

Cognitive and Noncognitive Changes From Participation in National Guard Youth ChalleNGe

Jennie W. Wenger • Jennifer R. Atkin



DRM-2013-U-005201-Final
October 2013

Members of Freestate ChalleNGe Academy class #34 march to their commencement at Aberdeen Proving Ground, Md., June 12. The academy, part of the National Guard Youth ChalleNGe Program, is a 22-week military orientated program that takes "at risk" youth and provides them with education, life-coping skills and other tools to become successful members of society. (U.S. Air Force photo by Tech. Sgt. Gareth Buckland/Released)

Approved for distribution:

October 2013

A handwritten signature in black ink that reads "David Rodney". The signature is written in a cursive style and is underlined with a single horizontal line.

David Rodney, Director
Fleet and Operational Manpower Team
Resource Analysis Division

This document represents the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Department of the Navy.

Approved for Public Release; Distribution Unlimited. Specific authority: **N00014-11-D-0323**
Copies of this document can be obtained through the Defense Technical Information Center at www.dtic.mil
or contact CNA Document Control and Distribution Section at 703-824-2123.

Copyright © 2013 CNA

This work was created in the performance of Federal Government Contract Number **N00014-11-D-0323**. Any copyright in this work is subject to the Government's Unlimited Rights license as defined in DFARS 252.227-7013 and/or DFARS 252.227-7014. The reproduction of this work for commercial purposes is strictly prohibited. Nongovernmental users may copy and distribute this document in any medium, either commercially or noncommercially, provided that this copyright notice is reproduced in all copies. Nongovernmental users may not use technical measures to obstruct or control the reading or further copying of the copies they make or distribute. Nongovernmental users may not accept compensation of any manner in exchange for copies. All other rights reserved.

Contents

Executive summary	1
Introduction and background	3
National Guard Youth Challenge	3
Noncognitive skills	4
The Washington Youth Academy	5
Data sources and methodology	7
Cognitive skills: TABE scores	7
Noncognitive skills: cadet survey	9
Results	13
Noncognitive skills	13
Predictive power of noncognitive measures	17
Adoption of the Khan Academy curriculum	19
Implications and recommendations	25
Appendix: Supplemental information	27
References	33
List of figures	35
List of tables	37

This page intentionally left blank.

Executive summary¹

In this research, we document our analysis of one National Guard Youth Challenge (ChalleNGe) program's data on participants' cognitive and noncognitive skills. We find that participants' (cadets') noncognitive skills increase substantially over the course of the five-month program. We also find that the program's recent adoption of an online math curriculum, presented through a facilitated instruction model, is associated with higher gains in math scores.

The ChalleNGe program serves 16- to 18-year-old high school dropouts and students at risk of dropping out. The program model includes substantial classroom instruction as well as a strong emphasis on noncognitive skills, such as leadership, planning, and determination. The Washington Youth ChalleNGe Academy (WYA) is part of the ChalleNGe program. WYA focuses on credit recovery—classroom instruction aimed at completing certified courses so that cadets can reenter their home high schools after ChalleNGe on track for graduation.

During the most recent program cycle (spring 2013), the WYA collected data on cadets' noncognitive skills by surveying cadets at the beginning and the end of the program. The survey included several potential measures of noncognitive skills, such as determination, confidence, and ability/willingness to follow directions. The program also collected data on cognitive skills from the Test of Adult Basic Education (TABE).

Our analyses indicate that cadets' noncognitive skills increased substantially during the program. At the beginning of the program, male and female cadets recorded different levels of various noncognitive

1. We are very grateful to Lauren Malone for providing a helpful review of the document, and to Molly McIntosh for her assistance during earlier phases of this project.

skills; female cadets showed higher levels of determination and a greater ability to follow directions, while male cadets showed higher levels of math confidence and locus of control (belief that one's actions influence eventual outcomes). By the end of the program, the measured noncognitive skills of both male and female cadets had improved on average, and the gender differences were no longer evident.

Particularly in the case of female cadets, we found that initial math confidence was a strong predictor of math success (measured by increases in math test scores). This suggests that focusing on math confidence at the beginning of the program could pay dividends, especially for female cadets.

The program's recent decision to adopt the Khan Academy math curriculum also is associated with increased gains on math test scores. In particular, applied math skills (measured by the test-taker's ability to solve math-based word problems) increased by an additional half a grade level compared with what we would have expected had the curriculum remained unchanged. Gains were larger for female cadets than for male cadets, and gains were larger for cadets who began the program with relatively low scores. Moreover, given the other cognitive outcomes, we believe that these results may understate the true effects somewhat. We suspect that these results are driven by the large variation in initial math skills across cadets who enter the program (standardized test scores suggest that cadets' initial math performance ranges from 1st- to 12th-grade level; providing cogent instruction to a group with such a range of backgrounds is extremely challenging, but the online curriculum allows each cadet to work at an appropriate level).

Of course, we would like to expand the dataset by adding data from upcoming classes. We also strongly recommend that the program collect more, and more detailed, information about the eventual success of cadets who return to their home high schools. At this point, however, we feel confident in stating that the program has a positive impact on noncognitive skills and in recommending that the program continue to use the online math curriculum.

Introduction and background

In this section, we provide some background information on the entire National Guard Youth Challenge (ChalleNGe) program. We also discuss some aspects of noncognitive skills. Finally, we provide information on the ChalleNGe site that is the focus of this analysis—the Washington Youth ChalleNGe Academy (WYA).

National Guard Youth Challenge

ChalleNGe is a quasi-military, 22-week residential program designed to serve 16- to 18-year-old high school dropouts and those at risk of dropping out.² (Students who have earned far fewer high school credits than expected are considered to be at risk.) A mentoring component follows the residential phase; participants (known as “cadets”) work with their mentors for at least another year.

The ChalleNGe program is funded jointly by DOD, the states, and the state National Guard units. Currently, there are 34 locations in 29 states, the District of Columbia, and the territory of Puerto Rico. Most ChalleNGe programs consider passing the General Educational Development (GED) tests to be the primary academic goal. However, some programs award alternate credentials, such as state high school diplomas to cadets who complete the program. Other programs focus on credit recovery so the cadet can reenroll in and graduate from his or her previous high school after completing ChalleNGe. In this model, programs provide coursework certified by state or local authorities; cadets who complete the program transfer these credits back to their home high schools. Finally, some programs are considered schools and award regular high school diplomas.

2. The program is quasi-military in the sense that participants live in barracks, wear uniforms, and take part in drills, marching, and regular physical training, but they are not military enlistees.

The ChalleNGe model is quite detailed; it includes eight core components: leadership/followership, responsible citizenship, service to community, life-coping skills, physical fitness, health and hygiene, job skills, and academic excellence. Academic progress can be followed through changes in standardized test scores, course completions, and credits/credentials awarded, but the other components are more difficult to measure. Indeed, many of these components depend heavily on the development of noncognitive skills. An emphasis on developing such noncognitive skills as long-term planning is a common aspect of many programs designed for preteens and teens.³ Given the ChalleNGe program's emphasis on noncognitive skills, it would be preferable to have a measure of such skills and, optimally, a measure of how they change during the course of the program.

