the grand mean, and

ζ_{0i} is the deviation of individual-specific mean from the grand mean.

We assume that the Level 1 and Level 2 residuals (ε_{it} and ζ_{0i}) are normally distributed, both with mean 0, and variance σ_ε^2 and σ_0^2 respectively, so that σ_ε^2 provides an estimate of the variability in the outcome of each individual around his/her own mean, and σ_0^2 summarizes the variability of individual-specific means around the grand mean.

Since the Level 2 equation cannot be estimated directly because of the structural parameter π_{0i}, we substitute [4.2] into [4.1] to obtain the reduced-form model for the observed responses Y_{it}, with one fixed component (Υ_{00}) and a composite residual (random component) as follows:

$$Y_{it} = \gamma_{00} + (\zeta_{0i} + \varepsilon_{it}) \tag{4.3}$$

Crowder and Hand (1990) refer to the fixed component as the

[4] We also note that a fixed effects panel regression model (which relies on individual variation overtime or ‘within’ variation) will fail to produce a coefficient for the treatment status inasmuch as it is a fixed characteristic of the individual student and remains constant overtime. However, one could estimate an average slope change using an interaction of the treatment status with time.

Table 4
Multi-level regression model for math grade.

Fixed Components		Parameter	Model 1 Coefficient (Robust Std. Error)	Model 2 Coefficient (Robust Std. Error)	Model 3 Coefficient (Robust Std. Error)	Model 4 Coefficient (Robust Std. Error)	Model 5 Coefficient (Robust Std. Error)
Initial Status (π_{0i})	Intercept	Υ_{00}	77.56***	80.33***	80.22***	80.56***	80.58***
			(0.34)	(0.37)	(0.42)	(1.33)	(1.65)
	NtN	Υ_{01}			0.47	0.56	0.41
					(0.87)	(0.83)	(0.82)
Rate of change (π_{1i})	Intercept	Υ_{10}		−1.46***	−1.60***	−1.13***	−1.14***
				(0.13)	(0.15)	(0.14)	(0.15)
	NtN	Υ_{11}			**0.60****	**0.55***	**0.57****
					(0.29)	**(0.29)**	**(0.28)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ_ε^2	56.38***	43.41***	43.31***	42.73***	42.70***
			(2.46)	(2.15)	(2.15)	(2.19)	(2.19)
Level 2	Initial status	σ_0^2	55.22***	49.42***	49.47***	43.33***	42.37***
			(3.82)	(5.17)	(5.20)	(4.87)	(4.88)
	Rate of change	σ_1^2		2.94***	2.90***	2.28***	2.28***
				(0.54)	(0.52)	(0.47)	(0.47)
	Covariance	σ_{01}		−0.85	−0.88	−1.00	−0.92
				(1.36)	(1.33)	(1.24)	(1.28)
Deviance			17,553.25	17,267.54	17,260.05	17,124.85	17,115.59
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2419	2419	2419	2419	2419

*** p-value < 0.01;.
** p-value < 0.05;.
* p-value < 0.10.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

"immutable constant of the universe," to ζ_{0i} as the "lasting characteristic of the individual" and to ε_{it} as the "fleeting aberration of the moment."

Model 2 estimates an unconditional growth model that introduces the predictor 'Time' at Level 1, allowing each student to have a distinct growth rate or trajectory π_{1i}, and enables us to examine whether inter-individual differences emanate from differences in the mean or the growth rate. Level 1, Level 2 and the reduced-form equations are specified as follows:

$$\text{Level 1:}\quad Y_{it} = \pi_{0i} + \pi_{1i}\,\text{Time}_{it} + \varepsilon_{it} \tag{5.1}$$

$$\text{Level 2:}\quad \pi_{0i} = \gamma_{00+}\zeta_{0i} \tag{5.2a}$$

$$\pi_{1i} = \gamma_{10+}\zeta_{1i} \tag{5.2b}$$

$$\text{Reduced-form: } Y_{it} = (\gamma_{00} + \gamma_{10}\text{Time}_{it}) + (\varepsilon_{it} + \zeta_{0i} + \zeta_{1i}\text{Time}_{it}) \tag{5.3}$$

We now have an additional structural parameter π_{1i} and a corresponding Level 2 equation [5.2b] that estimates inter-individual differences in the rates of change or growth trajectories. The fixed components Υ_{00} and Υ_{10} now estimate the mean intercept and mean growth rate, respectively; ζ_{0i} and ζ_{1i} are the deviations of each student from the group mean intercept and group mean growth rate; and the Level 1 residuals ε_{it} now tell us the individual deviation from his/her true growth trajectory. We continue to assume that both the Level 1 and Level 2 residuals have a normal distribution, with ζ_{0i} and ζ_{1i} now bivariate normal with mean 0 and variance $\sigma_0{}^2$ and $\sigma_1{}^2$. In addition, the covariance (σ_{01}) between ζ_{0i} and ζ_{1i} is also estimated in this model.

In Model 2, we have made the assumption that the time and individual-specific values of the outcome (Y_{ij}) are completely governed by the underlying trajectory process and any deviations of these values from the trajectory are treated as error. We now extend these models to capture situations in which we posit that the growth rates in outcomes are related only partly to the trajectory process but may also be influenced by their participation in the NtN program. We study the NtN impact in Model 3.

Level 1, Level 2 and the composite specifications of Model 3 are as follows:

$$\text{Level 1:}\quad Y_{it} = \pi_{0i} + \pi_{1i}\,\text{Time}_{it} + \varepsilon_{it} \tag{6.1}$$

$$\text{Level 2:}\quad \pi_{0i} = \gamma_{00+}\gamma_{01}\text{NtN} + \zeta_{0i} \tag{6.2a}$$

$$\pi_{1i} = \gamma_{10+}\gamma_{11}\text{NtN} + \zeta_{1i} \tag{6.2b}$$

$$\text{Reduced-form: } Y_{it} = (\gamma_{00} + \gamma_{01}\text{NtN} + \gamma_{10}\text{Time}_{it} + \gamma_{11}\text{NtN*Time}_{it}) + (\varepsilon_{it} + \zeta_{0i} + \zeta_{1i}\text{Time}_{it}) \tag{6.3}$$

Model 3 includes NtN participation as a predictor of both the initial or baseline outcome levels as well as the growth (change) in the outcomes. The Model contains four fixed components, Υ_{00}, the level of initial outcome of the average control group student; Υ_{01}, the difference in the initial outcome level between NtN and control students; Υ_{10}, the growth rate of the average control student; and finally Υ_{11}, the difference in the growth rate between the NtN and control students, which is the coefficient of interest that provides NtN program impact. The random effects parameters are specified as before.

Equations for Models 4 and 5 closely follow the specification used for Model 3, except in Level 2, we add demographic controls in Model 4 and cohort-specific controls in Model 5. In light of some descriptive evidence at baseline about the possibility of 'unhappy randomization' in some instances, and to make NtN impact estimates more precise, we consider Model 5, which controls for both student demographic characteristics and cohort-specific factors that remain time invariant, to be

Table 5
Multi-level regression model for language arts grade.

