

Appendix

Appendix A1 Study characteristics: Millenky, Bloom, and Dillon, 2010

Characteristic	Description
Study citation	Millenky, M., Bloom, D., & Dillon, C. (2010). <i>Making the transition: Interim results of the National Guard Youth ChalleNGe evaluation</i> . New York: MDRC.
Participants	<p>Between 2005 and 2007, a total of 3,074 high school dropouts between the ages of 16 and 18 were randomly assigned to the <i>National Guard Youth ChalleNGe Program</i> in 10 states; 2,320 to the intervention group and 754 to the control group. There was a single site in each state. A random subsample of 1,508 participants (916 in the intervention group and 592 in the control group) was selected to be given the 21-month follow-up survey. The analysis sample included 1,196 youths (736 in the intervention group and 460 in the control group) who responded to the follow-up survey; 80% of the intervention group and 78% of the control group responded to the survey.</p> <p>Most sample members (84%) were male. They ranged in age from 16 to 18 years old at program entry. The sample was racially and ethnically diverse: 41% were Caucasian; 40% were African-American; 14% were Hispanic; and 5% were from other racial and ethnic groups. Sample members had performed poorly in school before entering the program. About half reported receiving mostly Ds and Fs before dropping out.</p>
Setting	The study took place in 10 sites in California, Florida, Georgia, Illinois, Michigan, Mississippi, New Mexico, North Carolina, Texas, and Wisconsin.
Intervention	<p>The <i>National Guard Youth ChalleNGe Program</i> attempts to promote positive youth development for high school dropouts through a residential education and training program followed by a structured mentoring program. The intervention consisted of three phases.</p> <p>Phase 1: Two weeks of program orientation and physical and psychological assessment in a residential, quasi-military setting.</p> <p>Phase 2: Twenty weeks of education and training in a residential, quasi-military setting. During this phase, the majority of participants' time was spent on educational activities, including work toward a GED or high school diploma. Education and training covered the following eight areas: leadership, responsible citizenship, service to community, life-coping skills, physical fitness, health and hygiene, job skills, and academic excellence.</p> <p>Phase 3: One year postresidential phase. Participants worked with program staff to arrange a postresidential placement in employment, continued education, or military service. In addition to participating in the placement activity, each participant was supposed to maintain monthly contact with a mentor, who was chosen by the participant and trained by <i>ChalleNGe</i> program staff.</p>
Comparison	Control group members were not eligible to receive <i>National Guard Youth ChalleNGe Program</i> services but could receive other services available in the community. Contrary to the study design, eight control group members enrolled in the program (1% of the full control group). These enrollees were still assigned to the control group in all analyses.
Primary outcomes and measurement	The relevant study outcome included in this review is whether students earned a high school diploma or GED, based on student follow-up interviews. For a more detailed description of this outcome measure, see Appendix A2. The study also examined a number of other outcomes that are not within the scope of the Dropout Prevention protocol.
Staff/teacher training	Program staff included team leaders who directly supervised the students, teachers who provided classroom instruction, and counselors who provided individual, group, and career counseling. It was common for staff members, particularly team leaders, to be National Guard members or to have military experience. Teachers often came to the program from the local school district or from community colleges. Counselors typically had bachelor's or advanced degrees in psychology, social work, or other relevant fields.

Appendix A2 Outcome measure for the completing school domain

Outcome measure	Description
Earned a diploma or GED certificate	This binary outcome was measured based on a follow-up survey conducted 21 months after random assignment. As a part of this survey, youth reported whether they had earned a high school diploma or a GED certificate.

Appendix A3 Summary of study findings included in the rating for the completing school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study			WWC calculations		
			Mean outcome		Mean difference ² (<i>ChalleNGe</i> – comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
			<i>ChalleNGe</i> group	Comparison group				
Millenky, Bloom, & Dillon, 2010⁶								
Earned a Diploma or GED Certificate	Survey respondents	1,196	0.61	0.36	0.24	0.60	Statistically significant	+22
Domain average for completing school⁷						0.60	Statistically significant	+22