Noncognitive skills

Noncognitive skills, sometimes referred to as “soft skills,” include many aspects of personality and attitude, such as communication skills, determination, leadership, ability to make and carry out plans, and timeliness. These skills generally are acknowledged to be important in the job market and in life but often have taken a back seat to cognitive skills (skills that are academic in nature, such as reading and mathematics proficiency) in the education and the economics literatures. In recent years, however, research emphasis has shifted to include and sometimes even focus on noncognitive skills (e.g., see [1]). While the literature is fairly wide ranging, it is clear that noncognitive skills are strongly associated with a wide variety of highly relevant outcomes, such as dropping out of, versus completing, high school, attending college, participating in the labor market, and the probability of arrest/incarceration (see [2], [3], and [4]).

A key aspect of noncognitive skills is that they can be developed throughout childhood and the young adult years [5, 6]. Indeed, an

3. For example, the Job Corps model includes academic and vocational skills as well as “employability skills and social competencies.” For more details, see www.jobcorps.gov/AboutJobCorps/program_design.aspx (last accessed June 24, 2013).

increase in noncognitive skills is the most likely explanation for the long-term success of participants of early childhood interventions, such as the Perry Preschool Project [4]. While cognitive and noncognitive skills are not completely unrelated, correlations are far from perfect, suggesting that they measure different attributes [2].

The Washington Youth Academy

In this work, we focus on one program, the WYA in Bremerton, Washington. This program uses a credit recovery model; cadets who complete the program are awarded high school credits for coursework completed at the academy and then return to their home high schools having made substantial progress toward graduation. During the most recent cycle of the program (January through June 2013), WYA used a survey to measure cadets' noncognitive skills. Cadets completed the survey at the beginning and the end of the program. At the same time, WYA moved to a new math curriculum based on a facilitated online model; cadets work independently using computers to access modules developed by the Khan Academy, but a math teacher is in the room at all times and provides input, guidance, and assistance as needed.⁴ Although the research on online learning in the K–12 arena is still fairly limited, findings suggest that such a curriculum is likely to be more effective than unstructured online learning and could provide better opportunities than more traditional classroom approaches (see [7]).

We analyze data provided by WYA to determine the extent to which cadets' noncognitive skills changed over the course of the program, to explore the relationships between noncognitive skills and other outcomes of interest, and to test the correlation between the new curriculum and cadets' gains in math. In the next section, we provide detailed information on our data, including the noncognitive measures on the survey that WYA used. Later sections of the paper present our results and our recommendations.

4. The Khan Academy is a nonprofit website with several thousand short videos and practice problems on a wide range of topics. For more information, see www.khanacademy.org (last accessed June 13, 2013).

This page intentionally left blank.

Data sources and methodology

In this analysis, we use several sources of data provided by WYA. First, the program collected cadets' scores on the Test of Adult Basic Education (TABE) at the beginning and the end of the program. (The program also collected and provided us with TABE scores on cadets who attended past sessions; we use this information to analyze the effects of the shift to the online math curriculum as discussed below). In addition, the program collected data indicating which cadets completed ChalleNGe. Finally, all cadets completed a survey that was designed to measure noncognitive skills; they completed the survey twice—once in early February (at the end of pre-ChalleNGe, right before classroom instruction began) and again during the last week of classes. In this section, we provide more information on each data source.

Cognitive skills: TABE scores

Our measure of cognitive skills is formed from the TABE, which cadets take at the beginning and the end of ChalleNGe. The TABE was designed for placement of adult learners and is often used as an assessment tool in adult education programs with a focus on completing the GED tests. Each subsection of the TABE is scored to indicate grade level (for example, a score of 9.3 indicates performance at the 3rd month of 9th grade).

We focus on the four subsections of the TABE, as well as on the total score (formed from averaging scores on the subtests). The subtests include Math, Applied Math, Reading, and Language. The Math section is made up of computational problems requiring test-takers to perform addition, subtraction, multiplication, and division, to work with percentiles, fractions, and exponents, and to solve basic algebra problems. The Applied Math section is made up of word problems, which require the following abilities: chart and table comprehension,

basic equation setup, coordinate graphing, an understanding of some limited geometry, and application of the concepts of fractions, percentiles, and algebra in the context of word problems. The Language section includes questions on grammar and punctuation, combining sentences to preserve their meanings, and some basics of paragraph composition. The Reading section involves reading passages or detailed charts/tables and answering questions about the content.

ChalleNGe cadets attending the WYA usually enter the program around the 6th-grade level in Math and near the 9th-grade level in Applied Math.⁵ They tend to come into the program scoring near the 7th-grade level in Language and about halfway through 8th-grade in Reading. However, these scores are averages, and the variation across cadets is substantial. The average cadet gains over 2 years on the TABE during the course of the program (suggesting their achievement levels increase by more than 2 school years in 5.5 months). Based on all TABE data from 2009 through 2013, the average cadet gains 2.2 years in Math, 1.7 years in Applied Math, 1.7 years in Reading, and 2.5 years in Language.⁶ Thus, average scores are lowest and average gains are highest in (computational) Math and (grammar/compositional) Language. Some of this difference may be driven by ceiling effects (cadets who score at least 10.5 on the TABE are limited to lower-than-average gains because the maximum score is 12.9).⁷

-
5. While it seems counterintuitive that cadets tend to score higher on Applied Math than on (computational) Math, we believe this difference occurs because most 9th graders will have fairly strong computational skills, but many will lag in math applications; thus, cadets may be more typical in terms of Applied Math than computational Math.
 6. Gain scores can be calculated only for cadets who complete the program; the completion rate at WYA from 2009 to 2013 was about 78 percent. This is a relatively high completion rate; across all programs, the completion rate was about 67 percent between 2006 and 2012.
 7. In past classes, about 20 percent of cadets entered WYA with TABE scores of 10.5 or higher.

Noncognitive skills: cadet survey

Our data include several measures of noncognitive skills based on the survey completed by cadets. The cadets completed the survey at the beginning of the program (right after the initial two weeks known as “pre-ChalleNGe” but before beginning classroom work); they completed an identical survey during the last week of the program.⁸ The survey included the following measures:

- Grit scale⁹
- Locus-of-control scale¹⁰
- Efficacy measures to determine cadets’ confidence in their math and science abilities¹¹
- Time preference—would cadets prefer to be paid \$50 today or \$100 in 6 months?
- Following directions—cadets were asked to read and follow instructions on a question about why they left their previous high school

First, the survey included the 8-item *grit* scale, designed to measure the respondent’s determination/tenacity. The answers range from

8. We wish to thank the WYA program staff, especially Mike Mittleider, Larry Pierce, Lynn Caddell, and Chris Acuna, for providing the data used in our analyses and for cheerfully answering our queries. The appendix includes additional details on the survey and the measures used, as well as the distributions of initial and final grit and locus-of-control scores (figures 3 and 4).

9. The grit scale was developed by and used with the permission of Dr. Angela Duckworth, Department of Psychology, University of Pennsylvania.

10. The locus-of-control scale was developed by and used with the permission of Dr. Julian Rotter, Emeritus Professor, Department of Psychology, University of Connecticut.

11. Efficacy scales were adapted from *Middle and High School STEM-Student Survey*, 2012, Raleigh, North Carolina, and used by permission of the Friday Institute for Educational Innovation, NC State University.

“Very much like me” to “Not like me at all” in the form of a 5-point Likert (rating) scale. The grit score is calculated by awarding points for stated determination; for example, one statement is, “I am a hard worker,” and another is “I often set 1 goal but later choose to pursue a different goal.” For the first statement, cadets received 5 points for selecting “Very much like me” and decreasing numbers of points down to 1 point for “Not at all like me.” For the second statement, cadets received 1 point for choosing “Very much like me” and increasing numbers of points to 5 points for “Not at all like me.” Scores range from 8 to 40 with higher scores indicating higher levels of determination, or grit. Figure 3 (in the appendix) shows initial and final distributions of measured grit among cadets and indicates a shift toward higher levels of grit over the course of the program.