Fixed Components		**Parameter**	**Model 1 Coefficient (Robust Std. Error)**	**Model 2 Coefficient (Robust Std. Error)**	**Model 3 Coefficient (Robust Std. Error)**	**Model 4 Coefficient (Robust Std. Error)**	**Model 5 Coefficient (Robust Std. Error)**
Initial Status (π_{0i})	Intercept	Υ_{00}	77.89***	79.61***	79.33***	80.27***	80.42***
			(0.30)	(0.33)	(0.38)	(1.07)	(1.42)
	NtN	Υ_{01}			1.15	0.83	0.48
					(0.76)	(0.75)	(0.75)
Rate of change (π_{1i})	Intercept	Υ_{10}		−0.87***	−0.87***	−0.54***	−0.48***
				(0.11)	(0.12)	(0.12)	(0.13)
	NtN	Υ_{11}			**0.02**	**0.03**	**0.03**
					(0.23)	**(0.22)**	**(0.22)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ_ε^2	41.49***	35.04***	35.06***	35.05***	35.02***
			(1.93)	(1.92)	(1.92)	(1.90)	(1.90)
Level 2	Initial status	σ_0^2	41.25***	39.90***	39.65***	37.36***	36.98***
			(2.95)	(3.94)	(3.91)	(3.92)	(3.92)
	Rate of change	σ_1^2		1.71**	1.67**	1.17**	1.4**
				(0.44)	(0.44)	(0.35)	(0.35)
	Covariance	σ_{01}		−1.12	−1.06	−1.59	−1.76
				(1.10)	(1.08)	(1.02)	(1.04)
Deviance			16,754.30	16,605.06	16,601.71	16,490.74	16,477.29
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2410	2410	2410	2410	2410

*p-value < 0.10.
*** p-value < 0.01;.
** p-value < 0.05;.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

our final model. To assess model fit and improvement in model fit across models, we use the likelihood ratio test and the deviance statistic, respectively.

3. Results

Table 4 presents results of fitting multi-level trajectory models for math grades. Model 1 (Equations [4.1–4.3]) shows that the only fixed component parameter in the model (Υ_{00}), the average math score across all students over all time periods, is 77.56, and is significantly different from zero. The random components σ_ε^2 and σ_1^2 provide an estimate of the variability in math grades within and across students, and indicate that there is a significant amount of unexplained variability paving the way for inclusion of predictors. These variance estimates can also be used calculate an intra-class correlation coefficient, which provides us with an indication of how much variability in math grades is due to differences across students (Singer & Willett, 2003). Model 1 indicates that nearly 50 percent[5] of the variability in math grades is attributable to differences across students.

Model 2 presents the results of the unconditional growth model (Equations [5.1–5.3]) where the two fixed components Υ_{00} and Υ_{10} show that the estimated average starting point in math grade was 80.33, which was declining over time at a rate of 1.46 per year. The estimated Level 1 residual variance of 43.41 (σ_ε^2) shows the amount of average deviation of individual math grades from the student's own linear change trajectory, and when compared to Model 1, indicates that about 23 percent of the within-person variability in math grades (= (56.38–43.41)/56.38) is systematically related to *Time*, with a significant portion of the variability still left unexplained. The Level 2 residuals' variance of 49.42 and 2.94 summarize between-individual differences in the starting point and the rates of change, and their statistical significance suggests that there is still a substantial amount of unexplained variability in both the starting point and the growth rate and that there is benefit in adding substantive predictors to the model. The Model also estimates that the covariance between the Level 2 residuals (σ_{01}) is −0.85, indicating that student math scores that are higher in the beginning decline less rapidly over time, although this relationship is not significant.

In Model 3, we add NtN treatment as a substantive predictor in both the initial level of math grades and their growth over time, to assess whether the program served to shift the average trajectory upwards, or if it at least slowed down the decline in math grades. The estimated fixed components for levels of math grade reported in the top panel of the Table show that the average initial math grade for the control group students was 80.22, while for the NtN students it was about a half a point higher, although this difference was not significant. The estimated growth parameters on the other hand, show that there is a significant difference in the math grade trajectory between the NtN and control students. While both NtN and control students lost ground in math grades annually, the rate of decline for the NtN student was significantly slower (1.0 vs. 1.6 points), by about 38 percent.

The estimate of the within-variance component (σ_ε^2) in Model 3 remains similar to that of Model 2 indicating that the model could benefit from the inclusion of other predictors. Estimates of the Level 2 between-variance components also remain significant and about the same as the previous model suggesting the inclusion of additional predictors for both the level and trajectory in math grades.

Results from Model 4 that includes the students' personal fixed

[5] Since the total variability in math grades is the sum of two variance components - within and between variability, we can calculate the intra-class correlation, or that portion of the variability that is due to differences across individuals as: $\sigma_0^2/(\sigma_0^2 + \sigma_\varepsilon^2)$.

Table 6
Multi-level regression model for science grade.

Fixed Components		**Parameter**	**Model 1 Coefficient (Robust Std. Error)**	**Model 2 Coefficient (Robust Std. Error)**	**Model 3 Coefficient (Robust Std. Error)**	**Model 4 Coefficient (Robust Std. Error)**	**Model 5 Coefficient (Robust Std. Error)**
Initial Status (π_{0i})	Intercept	γ_{00}	81.50*** (0.27)	83.45*** (0.28)	83.22*** (0.32)	83.86*** (0.96)	81.81*** (1.27)
	NtN	γ_{01}			0.97 (0.63)	0.92 (0.59)	0.65 (0.58)
Rate of change (π_{1i})	Intercept	γ_{10}		−1.03*** (0.11)	−1.12*** (0.13)	−0.66*** (0.12)	−0.58*** (0.12)
	NtN	γ_{11}			**0.44*** (0.23)**	**0.45** (0.21)**	**0.43** (0.21)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ^2_ε	50.36*** (2.13)	45.20*** (2.27)	45.25*** (2.27)	43.78*** (1.93)	43.61*** (1.93)
Level 2	Initial status	σ^2_0	30.01*** (2.69)	15.15*** (3.06)	14.73*** (3.00)	12.70*** (2.23)	11.23*** (2.02)
	Rate of change	σ^2_1		1.19** (0.61)	1.10** (0.59)	0.43** (0.15)	0.44** (0.15)
	Covariance	σ_{01}		2.50** (1.15)	2.50** (1.10)	2.33*** (0.34)	2.23*** (0.30)
Deviance			16,762.90	16,586.67	16,576.71	16,391.46	16,354.00
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2376	2376	2376	2376	2376

*** p-value < 0.01;.
** p-value < 0.05;.
* p-value < 0.10.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

characteristics (gender, race, language spoken at home, and special education status) indicate a similar pattern to the previous model. The benefit NtN students accrue from program participation is reinforced in Model 4, with the average control student's math grade declining at an annual rate of 1.13 points and that of the average NtN student's declining at a rate of 0.58 (= −1.13 + 0.55), or about by half as much. Model 5 adds cohort fixed effects, that is, characteristics specific to each NtN cohort that remain time-invariant (e.g., teacher quality, school quality, or other unobserved/unmeasured factors). Results however, are virtually identical to the previous model, with a significant treatment effect in the form of an arrested decline in math grade each year by about 50 percent. Estimates of both the within- and between-variance components in Models 4 and 5 continue to indicate the presence of significant unexplained variance at both Level 1 and Level 2, however, the size of the variance estimate is somewhat smaller at Level 1 and considerably smaller at Level 2 with respect to both initial level and rate of change.