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the completing school domain. Subgroup findings from the same study are not included in these ratings but are reported in Appendices A4.1 and A4.2.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
3. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of binary outcome variables, effect sizes are calculated using the Cox effect-size index. The Cox index is based on logged odds ratios; therefore, standard deviations are not needed for the calculation of effect sizes for binary outcome measures, such as whether youth have earned a high school diploma or GED.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No corrections for clustering or multiple comparisons were needed.
7. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of study findings for separate outcomes for the completing school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study			WWC calculations		
			Mean outcome		Mean difference ² (ChalleNGe – comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
			ChalleNGe group	Comparison group				
Millenky, Bloom, & Dillon, 2010⁶								
Earned a Diploma	Survey respondents	1,196	0.22	0.16	0.06	0.22	Statistically significant	+9
Earned a GED Certificate	Survey respondents	1,196	0.48	0.22	0.26	0.73	Statistically significant	+27

1. This appendix presents findings for outcome measures that disaggregate whether a respondent earned a high school diploma or a GED certificate. The combined outcome measures were used for rating purposes and are presented in Appendix A3.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
3. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of binary outcome variables, effect sizes are calculated using the Cox effect-size index. The Cox index is based on logged odds ratios; therefore, standard deviations are not needed for the calculation of effect sizes for binary outcome measures, such as whether youth have earned a high school diploma or GED.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No corrections for clustering or multiple comparisons were needed.

Appendix A4.2 Summary of under age 17 and age 17 and older subgroup findings for the completing school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study			WWC calculations		
			Mean outcome		Mean difference ² (ChalleNGe – comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
			ChalleNGe group	Comparison group				
Millenky, Bloom, & Dillon, 2010⁶								
Earned a Diploma or GED Certificate	Under age 17	435	0.58	0.27	0.30	0.78	Statistically significant	+28
Earned a Diploma or GED Certificate	Age 17 and older	761	0.62	0.40	0.22	0.53	Statistically significant	+20

1. This appendix presents under 17 and 17 and older subgroup findings for measures that fall in the completing school domain. Total group outcomes were used for rating purposes and are presented in Appendix A3.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
3. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of binary outcome variables, effect sizes are calculated using the Cox effect-size index. The Cox index is based on logged odds ratios; therefore, standard deviations are not needed for the calculation of effect sizes for binary outcome measures, such as whether youth have earned a high school diploma or GED.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No corrections for clustering or multiple comparisons were needed.

Appendix A4.3 Summary of low-achieving and higher-achieving subgroup findings for the completing school domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study			WWC calculations		
			Mean outcome		Mean difference ² (<i>ChalleNGe</i> – comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
			<i>ChalleNGe</i> group	Comparison group				
Millenky, Bloom, & Dillon, 2010⁶								
Earned a Diploma or GED Certificate	Higher-achieving	621	0.65	0.38	0.28	0.68	Statistically significant	+25
Earned a Diploma or GED Certificate	Low-achieving	540	0.55	0.33	0.22	0.55	Statistically significant	+21

1. This appendix presents low-achieving and higher-achieving subgroup findings for measures that fall in the completing school domain. Low-achieving is a binary indicator for students who earned mostly Ds and Fs when in school. Higher-achieving is a binary indicator for students who earned mostly better than Ds and Fs when in school. Total group outcomes were used for rating purposes and are presented in Appendix A3.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
3. For an explanation of the effect-size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of binary outcome variables, effect sizes are calculated using the Cox effect-size index. The Cox index is based on logged odds ratios; therefore, standard deviations are not needed for the calculation of effect sizes for binary outcome measures, such as whether youth have earned a high school diploma or GED.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
6. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No corrections for clustering or multiple comparisons were needed.

Appendix A5 National Guard Youth ChalleNGe Program rating for the completing school domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of completing school, the WWC rated the *National Guard Youth ChalleNGe Program* as having potentially positive effects for at-risk youth. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as the *National Guard Youth ChalleNGe Program* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. One study of the *National Guard Youth ChalleNGe Program* reported a statistically significant and substantively important positive effect in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies found statistically significant or substantively important negative effects or indeterminate effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

Not met. Only one study of the *National Guard Youth ChalleNGe Program* reported a statistically significant and substantively important positive effect in this domain.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. No studies found statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Staying in school	na	na	na	na
Progressing in school	na	na	na	na
Completing school	1	>2	1,196	Small

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.