Locus of control measures the extent to which a person believes that his or her own actions (versus random factors or other powers) determine outcomes. Essentially, the scale measures the extent to which respondents believe that they can control their lives. Those who believe that their own actions have consequences are designated “internal”; those who believe that other factors determine outcomes are termed “external.” Each question is a forced-choice format; the respondent chooses which of two statements best describes his or her beliefs/feelings. Respondents receive 1 point each time they choose a statement indicating they have control over situations; the score ranges from 0 (completely “external,” failing to see a relationship between actions and consequences/reactions) to 13 (completely “internal,” giving no explanatory power to luck). Figure 4 (in the appendix) shows the initial and final distributions of locus of control among cadets and indicates a shift toward more internal scores.

Efficacy is measured using a 5-point Likert scale of responses to a series of statements about the cadet’s attitude toward and confidence in math and science. We calculate math and science efficacy as separate variables; in each case, the efficacy score is determined by awarding points for responses that exhibit positive attitude and confidence in the subject. Thus, cadets who select “Strongly agree” for such statements as “I know I can do well in science” receive 5 points, as do cadets who select “Strongly disagree” for such statements as “I can handle most subjects well, but I cannot do a good job in science.”

Total efficacy scores are determined by adding the total number of points for the eight math and nine science questions and taking the average; thus, each efficacy score indicates the average response on the Likert scale with higher scores indicating higher efficacy. Scores range from 1 to 5. Figures 5 and 6 (in the appendix) show initial and final efficacy distributions among cadets. As is the case for grit and locus of control, these figures indicate a shift toward higher levels of efficacy over the course of the program.

Time preference is the fourth measure of noncognitive skills. A simple question asks whether the cadet would prefer to be paid \$50 today or \$100 in 6 months. Indicating a preference for \$100 in 6 months suggests a level of determination, planning, and self control.

Following directions is the final measure. At one point in the survey, cadets are asked why they left their previous high school. The survey presents a variety of reasons; cadets are instructed to mark all that apply *and* to circle the most important reason. All cadets marked at least one reason. We considered those who also circled a reason to have followed the directions and those who did not circle a reason to have not followed the directions.

Of the 152 cadets who entered the classroom portion of WYA in January 2013, 151 filled out the initial survey. During the classroom phase, 19 cadets left the program; thus, 133 cadets completed the program and the final survey. We have no post-ChalleNGe information on cadets who left during the classroom phase. Also, due to missing information, it was not possible to match 13 of the initial surveys to final surveys. We do know, based on program information, that 8 of the 13 cadets completed the program. Therefore, we have 125 complete, matched surveys (including pre- and post-ChalleNGe information).

In a few cases, cadets skipped questions or sections of the survey, but, overall, cadets answered the vast majority of the questions on the pre- and post-ChalleNGe surveys. In each case, we present the most complete information possible.

This page intentionally left blank.

Results

In this section, we present our results based on the WYA data sources. First, we focus on survey results and analyze how cadets' noncognitive skills changed over the course of the program. We also examine the relationship between noncognitive skills and program outcomes. Then, we use the TABE results from multiple WYA sessions to examine how cadets' math achievement changed after the adoption of the new online math curriculum.

Noncognitive skills

As discussed earlier, the WYA survey included several measures of noncognitive skills. Table 1 presents average scores on each measure. We list initial scores for all cadets who took the survey, as well as initial and final scores for the cadets who completed ChalleNGe.

Table 1. Noncognitive measures, before and after ChalleNGe^a

Noncognitive measure	Initial score		Final score, graduates
	All cadets	Graduates	
Grit score	24.7	25.0	28.7 [^]
Math efficacy	2.71	2.73	3.23 [^]
Science efficacy	2.90	2.89	3.03 [*]
Locus of control (internal)	6.57	6.55	8.46 [^]
Chose \$100 in 6 months (%)	53.4	53.9	80.0 [^]
Followed directions (%)	17.2	21.1	30.9 [*]
Number of observations	151	125	125

a. Data are from surveys collected by WYA. Initial data are collected at the end of pre-ChalleNGe (2 weeks into the program, at the beginning of classroom instruction); final data are collected during the last week of classes. See the previous section as well as the appendix for explanations of each noncognitive measure.

[^] Differences between initial and final score among graduates are statistically significant at the 1-percent level (likelihood of occurring by chance less than 1 in 100).

^{*} Differences between initial and final score among graduates are statistically significant at the 5-percent level (likelihood of occurring by chance less than 1 in 20).

Table 1 demonstrates two main ideas. First, among cadets who completed the program, noncognitive skills improved over the course of ChalleNGe; this can be seen by comparing the final two columns of the table. On average, cadets who completed the program scored higher than they had at the beginning of the program on each measure. Cadets' grit (determination) improved, they reported being more internal (were more likely to believe their actions influenced outcomes), and they had higher levels of efficacy (confidence) in both math and science. In addition, at the end of the program, cadets were more likely to choose \$100 in 6 months over \$50 today, suggesting an increase in self-control. Finally, based on one portion of the survey, cadets who completed ChalleNGe were more likely to read and follow directions than they had been at the beginning of the program. In each case, the differences were statistically significant, implying that the differences are unlikely to have occurred by chance.

Some of the measures on the WYA survey were adopted from existing instruments; exceptions are the questions about self-control and following directions. Data on the efficacy scores are limited, but some research on the grit scale and decades of research on the locus-of-control scale exist. Cadets' initial grit scales are lower than those of any other population reported, but many of the groups tested could be expected to have high levels of determination (e.g., Ivy League undergraduates, West Point cadets, and National Spelling Bee finalists). By the end of WYA, graduates' scales increased into the range found among all adults, as well as National Spelling Bee finalists; this suggests that the grit of cadets at the end of the program is likely to compare favorably with that of many of their peers (see [8]). Cadets' locus-of-control levels are considerably lower than levels reported for most groups; however, this may reflect the reality that the ability of most teens to control their lives is, in fact, quite limited (see [9]).

While noncognitive skills improve during the program, table 1 also suggests that most initial measures of noncognitive skills are unlikely to predict which cadets will complete ChalleNGe. This can be seen by comparing the initial scores for all cadets with the initial scores for those cadets who completed ChalleNGe (the first two columns of table 1). In the cases of grit, efficacy, locus of control (internality), and choice between \$50 today versus \$100 in 6 months, the average

initial scores for all cadets are very similar to the average initial scores for cadets who complete the program. This suggests that these measures are not predictive of success (if having higher levels of these measures were predictive of success, we would expect graduates to have significantly higher initial scores than others). In the case of reading directions, however, only 17 percent of all cadets initially read directions, but nearly 21 percent of those cadets who would go on to complete the program initially read the directions. While 21 percent is still quite low, the difference between these figures suggests that those who did not complete ChalleNGe were very likely *not* to have followed directions.¹² Finally, differences between initial scores of all cadets and initial scores of eventual graduates are insignificant at the 5-percent level, suggesting the differences are due to chance.