All five models show good fit as indicated by the significant likelihood ratio test. Progressive reductions in the deviance statistic in each model relative to the previous model point to the usefulness of the predictors added. The final model (Model 5) also shows moderate to considerable reductions in the within-individual and between-individual error variances relative to the baseline unconditional models (Models 1 and 2), confirming the conclusions indicated by the deviance statistic with respect to improvements in model fit.

In Table 5, we provide the results from our multi-level regression analyses of changes in language arts grade. None of the models that include the treatment (Models 3, 4, and 5) shows that NtN altered the course of the overall negative trajectory of language arts grade (an average annual decline of about half a grade point according to Model 5).

When we fit our series of multi-level models for science grades Table 6), we once again find an effect of NtN participation. As in the case of mathematics grades, NtN appears to significantly slow down the negative trajectory of grade performance. Model 5 shows that the growth parameters estimated from this model demonstrate a significant decline in science grade at a rate of 0.58 per year for the control students, with NtN student grade declining at a significantly lower rate of 0.15 (= −0.58 + 0.43). The bottom half of the Table indicates that even though considerable intra and inter-individual variability remains unexplained, the substantial and progressive reduction in these variance estimates across models is noteworthy, as is the improvement in model fit. In addition, we note that the covariance estimate shows significance in all models, implying that higher initial grades are associated with a lower rate of decline in grade.

In Tables 7 through 10, we present results on NtN impact on student soft skills. We continue to interpret NtN impact in the same fashion as we have in the case of hard skills presented in Tables 4, 5, and 6, primarily by focusing on our final model (Model 5), and on (1) any significant baseline differences between NtN and control group students, (2) annual rate of change in the outcome, (3) the adjustment that NtN treatment has on that trajectory, and (4) indications of unexplained variation.

Table 7 shows results for overall soft skills. These results indicate that while there was a non-significant decline in these skills annually for the control group on average, the NtN students experienced a *significant* and net *positive* growth in their trajectories where their soft skills grew at a rate of 1.95 (= −0.20 + 2.15) on average. While individual variability over time remains more or less invariant to addition of predictors across models, between individual differences initially and over time are reduced rather dramatically from the initial to the final models as more predictors are added.

In Table 8, we see that the results on pro-social behavior (communication, teamwork, empathy) are very similar to overall soft skills

Table 7
Multi-level regression model for overall soft skills.

Fixed Componentss		Parameter	Model 1 Coefficient (Robust Std. Error)	Model 2 Coefficient (Robust Std. Error)	Model 3 Coefficient (Robust Std. Error)	Model 4 Coefficient (Robust Std. Error)	Model 5 Coefficient (Robust Std. Error)
Initial Status (π_{0i})	Intercept	γ_{00}	22.44***	24.08***	24.06***	23.33***	7.67***
			(0.51)	(0.74)	(0.84)	(2.13)	(1.90)
	NtN	γ_{01}			0.02	−0.06	−1.10
					(1.69)	(1.69)	(1.39)
Rate of change (π_{1i})	Intercept	γ_{10}		−0.88***	−1.31***	−1.10***	−0.20
				(0.27)	(0.31)	(0.31)	(0.33)
	NtN	γ_{11}			**1.90*****	**1.80*****	**2.15*****
					(0.58)	**(0.57)**	**(0.58)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ_ε^2	364.36***	348.95***	348.06***	348.51***	345.94***
			(9.13)	(10.56)	(10.50)	(10.43)	(10.38)
Level 2	Initial status	σ_0^2	65.68***	102.94***	97.16***	90.14***	5.23
			(9.32)	(17.79)	(17.65)	(17.60)	(16.12)
	Rate of change	σ_1^2		5.81**	4.24**	3.87**	3.81
				(2.00)	(1.98)	(1.87)	(2.07)
	Covariance	σ_{01}		−15.75***	−12.64**	−12.38**	0.26
				(5.37)	(5.40)	(5.36)	(5.42)
Deviance			22,020.83	22,004.06	21,983.59	21,944.48	21,748.50
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2483	2483	2483	2483	2483

*p-value < 0.10.
*** p-value < 0.01;.
** p-value < 0.05;.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

Table 8
Multi-level regression model for pro-social behavior.

Fixed Components		Parameter	Model 1 Coefficient (Robust Std. Error)	Model 2 Coefficient (Robust Std. Error)	Model 3 Coefficient (Robust Std. Error)	Model 4 Coefficient (Robust Std. Error)	Model 5 Coefficient (Robust Std. Error)
Initial Status (π_{0i})	Intercept	γ_{00}	8.82***	9.44***	9.46***	9.13***	3.01***
			(0.20)	(0.29)	(0.33)	(0.84)	(0.76)
	NtN	γ_{01}			−0.16	−0.20	−0.57
					(0.66)	(0.66)	(0.54)
Rate of change (π_{1i})	Intercept	γ_{10}		−0.34***	−0.50***	−0.43***	−0.07
				(0.11)	(0.12)	(0.12)	(0.13)
	NtN	γ_{11}			**0.74*****	**0.70*****	**0.84*****
					(0.23)	**(0.23)**	**(0.23)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ_ε^2	57.44***	54.87***	54.88***	54.87***	54.55***
			(1.42)	(1.64)	(1.63)	(1.62)	(1.63)
Level 2	Initial status	σ_0^2	9.48***	15.71***	14.83***	13.53***	0.33
			(1.41)	(2.78)	(2.76)	(2.73)	(2.57)
	Rate of change	σ_1^2		0.96**	0.71**	0.64**	0.60*
				(0.31)	(0.31)	(0.30)	(0.33)
	Covariance	σ_{01}		−2.59***	−2.09**	−2.00**	0.08
				(0.85)	(0.85)	(0.84)	(0.87)
Deviance			17,412.01	17,395.33	17,377.47	17,337.88	17,150.25
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2483	2483	2483	2483	2483

*** p-value < 0.01;.
** p-value < 0.05;.
* p-value < 0.10.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

Table 9
Multi-level regression model for higher order thinking.