These results suggest that ChalleNGe has a substantial impact on cadets' noncognitive skills but that initial noncognitive skills in most cases do not predict program success. Thus, while the survey provides potential measures of ChalleNGe's influence on cadets, there is little reason to believe that cadets who initially have strong noncognitive skills (at least by most survey measures) will be more successful than others in the program. Ability to follow directions is an exception, perhaps because of the highly structured nature of the ChalleNGe program—that is, noncognitive skills may be much more likely to affect outcomes after completion of ChalleNGe.¹³ For example, cadets at ChalleNGe follow the program schedule and attend class as a group; after completing ChalleNGe, cadets return home and must take much more responsibility for attaining their educational goals. For this reason, the program's impact on cadets' noncognitive skills is likely to be a key outcome and is likely to be predictive over a range of longer term outcomes, such as completing high school, obtaining postsecondary education, and participating in the labor force. This

12. Many cadets who *did* complete ChalleNGe failed to follow directions on this section of the survey. But every single cadet who did *not* complete ChalleNGe failed to follow directions on this portion of the survey.

13. Consistent with this, differences in noncognitive skills are thought to be an explanation for differences in performance among graduates and nongraduates who enlist in the armed forces (e.g., see [10]).

would be consistent with the literature on noncognitive skills; see the foregoing discussion.

Next, we examine these measures by gender. Table 2 demonstrates that female cadets began the program with lower measures of efficacy (in science and in math) and were less internal than male cadets. However, female cadets began the program with higher levels of grit, and were more likely to read and follow directions on the first survey.

Table 2. Initial and final scores on noncognitive measures, by gender^a

Noncognitive measure	Initial score of cadets		Final score of cadets	
	Female	Male	Female	Male
Grit score	26.6	23.8 [^]	28.3	28.8
Math efficacy	2.49	2.81 [*]	3.00	3.35
Science efficacy	2.86	2.92	3.13	2.99
Locus of control (internal)	5.89	7.07 [^]	8.44	8.48
Chose \$100 in 6 months (%)	50.0	55.5	74.4	82.9
Followed directions (%)	27.1	12.6 [^]	35.0	28.9

a. Data are from surveys collected by WYA. Initial data are collected at the end of pre-ChalleNGe (2 weeks into the program, at the beginning of classroom instruction); final data are collected during the last week of classes. See the previous section as well as the appendix for explanations of each noncognitive measure.

[^] Differences between men and women are statistically significant at the 1-percent level (likelihood of occurring by chance is less than 1 in 100).

^{*} Differences between men and women are statistically significant at the 5-percent level (likelihood of occurring by chance is less than 1 in 20).

Female cadets experienced large gains in terms of locus of control; male cadets had very large gains in terms of grit. By the end of the program, average measures of these two skills were very similar between men and women. Male cadets had very large gains in terms of following directions (recall that this is the only noncognitive skill that is obviously related to program success; refer to table 1).

Our findings indicate that cadets' noncognitive skills increased substantially over the course of WYA. However, initial skills generally are not related to program success as defined by graduation. Next, we take a closer look at success and the potential relationships between test scores, test score gains, noncognitive skills, and noncognitive skill gains.

Predictive power of noncognitive measures

While it is interesting and instructive to examine the differences among our noncognitive measures, an exploration of how these measures are related to program success is likely to yield the most actionable set of recommendations for the ChalleNGe program. Therefore, we next model program success as a function of individual characteristics, including TABE scores and noncognitive measures. Our first outcome variable is completion of the ChalleNGe program.¹⁴

Because we have a small dataset, we estimate a number of very parsimonious equations, including only a few variables. Also, recall from table 1 that initial noncognitive measures are very similar between all cadets and those cadets who go on to complete the program. This suggests that most noncognitive measures are likely to have relatively small impacts on program completion.

We experimented with numerous specifications. As suggested by the descriptive statistics above, noncognitive measures have little impact on program completion, together or separately. We could not test the effect of “reads directions” on program completion since every cadet who did not complete the program also failed to follow directions on the initial survey.¹⁵ (Missing information meant that some of the pre- and post-ChalleNGe surveys could not be matched; for this reason, we urge caution in interpreting our results on program completion).

We were, however, able to test specifications including both noncognitive measures and cognitive (various TABE) measures. We found

14. Our outcome of interest (“dependent variable”) is a dichotomous variable: cadets complete ChalleNGe or they do not. Therefore, we use a logistic (logit) regression. Because logit regressions yield coefficients that are related to marginal effects in a nonlinear manner, interpreting the regression results is not straightforward. Thus, we calculate and present marginal effects holding all other variables constant; the appendix contains regression results.

15. This creates a situation of perfect collinearity; the correlation between not reading directions and not completing the program is 1.0. Therefore, we cannot estimate an effect size. This does suggest, however, that reading directions is likely to be an important explanatory variable.

that one TABE score—the score on Reading—had explanatory power over program completion; cadets with lower Reading scores were less likely to complete the program. Language scores had similar effects but did not achieve statistical significance; neither Math nor Applied Math scores predicted program completion. Specification tests suggest that the effect is linear—that cadets who enter the program reading one grade level higher are about 2 percentage points more likely to complete the program.¹⁶ While this may not sound like a large effect, the results suggest that a cadet who enters the program at the 6th-grade reading level is more than twice as likely *not* to complete the program as a cadet who enters at the 10th-grade reading level. Between those entering at the 7th-grade versus the 9th-grade level, the chance of leaving the program varies from 10 percent to 6.5 percent. Thus, initial reading level is highly correlated with success. This suggests that WYA could place more emphasis on initial reading skills in making program acceptance decisions. Of course, cadets who enter with relatively low reading levels could have other characteristics that decrease the chance of success; we recommend that the program carefully track the relationship between initial reading levels and success in future classes.

Finally, we analyzed the relationships between noncognitive and cognitive skills. Specifically, we wondered whether cadets with higher levels of noncognitive skills would make more cognitive progress. The most obvious measure of this is provided by increases in TABE scores. Again, we tested many specifications, keeping each as parsimonious as possible because of the limited sample size.

We found that, in the case of Math scores, one noncognitive measure is especially important: initial math efficacy offers substantial explanatory power over the gains that cadets made in Math while at WYA. (The effect holds for the Math test, but *not* for the Applied Math test. Therefore, the effects occur in basic computation, rather than in applied word problems.)

Our results indicate that this effect is driven largely by female cadets. In addition, our results suggest that the initial efficacy score has a

16. Complete regression results appear in table 3 of the appendix.

larger effect on the final Math score than the initial Math score.¹⁷ Note that the cadets, especially the female cadets, in this class entered with very low Math scores, and with quite low math efficacy scores (refer to table 2). We found that the *change* in efficacy over the course of the program had little impact.¹⁸ This suggests that, especially for female cadets, working to educate and convince them of their ability to achieve in mathematics *before* beginning classroom work could pay off.

Of course, we measured this effect while WYA was using the Khan Academy online math curriculum; it is not clear how results would have differed under another curriculum. Next, we discuss the adoption of the Khan curriculum and how test scores have changed with this adoption.

Adoption of the Khan Academy curriculum

WYA has moved toward an online curriculum; in particular, math classes now use online materials. Because cadets who enter ChalleNGe initially test at a wide variety of grade levels, especially in math, presenting material of appropriate difficulty for an entire class can be difficult to impossible. The Khan Academy provides math modules at a variety of levels; cadets work through the modules at their own pace, but do so in a classroom with a teacher available for guidance and to work with individuals.

The ideal manner in which to measure the effect of a new curriculum would be to randomly divide the cadets in a given class into two groups and expose one group to the new curriculum while instructing the other group via the old curriculum. This is impractical for a small

17. The two scores are measured in different units, but this still suggests that initial efficacy is an important determinant of Math improvement. See table 4, appendix, for regression results.

18. We also found that this is not an effect of general confidence; science efficacy had no explanatory power in this regression. Other noncognitive measures do not add explanatory power either. Finally, although the effect of initial math efficacy on Applied Math is positive, the result is not statistically significant (and, therefore, is somewhat likely to have occurred by chance).

program. Therefore, we measure the effect of the Khan curriculum by comparing the math outcomes of cadets in the current class (who used the Khan curriculum) with those in prior classes (who did not). Our “control group” is made up of cadets at the WYA program in the cycles commencing in 2009. (We compare current gains with those in the seven classes commencing from January 2009 to January 2012; we consider the cadets in the fall cycle of 2012 to have been in transition because adoption of the Khan program was taking place during that cycle.)

Figures 1 and 2 present initial and final test scores in Math/Applied Math and Language/Reading, respectively. The figures demonstrate that, over time, both initial and final test scores have been fairly constant, with a slight downward trend. In particular, figure 1 suggests that math gains were higher for the most recent class than for many of the earlier classes, but figure 2 suggests that gains in Reading and Language were perhaps lower, and surely no higher, than for earlier classes.

Figure 1. Initial and final scores in Math and Applied Math, by WYA class

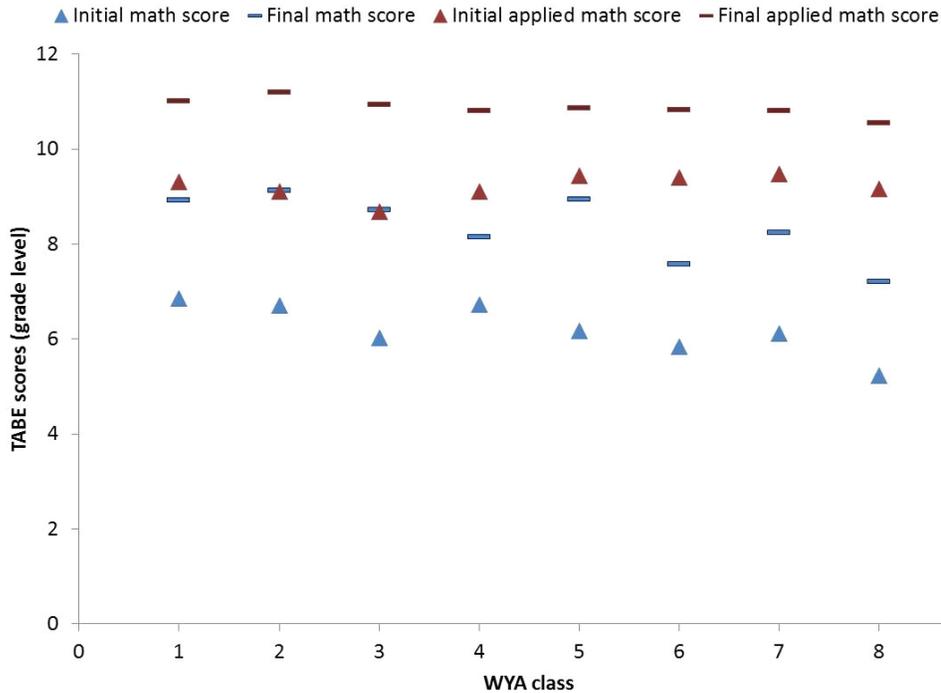
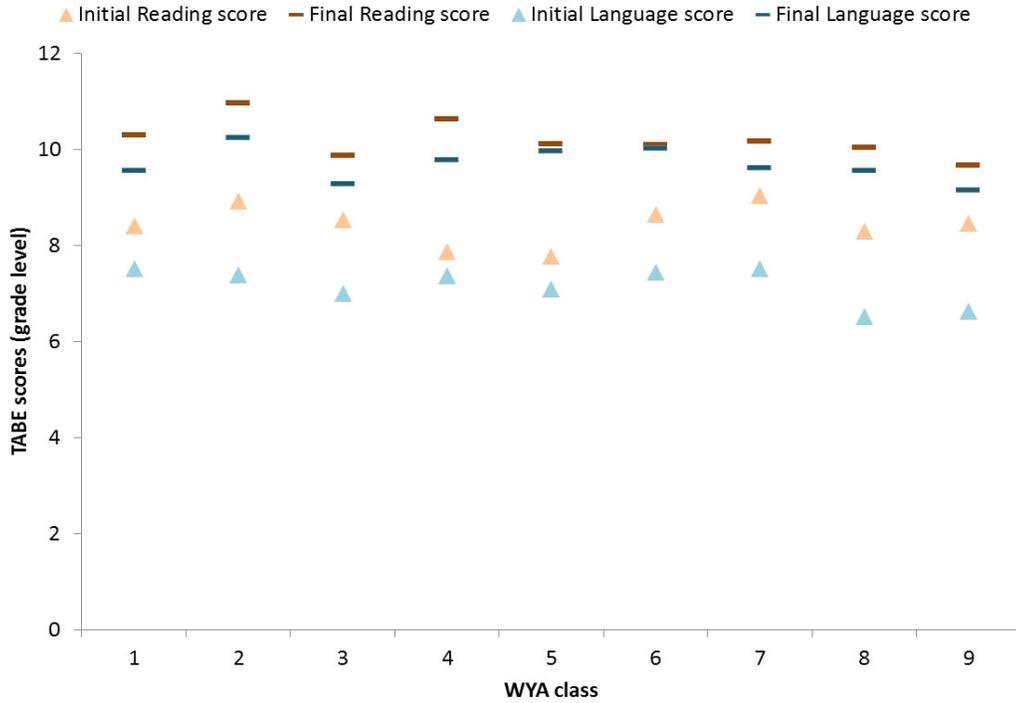


Figure 2. Initial and final Reading/Language scores, by WYA class



To test our results more formally, we pool the data and run regressions explaining final scores as a function of initial scores, gender, fall versus spring class, and having the Khan Academy mathematics curriculum in place (we ran models with each subscore but include only the most relevant here).¹⁹

Our results indicate that Applied Math gains were higher in the class that used the Khan Academy curriculum. Our results also suggest that Math gains may have been higher, but the results do not achieve statistical significance. In contrast, gains in Reading and Language appear *lower* in the class using the Khan curriculum than in other

19. See table 5 in the appendix for complete regression results. To allow for nonlinearities, we model initial scores using several categories. We include the “fall” indicator variable because program staff report that cadets in fall versus spring sessions differ somewhat in terms of preparation and attitude.

classes. (The result achieves significance at the 10-percent level for Reading, but not for Language.) These results suggest that the Khan Academy curriculum has increased Math achievement as measured by TABE scores; Applied Math scores increased about 0.5 year more than they would have under the previous curriculum. The unusually low gains in Reading/Language suggest that cadets in this class might have had lower-than-expected gains in Math had it not been for the introduction of the new curriculum. Therefore, our results may be an understatement of the true effect of the Khan curriculum, which could be as large as 0.6 to 0.8 year. This suggests that the cadets gained *at least* half a year in terms of Applied Math after the new curriculum was put into place.