Fixed Components		**Parameter**	**Model 1 Coefficient (Robust Std. Error)**	**Model 2 Coefficient (Robust Std. Error)**	**Model 3 Coefficient (Robust Std. Error)**	**Model 4 Coefficient (Robust Std. Error)**	**Model 5 Coefficient (Robust Std. Error)**
Initial Status (π_{0i})	Intercept	γ_{00}	4.95***	5.32***	5.26***	5.10***	1.66***
			(0.12)	(0.17)	(0.19)	(0.50)	(0.44)
	NtN	γ_{01}			0.24	0.20	−0.08
					(0.38)	(0.38)	(0.32)
Rate of change (π_{1i})	Intercept	γ_{10}		−0.20***	−0.30***	−0.25***	−0.04
				(0.06)	(0.07)	(0.07)	(0.07)
	NtN	γ_{11}			**0.41*** **	**0.39*** **	**0.47*** **
					(0.13)	**(0.13)**	**(0.13)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ^2_ε	18.63***	18.13***	18.18***	18.10***	17.95***
			(0.49)	(0.57)	(0.57)	(0.56)	(0.55)
Level 2	Initial status	σ^2_0	3.57***	5.04***	4.65***	4.45***	0.31
			(0.49)	(0.91)	(0.91)	(0.91)	(0.81)
	Rate of change	σ^2_1		0.20**	0.10	0.10	0.10
				(0.10)	(0.10)	(0.09)	(0.10)
	Covariance	σ_{01}		−0.61***	−0.44**	−0.46**	0.12
				(0.27)	(0.26)	(0.27)	(0.26)
Deviance			14,652.97	14,639.24	14,613.07	14,579.04	14,387.31
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2483	2483	2483	2483	2483

*p-value < 0.10.
*** p-value < 0.01;.
** p-value < 0.05;.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

Table 10
Multi-level regression model for conscientiousness.

Fixed Components		**Parameter**	**Model 1 Coefficient (Robust Std. Error)**	**Model 2 Coefficient (Robust Std. Error)**	**Model 3 Coefficient (Robust Std. Error)**	**Model 4 Coefficient (Robust Std. Error)**	**Model 5 Coefficient (Robust Std. Error)**
Initial Status (π_{0i})	Intercept	γ_{00}	8.65***	9.32***	9.30***	9.00***	3.01***
			(0.20)	(0.29)	(0.30)	(0.82)	(0.74)
	NtN	γ_{01}			0.06	0.06	−0.36
					(0.67)	(0.67)	(0.55)
Rate of change (π_{1i})	Intercept	γ_{10}		−0.36***	−0.52***	−0.43***	−0.09
				(0.11)	(0.12)	(0.12)	(0.13)
	NtN	γ_{11}			**0.71*** **	**0.66*** **	**0.80*** **
					(0.23)	**(0.23)**	**(0.23)**
Individual controls[a]			No	No	No	Yes	Yes
Cohort fixed effects[b]			No	No	No	No	Yes
Random Components			**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**	**Estimate (Robust Std. Error)**
Level 1	Within person	σ^2_ε	55.37***	52.81***	52.78***	52.71***	52.29***
			(1.42)	(1.61)	(1.60)	(1.59)	(1.58)
Level 2	Initial status	σ^2_0	10.19***	16.62***	15.79***	14.52***	2.19
			(1.42)	(2.76)	(2.74)	(2.72)	(2.48)
	Rate of change	σ^2_1		0.96**	0.75**	0.69**	0.70**
				(0.31)	(0.31)	(0.30)	(0.33)
	Covariance	σ_{01}		−2.68***	−2.23**	−2.18**	0.40
				(0.85)	(0.85)	(0.84)	(0.84)
Deviance			17,347.94	17,328.46	17,309.18	17,265.44	17,078.38
P (LR Chisquared Test)			0.00	0.00	0.00	0.00	0.00
N			2483	2483	2483	2483	2483

*p-value < 0.10.
*** p-value < 0.01;.
** p-value < 0.05;.
[a] Models 4 and 5 control for race, gender, language spoken at home, and special education status.
[b] Model 5 controls for cohort specific characteristics that are time invariant.

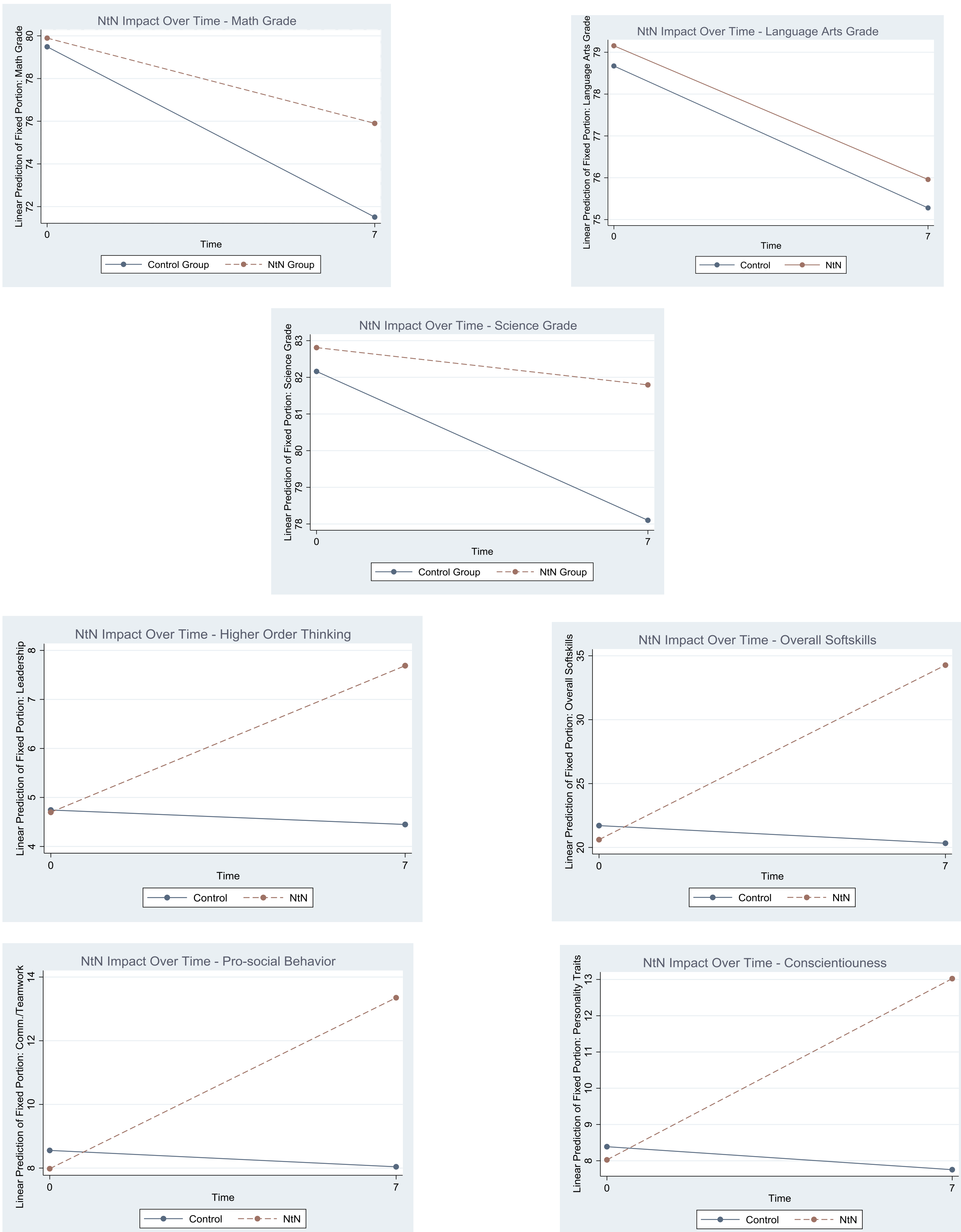

Fig. 2. NtN Impact on various outcomes over time.

Table 11
Sensitivity Analyses: Effect Size[a] Estimates for Various Model Specifications.

Outcome	Basic Findings	Including Quadratic Time	Years of Exposure		
			3 Years	4 Years	5 Years
Math grade	0.43**	0.46**	0.24*	0.32**	0.34**
Language Arts grade	0.03	0.03	0.01	0.01	0.01
Science grade	0.39**	0.40**	0.16	0.23*	0.22*
Overall Softskills	0.83***	0.85***	0.81***	0.68***	0.50**
Pro-social behavior	0.82***	0.84***	0.76***	0.65***	0.81***
Higher order thinking	0.80***	0.82***	0.35**	0.51***	0.56***
Conscientiousness	0.79***	0.81***	0.30***	0.52***	0.59***

Statistical significance displayed here pertain to impact coefficient from the final models.

[a] Calculated by taking the treatment effect coefficient from the final model, multiplying it by the number of student years in the program and dividing by the standard deviation of outcome given in Table 3.

*** p-value < 0.01;.

** p-value < 0.05;.

* p-value < 0.10.

measure with respect to NtN impact as well as variance estimates. Here, NtN participation significantly and positively affects students' ability in communication, teamwork, and empathy (pro-social behavior). NtN students experienced an average net positive growth of about 0.77 points on this scale annually, while the control group experienced a non-significant decline in these skills. We again see substantial reductions in estimated inter-individual variances and corresponding improvements in model fit across the five models.

Table 9 shows that NtN also impacts the growth trajectory of higher order thinking skills/problem solving among participants. The average NtN student experienced a growth rate of 0.43 points on this scale measure annually while the control student exhibited a decline at a rate of 0.04. Variance estimates and model fit mimic the patterns we found in the previous two outcomes.

Finally, Table 10 illustrates NtN's impact on enhancing students' conscientiousness, a key skill for labor force success. We see that control students experienced an insignificant negative growth, but participation in NtN significantly bolstered the trajectory of participants, increasing their rate of growth by 0.71 points on this scale measure. The bottom portion of the Table also makes evident, the reductions in inter-individual variability and corresponding improvement in model fit with additional predictors across models.

We should reiterate that all of our multivariate analyses (presented in Tables 4–10) demonstrate the baseline equivalence between the experimental and control groups, and by inference, the validity of random assignment. We provide a visual summary of NtN impact on both the hard and soft skills outcomes we have considered in Fig. 2. These graphs clearly show NtN's role in slowing the decline in cognitive skills and enhancing the growth in soft skills.

In Table 11, we present NtN impacts in the form of effect sizes which is a common practice in the education literature. These effect sizes[6](ES) represent the NtN program impacts provided by the final models shown in Tables 4–10. In this Table, we also test the robustness of our basic findings by (a) assuming that outcomes do not always follow a linear trend by including a quadratic term for time[7]; and (b) re-estimating the final model for the first 3, 4, and 5 years of program exposure. Table 11 shows that NtN increases student math grades relative to the control group by 0.43 standard deviations; this ES increases to about a half a standard deviation when we include quadratic time. ESs of this magnitude would be considered moderately high in the educational psychology literature (Cohen, 1988; Lipsey, 1990). When student exposure is taken into account, we see that effect sizes for math now move into the small to moderate range, linearly increasing with time. Science grades also exhibit comparable ESs to that of math grades, but also drop in size when the data are limited to 3, 4, or 5 years of student exposure.

Effect sizes are quite large with respect to soft skills overall and, perhaps more importantly, for the three more specific measures of soft skills (pro-social behavior, higher order thinking, and conscientiousness), typically in the order of 0.8 standard deviations in the basic model and with quadratic time. These ESs also decrease in magnitude when the data are restricted to years of program exposure, but still remain in the moderate to large range.

4. Discussion and conclusions

The low number of U.S. students pursuing careers in STEM disciplines has grave implications for the health of our nation's economy and democracy. Inadequate academic preparation of students in our elementary and middle schools, especially those living in disadvantaged school district, increase the likelihood that these children will have difficulties in high school and, if they graduate, in higher education and the labor market. Failure in science and math courses is almost certain to eliminate these students from high-paying STEM jobs and careers in the health professions. Complicating matters even more is the evidence that as STEM technical skills freeze or erode, so do soft skills and the social capital that is necessary for labor market success (Heckman, 2013; Heckman & Masterov, 2007).

The moderate impact of NtN on cognitive skills and the larger effects on soft, socio-emotional competencies is a departure from much of the research we have discussed, that has focused on extra-school, STEM enhancement programing directed at minority and disadvantaged youth. We believe there are several reasons that contribute to the magnitude of our findings. First, the NtN intervention starts at a young age, the beginning of 4th grade, when interventions have been consistently found to yield their highest returns to investment (Cunha & Heckman, 2008; Maltese & Tai, 2011; Ramey, Bryant, Campbell, Sparling, & Wasik, 1988; Schweinhart, Barnes, & Weikart, 1993; Tai et al., 2006). Many of the rigorous evaluations of extra-school program that we have cited earlier in this paper (e.g., Career Academies, the Quantum Opportunities Program, ECHSI, MESA, and BELL) begin in late adolescence or early high school, at a time when STEM identities are more difficult to mold (Orr, 1992; Lauer et al., 2006; Ord & Leather, 2011).

The second reason for the relative success of NtN is its implementation of a Dewian active learning philosophy (Dewey, 1976) within the framework of outdoor education (Ord & Leather, 2011; Quay & Seaman, 2012) and the "wonders of nature" curriculum and teaching model (Jagannathan et al., 2018). Through a program that promotes the conjoint reinforcement of after-school, summer immersion, in-school curriculum extension, involved parents and dedicated NtN teachers who are proficient in math and science, NtN actively confronts the issues of "low dosage" and underpowered treatment that often affect STEM enhancement programs targeting disadvantaged students

[6] Calculated by taking the treatment effect coefficient from the final model, multiplying it by the number of years of student time in the program and dividing by the standard deviation of outcome given in Table 3 (Cohen, 1988; Lipsey, 1990).