When we split the sample by gender, we found that the effects were bigger for female cadets than for male cadets. Specifically, female cadets instructed using the Khan curriculum experienced a 0.8-year advantage over other female cadets, versus a 0.3-year advantage for male cadets. In other words, female cadets in the current class gained an additional 0.8 year in Applied Math compared with the gains of female cadets in earlier classes. Recall that entering Math scores, and math efficacy scores, were lower among female cadets than among male cadets. Therefore, this effect of utilizing the Khan Academy curriculum is particularly relevant for raising overall test scores and cognitive skills.

When we split the sample by initial TABE scores, we found that cadets who entered with the lowest TABE scores had the largest gains in terms of mathematics. Specifically, those whose initial TABE scores were below 6.0 (indicating they were performing below the 6th-grade level) gained an additional full year in terms of Applied Math compared with similar cadets in earlier classes. Our models suggest that the effect of the Khan Academy curriculum on the cadets who initially perform at or above the 9th-grade level is zero. This does not imply that these cadets' test scores are unchanged throughout the program. Rather, it implies that the highest performing cadets achieve about what they would have under the earlier curriculum. This could represent a ceiling effect; it is quite likely that these cadets work on material that is more advanced than that on the TABE. In this case, we would expect that these cadets' future performance will be higher than it

would have been under the previous curriculum, but the TABE may not reflect this.

We would like to examine data from future classes to make sure that factors particular to this class are not driving our results. However, our results at this point strongly suggest that the Khan Academy math curriculum should be kept in place.

This page intentionally left blank.

Implications and recommendations

Our findings suggest that the WYA ChalleNGe program has a substantive impact on cadets' noncognitive skills. Over the course of the program, cadets' stated grit (determination), locus of control, academic efficacy (confidence), willingness to wait for long-term payoffs, and ability to follow directions all increased significantly. Given the common aspects of the ChalleNGe model across the 34 programs, we would expect that cadets in other programs would experience similar gains. However, our data include only information on the WYA program.

At the same time, we find only limited evidence that initial measures of noncognitive skills predict successful completion of ChalleNGe. We urge caution in interpreting these findings because our information indicating which cadets completed the program was incomplete due to an inability to match some pre- and post-ChalleNGe survey data. However, our findings at this time suggest that selecting potential cadets based on noncognitive skills is not likely to be beneficial to the program or the cadets.

We suspect that the increases in noncognitive skills that occur during ChalleNGe will have large and substantial impacts on the cadets after they have left the program. At this point, however, we do not have any data to test this hypothesis. Partly for this reason, we strongly recommend that WYA (and other programs that focus on credit recovery) begin to keep detailed records on the progress of cadets after they leave the program and return to their home high schools. Indeed, we suspect that several of the noncognitive measures on the survey will have predictive power over cadets' likelihood of completing high school and attending postsecondary institutions, but testing this would require detailed data on cadets throughout the mentoring phase of the program and beyond.

While the noncognitive measures do not affect program success, we did find that the initial reading level was an important predictor of success. In particular, cadets who enter the program at less than a 6th-grade level struggle to complete WYA. This finding suggests either the need for an additional intervention to raise reading levels of some cadets before entering the classroom or additional work with these cadets early in the classroom phase. Supplementing classroom work with an online curriculum may be helpful; it is also possible that these cadets have specific learning disabilities that would require different interventions. (Our data included no information on prior individualized education programs (IEPs), etc.)

We found that math efficacy is a strong predictor of gains in Math scores. (Recall that the Math subtest on the TABE focuses on simple computational problems.) Especially among female cadets, math efficacy seems to be an important indicator of math gains during the program. However, female cadets have quite low levels of math efficacy at the beginning of the program (see table 2). This suggests the need for a specific intervention to increase math confidence, *before* beginning classroom work in math. A short instruction unit on the various ways in which people gain math skills might be helpful; information on noncognitive skills and the relationship between confidence and performance could be helpful as well.

Finally, our results suggest that adoption of the online Khan Academy math curriculum has had positive effects on test score gains, especially in Applied Math. Recall that the Applied Math subtest requires cadets to use math concepts to solve word problems, sometimes involving charts and tables. These problems are designed to mimic uses of math in the real world versus the classroom. It is particularly interesting to note that the gains in Math were concentrated among female cadets, and among cadets whose initial TABE scores indicated that they performed below the 6th-grade level on entering WYA. In addition, the Khan curriculum is not harmful to the highest achieving cadets; their progress on the TABE tests is similar to that shown by earlier classes, and they may be gaining math skills not included on the TABE. Although adopting an online curriculum implies technical challenges, the results so far suggest that cadets benefited from this change.

Appendix: Supplemental information

Figures 3, 4, 5, and 6 present more detailed information on non-cognitive measures included in the pre- and post-ChalleNGe WYA surveys.

Figure 3. Grit (i.e., determination) scores among cadets in pre- and post-ChalleNGe surveys

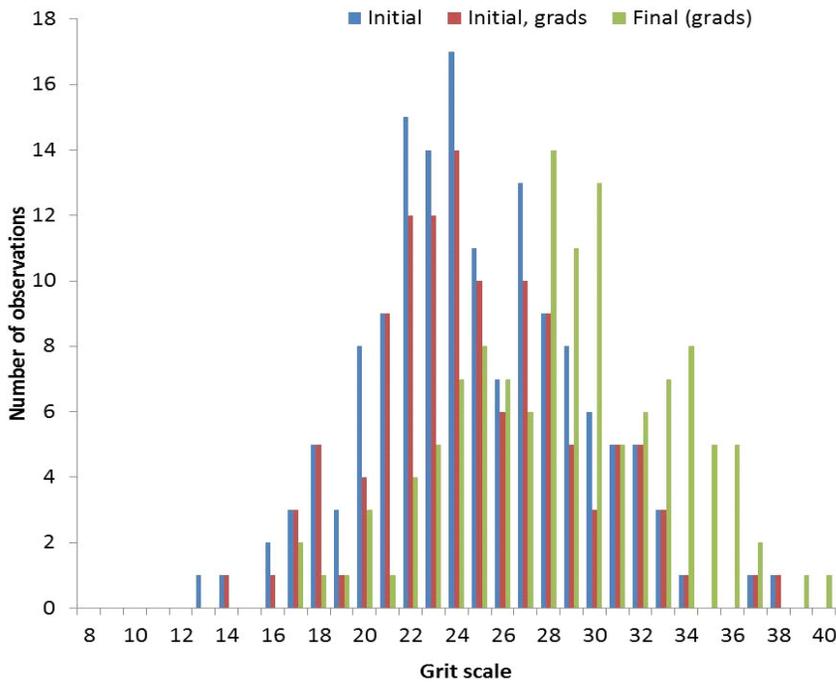
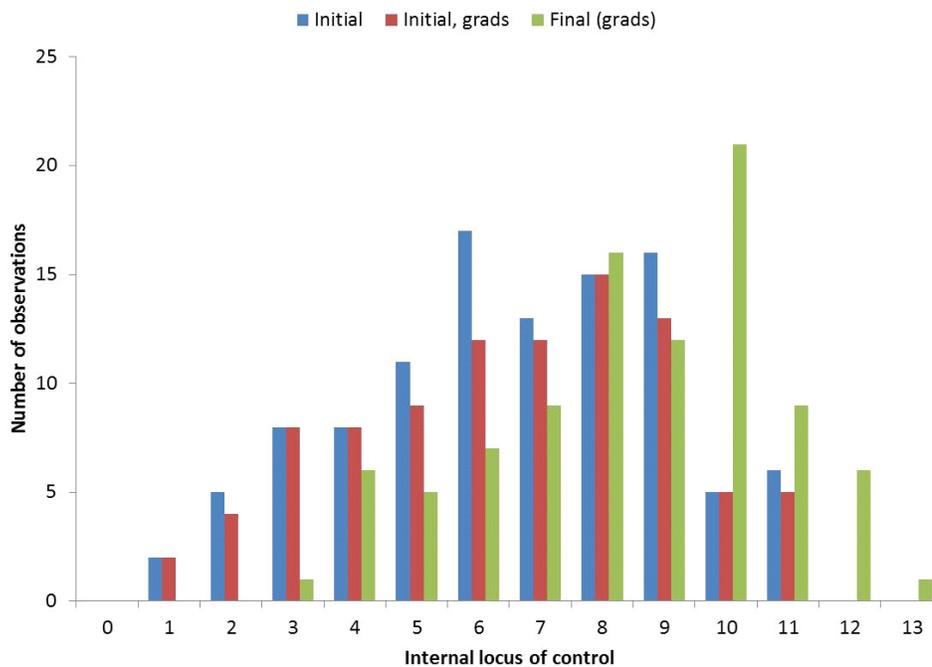