[7] The coefficient for the quadratic term for time is not statistically significant in any of the models for cognitive outcomes. It is significant in some of the soft skills models, but as can be seen from Table 11, it does not much change the treatment effect.

(Levine & Zimmerman, 2010; McCombs et al., 2011).

The third determinant of NtN impact, we believe, resides in the organizational partnership of Rutgers University, J & J and NBPS alluded to earlier. Active involvement of the University and J & J, through availability of scientists, laboratory facilities, careers exposure and (in the case of Rutgers) college students actively pursuing STEM and health care education and emphasizing the importance of technical and soft skill development has been critical in the process of creating STEM identities in NtN students. A similar partnership between Northwestern University, the Boys and Girls Club, and Chicago Public Schools also reports positive technical and soft skill improvement after a RCT evaluation.

Of course, NtN effects are subject to the vagaries of time and future circumstance. Like effects in the 21st CCLS program discussed earlier, or the Highscope Perry Preschool or Abecedarian Project (Heckman, 2000), NtN impacts, too, may diminish or even extinguish when these students begin their college or vocational careers. It is apparent for this research, that cognitive skills of NtN students (measured as math and science grades) decline as they advance through an increasingly more difficult curriculum. This finding with a sample of disadvantaged youth is not new with similar results reported by Alexander et al. (2007), Hanushek and Rivkin (2009), Wai et al. (2010), and Hill (2017). NtN's capacity to attenuate this decline is a finding that merits additional verification, ideally employing longitudinal data from other, natural science based extra-school programs.

Our findings with respect to the acquisition of soft skills are also encouraging. NtN students performed significantly better than controls on pro-social skills (communication, teamwork, empathy), higher order thinking and problem solving, and conscientiousness. If previous research is to serve as a guide, we would expect these skills to exhibit greater endurance, requiring less reinforcement learning through adulthood (Boscia, 2013: Kemple & Willner, 2008; Heckman, 2000; James-Burdumy et al., 2005; Levine & Zimmerman, 2010).

NtN makes a conscious attempt to bring to life the "University Elementary School" envisaged by John Dewey (Dewey, 1976; p.92–95). Of course NtN's attempt to operationalize the University Elementary School is subject to challenges that were all but impossible to see a century ago when Dewey first proposed the idea. Skyrocketing numbers of single parent homes, escalating levels of drug use and gang violence, etc., have made it more difficult to transport young students into a natural science world of exploration, excitement, and wonder. Notwithstanding these obstacles this journey remains as important as it ever has been.

Declaration of interest

None

Acknowledgments

Authors would like to acknowledge Bonnie Petrauskas and Joanne Fillweber from Johnson & Johnson Worldwide, Chancellor Dick Edwards, Mark Robson, and Dorothea Berkhout from Rutgers University, and Superintendent Aubrey Johnson and Dr. John Anzul of the New Brunswick Public Schools, for their steadfast support of the Nurture thru Nature program. We would like to thank German Rodriguez, Princeton University for his helpful comments on a previous version of the paper. We would also like to thank the Journal Editor and two anonymous referees for their insightful comments and suggestions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econedurev.2019.04.005.

Appendix A

Nurture thru Nature Knowledge, Skills and Abilities Inventory (NtN-KSAI)

Q4. Check the box that best describes how good you are at each skill. *(Put an X in the box that describes your skill level.)*

How good are you in:	Very Good	Good	Bad	Very Bad
a. Solving problems				
b. Listening to others				
c. Talking to others				
d. Working on your own				
e. Working with others				
f. Asking questions and gathering information to solve problems				
g. Reading and understanding written text/instructions				
h. Writing reports				
i. Making presentations				
j. Thinking creatively and coming up with new ideas				
k. Testing ideas about science				
l. Being sensitive to others' feelings				
m. Solving math problems				
n. Conducting science labs/experiments				
o. Using computers				
p. Making decisions				
q. Leading a group				
r. Being on time				
s. Always doing what you said you were going to do				
t. Not giving up on a task that is too hard to finish				

Appendix B

Composite Measure	Variable components and Cronbach's alpha
Overall Soft skills	How good the student feels he/she is at each of the following skills measured on a scale of 1–4 (1 = Very bad, 2 = Bad, 3 = Good, 4 = Very good): Solving problems, Listening to others, talking to others, working on their own, working with others, asking questions and gathering information to solve problems, making presentations, thinking creatively and coming up with new ideas, making decisions, being sensitive to others' feelings, leading a group, being on time, always doing what you said you were going to do, and not giving up on a task that is too hard to finish. Cronbach's alpha: 0.8
Pro-social behavior (Communication, Teamwork, Empathy)	How good the student feels he/she is at each of the following skills measured on a scale of 1–4 (1 = Very bad, 2 = Bad, 3 = Good, 4 = Very good): Listening to others, asking questions and gathering information to solve problems, talking to others, working with others, and being sensitive to others' feelings. Cronbach's alpha: 0.7
Higher order thinking	Solving problems, leading a group, thinking creatively and coming up with new ideas and making presentations. Cronbach's alpha: 0.6
Conscientiousness	Being sensitive to others' feelings, being on time, always doing what you said you were going to do, not giving up on a task that is too hard to finish, and working on their own. Cronbach's alpha: 0.7

Appendix C

Logistic Regression of Student Attrition

Characteristic	Coefficient (Robust Standard Error)
NtN Status	−0.16 (0.48)
Male	−0.38 (0.27)
Hispanic	0.56 (0.43)
English language	0.66 (0.37)
Special education	0.03 (0.35)
NtN Status * Male	−0.09 (0.76)
NtN Status * Hispanic	−0.02 (0.07)
NtN Status * English language	−2.15 (1.92)
NtN Status * Special education	−0.46 (1.19)
Number of observations	625
Log pseudo likelihood	−225.63
Pseudo R^2	0.03

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society Series A (General), 149*, 1–43.

Alexander, K. L., Entwisle, D. R., & Olsen, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*, 167–180.

Alvarado, S. A., & Muniz, P. (2018). Racial and ethnic heterogeneity in the effect of MESA on AP STEM coursework and college STEM major aspirations. *Research in Higher Education, 50*, 933–957.

American Institutes for Research (AIR). (2018). *What we know about the impact of the 21st CCLC program.* Chicago, IL: AIR. Available at www.air.org/page/afterschool-and-expanded-learning.

Attanasio, O. P. (2015). The determinants of human capital formation during the early years of life: Theory, measurement and policies. *Journal of the European Economic Association, 13*, 949–997.

Attanasio, O. P., Meghir, C., & Nix, E. (2017). Human capital development and parental investment in India. NBER Working Paper 21740. Retrieved at: http://www.nber.org/papers/w21740 Authors (2017) (2018).