Figure 3 demonstrates that cadets who complete the program had higher measured levels of grit than they did at the beginning of the program. This can be seen by comparing the distribution of the red and green bars in the figure—the green bars (representing the final grit scores for graduates) are shifted farther right than the red bars (representing the initial grit scores for those who go on to graduate),

indicating a greater prevalence of higher grit scores among cadets at the end of the program than the beginning.

Figure 4 shows the distribution of locus-of-control scores among cadets at WYA at the beginning and the end of the program. Consistent with table 1, figure 4 demonstrates that cadets who completed the program had higher measured levels of locus of control (a more “internal” world view) than they did at the beginning of the program.

Figure 4. Locus of control among cadets in pre- and post-ChalleNGe surveys^a



a. In some cases, we were unable to match pre- and post-ChalleNGe surveys; see appendix for more details. Therefore, final (grads) data include only data for which we also have initial observations.

Also consistent with table 1, figures 5 and 6 show that cadets’ math and science efficacy (confidence) are higher at the end of the program than at the beginning.

Figure 5. Math efficacy scores among cadets in pre- and post-ChalleNGe surveys

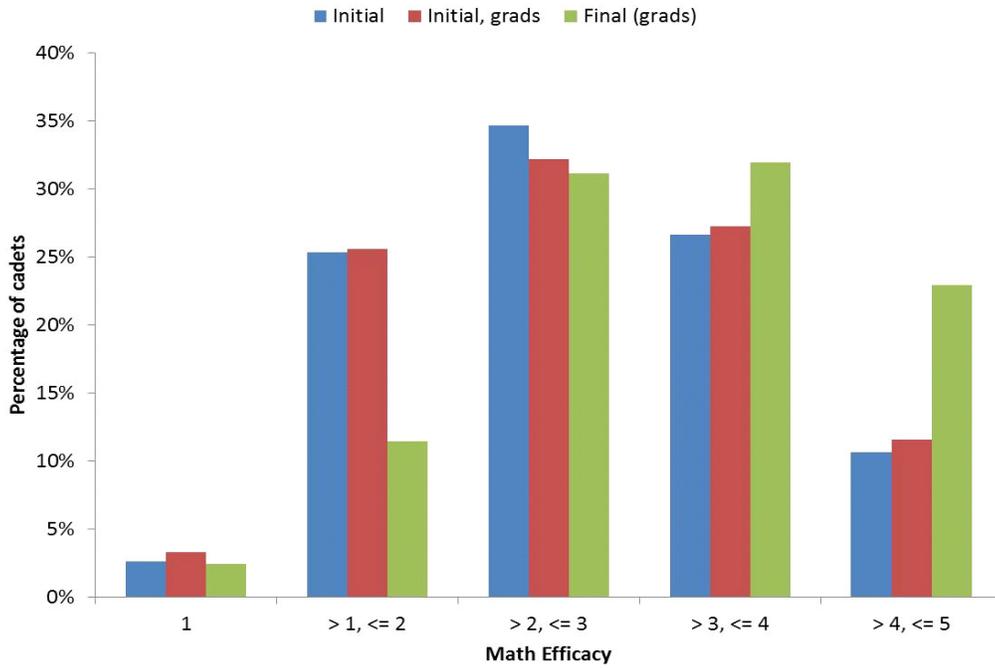
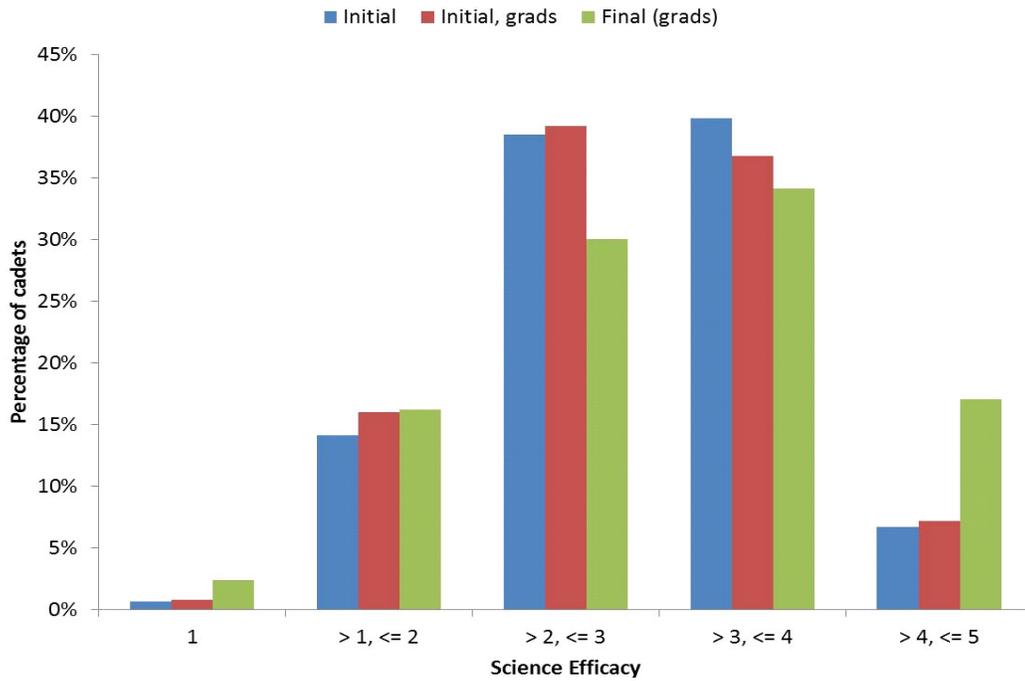


Figure 6. Science efficacy scores among cadets in pre- and post-ChalleNGe surveys



Tables 3, 4, and 5 include regression results discussed in the main text. Table 3 demonstrates that initial reading (as measured by the Reading TABE score) is negatively associated with dismissal and, thus, positively associated with completion of the program. In contrast, there is no relationship between program dismissal and initial math (measured by TABE Math test). We tested alternate specifications, including indication of gender, but we consider the results in table 3 to be our preferred specification because male cadets were disproportionately likely to leave the program before the classroom phase (and the pre-ChalleNGe survey) and because some of the male cadets' survey information could not be matched to test scores and other outcomes.

Table 3. Regression results: Outcome—ChalleNGe dismissal (noncompletion)^a

Variable	Coefficient	Standard error	Marginal effect
Initial grit	-0.0059	0.081	-.04%
Initial math	0.077	0.15	0.5%
Initial reading	-0.26*	0.12	-1.7%
Constant	-0.67	1.93	~

a. Regression includes 138 observations (all cadets with complete matched test score and survey data). Pseudo R-squared = 0.14. Initial grit measured by grit scale, developed by Dr. Angela Duckworth. Initial reading measured by TABE Reading test, initial math measured by TABE Math test.

* Indicates coefficient is significant at the 5-percent level or better and, thus, is likely to occur by chance fewer than 1 time in 20.

Table 4 includes results of a simple model of final math scores as a function of initial scores and initial math efficacy. We ran this model separately for male and female cadets. We do not include marginal effect columns because in a linear model such as this one marginal effects are equal to the estimated coefficients. Thus, for every 1-year increase in the initial Math TABE score, female cadets are expected to gain an additional 0.4 year in the final Math TABE score and male cadets are expected to gain 0.7 year. Initial math efficacy is positively associated with final Math TABE scores, but only for female cadets.

Table 4. Regression results: Outcome—Final Math TABE score^a

Variable	Female cadets		Male cadets	
	Coefficient	Standard error	Coefficient	Standard error
Initial math score	0.417**	0.131	0.697**	0.075
Initial math efficacy	0.747**	0.304	0.318	0.210
Constant	3.79**	0.88	3.78**	0.72

a. Regression includes 79 observations on male cadets and 39 on female cadets (all cadets with complete matched test score and survey data). Pseudo R-squared = 0.53 for men and 0.31 for women. Math efficacy measured by scale developed by Friday Institute, NC State University. Initial math measured by TABE Math test.

** Indicates coefficient is significant at the 2-percent level or better and, thus, is likely to occur by chance fewer than 1 time in 50.

Table 5 presents regression results explaining final Applied Math (TABE subtest) score as a function of initial Applied Math score, gender (in the initial specification), fall versus spring session, and use of the Khan Academy online curriculum.

Table 5. Regression results: Outcome—Final Applied Math score^a

Variable	Cadets						Initial TABE < 6.0		Initial TABE > 9.0	
	All		Female		Male		Coef.	SE	Coef.	SE
Initial < 4	2.12	1.80	-3.24**	0.64	2.10	1.80	1.89	2.39	~	~
Initial 4–5	1.37	1.80	-3.59**	0.73	1.28	1.79	1.10	2.40	~	~
Initial 5–6	3.02^	1.80	-2.30**	0.60	3.03^	1.78	2.73	2.39	~	~
Initial 6–7	3.28^	1.79	-2.14**	0.55	3.35*	1.77	2.77	2.39	-1.37**	0.46
Initial 7–8	3.87*	1.79	-1.86**	0.65	4.04*	1.78	3.17	2.42	~	~
Initial 8–9	4.47**	1.80	-1.00^	0.61	4.58**	1.78	4.20^	2.43	-0.25	0.44
Initial 9–10	5.09**	1.80	~	~	5.04**	1.78	4.88*	2.45	0.21	0.40
Initial 10–11	5.46**	1.79	-0.09	0.60	5.57**	1.77	5.31*	2.52	0.017	0.36
Initial > 11	6.53**	1.79	1.29**	0.54	6.52**	1.76	5.54*	2.49	0.78**	0.34
Male	0.20^	0.12	~	~	~	~	~	~	~	~
Fall	-0.034	0.12	0.012	0.23	-0.046	0.136	-0.011	0.27	-0.039	0.081
Khan	0.45**	0.18	0.76**	0.35	0.35^	0.21	0.93*	0.41	-0.045	0.14
Constant	5.90**	1.79	11.2**	0.51	6.10**	1.76	6.10**	2.37	12.0**	0.34

a. Regressions include 1,077 observations: 763 males, 314 females, 353 with initial TABE scores below 6.0, 384 with initial TABE scores above 9.0 Coef. = coefficient; SE = standard error.

^ Indicates coefficients is significant at the 10 percent level.

* Indicates coefficient is significant at the 5 percent level.

** Indicates coefficient is significant at the 1 percent level.

We find that using the Khan curriculum was associated with higher final Applied Math scores and, thus, with higher gains in Applied Math. This effect is larger for female cadets, and for cadets whose initial average TABE scores were less than 6.0, indicating that they entered the program below the 6th grade level. The use of the Khan Academy has a very small, negative, and insignificant effect on the final Applied Math score of cadets who begin the program with TABE scores of at least 9.0. This indicates that such cadets gain almost the same amount from using Khan as they would in other circumstances; given the wide scope of material contained in the Khan Academy, it is also likely that these cadets make gains in math that are not measured on the TABE.

References

- [1] Flavio Cunha and James J. Heckman. *Investing in Our Young People*. National Bureau of Economic Research (NBER) Working Paper No. 16201. Jul. 2010.
- [2] Margo Coleman and Thomas DeLeire. “An Economic Model of Locus of Control and the Human Capital Investment Decision.” *The Journal of Human Resources* 38, no. 3, 2003: 701–721.
- [3] Samuel Bowles, Herbert Gintis, and Melissa Osborne. “The Determinants of Earnings: A Behavioral Approach.” *Journal of Economic Literature* 39, no. 4, 2001: 1137–1176.
- [4] James Heckman et al. “The Rate of Return to the HighScope Perry Preschool Program.” *Journal of Public Economics* 94, no. 1-2, 2010:114–128.
- [5] Flavio Cunha et al. “Interpreting the Evidence on Life Cycle Skill Formation.” In Eric Hanushek and Finis Welch, eds. *Handbook of the Economics of Education*. Amsterdam: North-Holland, 2006. 697–812.
- [6] Paul Tough. *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. New York, New York: Houghton Mifflin Harcourt, 2012.
- [7] Angela Duckworth et al. “Grit: Perseverance and Passion for Long-Term Goals.” *Journal of Personality and Social Psychology* 92, no. 6, 2007: 1087–1101.
- [8] Julian Rotter. “Generalized Expectancies for Internal Versus External Control of Reinforcement.” *Psychological Monograph* 80, no. 1, 1966.
- [9] Barbara Means et al. *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*. U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. 2010.

- [10] Janice H. Laurence, Peter F. Ramsberger, and Jane Arabian.
Education Credential Tier Evaluation. Human Resources
Research Organization Final Report FR-PRD-844. Sep 1996.

List of figures

Figure 1.	Initial and final scores in Math and Applied Math, by WYA class	20
Figure 2.	Initial and final Reading/Language scores, by WYA class	21
Figure 3.	Grit (i.e., determination) scores among cadets in pre- and post-ChalleNGe surveys	27
Figure 4.	Locus of control among cadets in pre- and post-ChalleNGe surveys	28
Figure 5.	Math efficacy scores among cadets in pre- and post-ChalleNGe surveys	29
Figure 6.	Science efficacy scores among cadets in pre- and post-ChalleNGe surveys	29

This page intentionally left blank.

List of tables

Table 1.	Noncognitive measures, before and after ChalleNGe	13
Table 2.	Initial and final scores on noncognitive measures, by gender	16
Table 3.	Regression results: Outcome—ChalleNGe dismissal (noncompletion)	30
Table 5.	Regression results: Outcome—Final Applied Math score	31
Table 4.	Regression results: Outcome—Final Math TABE score	31

This page intentionally left blank.