Bagiati, A., Yoon, S. Y., Evangelou, D., & Ngambeki, I. (2010). Engineering curricula in early education: Describing the landscape of open resources. *Early Childhood Research & Practice, 12*(2), 2–15.

Berger, A., Turk-Bicakci, L., & Garett, M. (2013). *Early college, early successes: early college high school initiative impact study.* Washington, DC: American Institutes for Research.

Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences, 113*(47), 13354–13359.

Boscia, T. (2013). Serious about STEM. *Human Ecology, 41*(2), 9–11.

Briggs, D. C., & Peck, F. A. (2015). Using learning progressions to design vertical scales that support coherent inference about student growth. *Measurement, 13*, 75099.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of Hierarchical Linear Models to Assessing change. *Psychological Bulletin, 101*(1), 147–158.

Business Roundtable. (2017). *Work-in-progress - How CEO's are helping close america's skills gap.* Washington D.C: The Business Roundtable.

Camasso, M. J., & Jagannathan, R. (2017a). Experimental evidence for the effectiveness of nature-based learning on academic outcomes in poor urban schools. *Journal of Environmental Education, 49*, 30–42.

Camasso, M. J., & Jagannathan, R. (2017b). The Nurture thru Nature Program: Creating natural science identities in populations of at-risk children. *Cambridge Journal of Education, 48*, 263–277.

Cameron, A., & Trivedi, P. (2005). *Microeconometrics: methods and applications.* Cambridge University Press.

Carneiro, P., & Heckman, J. J. (2003). Human capital policy. In B. M. Friedman, J. J. Heckman, & A. Krueger (Eds.). *Inequality in America: what role for human capital policies?*. Cambridge, MA: MIT Press.

Center for the Economics of Human Development. (2015). *Conference on measuring and assessing skills report*University of Chicago.

Cheslock, J. J., & Rios-Aguilar, C. (2011). *Multilevel analysis in higher education research: A multidisciplinary approach. Higher Education: Handbook of Theory and Research, 25*.

Clark, P., Crawford, C., Steele, F., & Vignoles, A. (2015). Revisiting fixed- and random effects models: Some considerations for policy-relevant research. *Education Economics, 23*, 259–277.

Clarke, P. (2012). *Education for sustainability: becoming naturally smart.* London: Routledge.

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

Coleman, J. S. (1990). *Foundations of social theory.* Cambridge, MA: Harvard University Press.
Committee on Equal Opportunities in Science and Engineering. (2015). Report to Congress. http://www.nsf.gov/od/oia/activities/ceose/index.jsp.
Committee on Highly Successful Schools or Programs for K-12 STEM Education. (2011). *Successful STEM education: a workshop.* Washington D.C.: National Academic Press.
Crowder, M. J., & Hand, D. J. (1990). *Analysis of repeated measures.* London: Chapman and Hall257 pp.
Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica, 78*(3), 883–931.
Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources, 43*(4), 738–782.
Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*(4), 526–544.
Dewey, J. (1976). The school and society. In J. A. Boydston (Vol. Ed.), *John Dewey: the middle works: 1*, (pp. 1–96). Carbondale, IL: Southern Illinois University Press.
Dewey, J. (1990). Between two worlds. In J. A. Boydston (Vol. Ed.), *John Dewey: the later works: 17*, (pp. 451–465). Carbondale, IL: Southern Illinois University Press.
Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences.* Oxford, England: Oxford University Press.
Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness.* Washington, D.C.: Center for Research on the Education of Students Placed at Risk.
FHI-360. (2017). Celebrating the BTE Story: 25 Years, 25 Lessons: Retrieved from http://www.bridge2employment.org/wp-content/uploads/2017/03/25-Celebrating_the_BTE_Story-25Years-25Lessons.pdf.
Fryer, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics, 86*, 447–464.
Galloway, T., Lippman, L., Burke, H., Diener, O., & Gates, S. (2017). *Measuring soft skills and life skills in international youth development programs: a review and inventory of tools.* Washington, DC: USAID's Youth Power Implementation IDIQ.
Garcia, E. (2014). *The need to address noncognitive skills in the education policy agenda.* Washington, D.C: Economic Policy Institute, Briefing Paper #386. http://www.epi.org/publication/the-need-to-address-noncognitive-skills-in-the-education-policy-agenda/.
Goldin, C. (2016). Human capital. In C. Diebolt, & M. Haupert (Eds.). *Handbook of Cliometrics* (pp. 55–86). Berlin, Heidelberg: Springer.
Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement, 8*, 369–395.
Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). London: Wiley.
Gutman, L. M. Schoon. (2013). *The impact of non-cognitive skills on outcomes for young people. Institute of education.* London: University of London. https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Lit_Review_Non-CognitiveSkills.pdf.
Hanushek, E. A., & Rivkin, S. G. (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of Policy Analysis and Management, 28*, 366–393.
Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics, 54*, 3–56.
Heckman, J. J. (2013). *Giving kids a fair chance.* Cambridge, MA: MIT Press.
Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics, 19*(4), 451–464.
Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics, 29*, 446–493.
Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.
Hill, H. (2017). The Coleman Report, 50 years on: What do we know about the role of schools in academic inequality. *The ANNALS of the American Academy of Political and Social Science, 674*, 9–26.
Hirsch, B. J. (2005). *A place to call home: After-school programs for urban youth.* Teachers College Press.
Hollister, R. (2003). The growth in after-school programs and their impact. Washington, DC: Brookings Institution.
Ibarraran, P., Ripani, L., Tabooda, B., Villa, J. M., & Garcia, B. (2014). Life skills, employability and training for disadvantaged youth: Evidence from a randomized evaluation design. *IZA Journal of Labor and Development, 3*, 1–24.
Iowa Testing Programs (ITP). (2018). *Vertical scaling and the assessment of growth.* Iowa City, IA: University of Iowa. Retrieved at http://www.education.uiowa.edu/itp/.
Jagannathan, R., Camasso, M. J., & Delacalle, M. (2018). The effectiveness of the head-heart-hands model for natural and environmental science learning in urban schools. *Evaluation and Program Planning, 66*, 53–62.
James-Burdumy, S., Dynarski, M., Moor, M., Deke, J., & Mansfield, W. (2005). *When schools stay open late: the national evaluation of the 21st century community learning centers program.* U.S. Department of Education, Institute of Education Services. Available at http://www.ed.gov.ies.nces.
Jobs For the Future. (2017). *Bridging to a better future: findings from an evaluation of Bridge-to-College programs for english language learners.* Boston, MA: Jobs For the Future.
Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability and career success across the life span. *Personnel Psychology, 52*, 621–652.
Kemple, J., & Willner, C. (2008). *Career Academies: long term impacts on labor market outcomes, educational attainment and transitions to adulthood.* New York: MDRC.
Kerlinger, F. N. (1986). *Foundations of behavioral research.* New York, NY: Holt, Rinehart and Winston.
Klemmer, C. D., Waliczek, T. M., & Zajicek, J. M. (2005). Growing minds: The effect of a school gardening program on the science achievement of elementary students. *HortTechnology, 15*(3), 448–452.
Kline, P. (1998). *The new psychometrics: Science, psychology and measurement.* New York, NY: Routledge.
Krishnamurti, A., Ballard, M., & Noam, G. G. (2014). Examining the impact of afterschool STEM programs. *Afterschool Alliance*. Retrieved at: http://afterschoolalliance.org/ExaminingtheImpactof AfterschoolSTEMPrograms.pdf .
Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*(2), 275–313.
Levine, P. B., & Zimmerman, D. S. (2010). *Targeting investments in children: fighting poverty when resources are limited.* Chicago: University of Chicago Press.
Lieberman, G. A., & Hoody, L. L. (1998). *Closing the achievement gap: using the environment as an integrating context for learning - Results of a national study.* Washington D.C: Council of Chief State School Officers. Retrieved from http://www.seer.org/extras/execsum.pdf.
Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology,* 1–55.
Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Thousand Oaks, CA, US: Sage Publications, Inc.
Longford, N. T. (1993). *Random coefficient models.* Oxford: Clarendon.
Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education, 95*(5), 877–907.
Manufacturing Institute (2016). Manufacturing jobs over the next decade. Retrieved at: http://www.themanufacturinginstitute.org/Research/Other-Institute-Reports/~/media/24ECD17A43324F7696BF758A9A116B6F.ashx.
Marsh, H. W., Richards, G. E., & Barnes, J. (1986). Multi-dimensional self-concepts: A long term follow-up of the effect of participation in an Outward Bound Program. *Personality and Social Psychology Bulletin, 12*, 146–167.
McCombs, J. S., Augustine, C. H., Schwartz, H. L., Bodilly, S. J., McInnis, B., Lichter, D. S., & Cross, A. B. (2011). *Making summer count: how summer programs can boost children's learning.* Santa Monica, CA: The Wallace Foundation. Published by Rand.
Menard, S. W. (2002). *Longitudinal research* (second edition). Newbury Park, CA: Sage.
National 4-H Council (2016). Grow true leaders: National 4-H Council 2016 Annual Report. Retrieved at: https://4-h.org/wp-content/uploads/2016/03/2016-Annual-Report.pdf.
National Science Foundation (2016). Under-representation of Minorities in Science and Engineering occupations. Retrieved at: https://www.nsf.gov/statistics/2016/nsb20161/#/report/chapter-3/women-and-minorities-in-the-s-e-workforce.
National Science Board. (2016). *Science and engineering indicators 2016.* Arlington, VA: National Science Foundation. Retrieved at https://www.nsf.gov/statistics/2016/nsb20161/#/.
Oppenheim, A. N. (1992). *Questionnaire design, interviewing, and attitude measurement.* New York, NY: Printer Publishers.
Ord, J., & Leather, M. (2011). The substance beneath the labels of experiential learning: The importance of John Dewey for outdoor educators. *Australian Journal of Outdoor Education, 15*(2), 13–23.
Orr, D. W. (1992). Ecological Literacy: Education for a Post Modern World. Albany, NY: State University of New York.
Platt, G. (2008). The hard facts about soft skills measurement. *Training Journal, 8*, 53–56.
Quay, J., & Seaman, J. (2012). *John dewey and education outdoors.* Rotterdam: Sense Publishers.
Ramey, C., Bryant, D., Campbell, F., Sparling, J., & Wasik, B. (1988). Early intervention for high-risk children: The Carolina early intervention program. In R. Price, E. Cowen, R. Lorion, & M. Ramos-McKay (Eds.). *14 ounces of prevention: a casebook for practioners* (pp. 32–43). Washington, DC: American Psychological Association.
Rothwell, J. (2013). *The hidden STEM economy.* Washington, DC: Brookings Institution Metropolitan Policy Program. Retrieved at http://www.brookings.edu/research/the-hidden-stem-economy/.
Royal Horticulture Society. (2010). *Gardening in schools: a vital tool for children's learning. RHS campaign for school gardening.* London: Royal Horticulture Society.
Schweinhart, L., Barnes, H., & Weikart, D. (1993). *Significant benefits: the high/scope perry pre-school study through age 27.* Ypsilanti, MI: Hich Scope Press.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.
Singer, Joel. D. (1998). Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics, 23*(4), 323–355.
Smith, L. L., & Matsenbocker, C. E. (2005). Impact of hands-on science through school gardening in Louisiana public elementary schools. *HortTechnology, 15,* 439–443.
Snijders, Tom, & Bosker, Roel (1999). *Multilevel Analysis: an introduction to basic and advanced multilevel modeling.* London: Sage.
Soares, F., Babb, S., Diener, O., Gates, S., & Ignatowski, C. (2017). *Guiding principles for building soft skills among adolescents and young adults.* Washington, DC: USAID's Youth Power Action.
Somers, M-A., Welbeck, R., Grossman, J. B., & Gooden, S. (2015). *An analysis of the effects of an academic summer program for middle school students.* NY: MDRC.
Springer, K., & Diffily, D. (2012). The relationship between intensity and breadth of after-school program participation and academic achievement: Evidence from a short-term longitudinal study. *Journal of Community Psychology, 40*(7), 785–798.
State of New Jersey, Department of Education. (2018). Common core standards. Retrived at http://www.state.nj.us/education/archive/SCA/.
Steinberg, A. (1998). *Real learning.* New York: Real work.
Stewart, F. (2018). *The STEM dilemma: skills that matter to regions.* Kalamazoo, MI: W. E. Upjohn Institute.
Tai, R. H. (2012). *An examination of the research literature on project lead the way.*

Indianapolis, IN: Project Lead the Way.
Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, New Series, 312*(5777), 1143–1144.
Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal, 113*(485), F3–33.
U. S. Congress Joint Economic Committee Report. (2012). The 2012 Joint Economic Committee Report. https://www.jec.senate.gov/public/index.cfm/2012/12/report-142401bb-daa2-47f3-a550-3dba6c0a0b1e.
U. S. Department of Commerce. (2011). Education supports racial and ethnic equality in STEM" ESA Issue Brief #05-11, Retrieved at: http://www.esa.doc.gov/sites/default/files/education_supports_racial_and_ethnic_equality_in_stem.pdf.
U. S. Department of Education (2008). A nation accountable: Twenty-five years after A Nation of Risk. Retrieved at: https://www2.ed.gov/rschstat/research/pubs/accountable/accountable.pdf.
U. S. Department of Education Green Ribbon Schools. (2018). Green ribbon schools overview. Retrieved from https://www2.ed.gov/programs/green-ribbon-schools/index.html.
Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering and mathematics (STEM) and its relation to STEM education dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102*, 860–869.
Wooldridge, J. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.