Appendix

Appendix A1 Study characteristics: Snipes, Holton, Doolittle, & Sztejnberg, 2006—Houston study (quasi-experimental design)

Characteristic	Description
Study citation	Snipes, J. C., Holton, G. I., Doolittle, F., & Sztejnberg, L. (2006). *Striving for student success: The effect of Project GRAD on high school student outcomes in three urban school districts.* New York, NY: MDRC.
Participants	The main analysis sample included a series of cohorts of entering ninth grade students from three high schools implementing *Project GRAD* between 1998 and 2004 and ten matched comparison high schools. The sample consists of students for whom administrative records exist over the time period of the study.[1] A group of comparison schools was matched to each *Project GRAD* school based on performance on standardized achievement tests and demographic composition. The result was a sample of three *Project GRAD* and ten comparison high schools. The study followed cohorts of students. Cohort 1 included students in the intervention and matched comparison schools who enrolled in the ninth grade during the first year of *Project GRAD* implementation at the intervention schools. Similarly, Cohort 2 included students in the intervention and comparison schools who were enrolled in the ninth grade during the second year of implementation, Cohort 3 included students who enrolled during the third year, and so on. Given the fixed period for data collection, later cohorts had shorter follow-up periods. To ensure both an adequate follow-up and an adequate sample size for measuring impacts, the WWC used results based on either Cohorts 1 through 4 (for most outcomes) or Cohorts 1 through 3 (for ever graduated, looking ahead at least three years—the number of cohorts was limited by the definition of the outcome measure) to rate the effectiveness of *Project GRAD*. Results for later cohorts that were followed over a shorter follow-up period are reported in Appendix A4. On average, the three *Project GRAD* and ten comparison high schools served students who had similar test scores, similar attendance patterns, and similar rates of promotion. There were some differences between the schools, however. *Project GRAD* schools were smaller than comparison schools (1,333 versus 2,158 students on average). In addition, *Project GRAD* schools served a larger share of African-American students than comparison schools did (56% versus 44%) and a smaller share of white students (1% versus 10%). The proportion of students who were Hispanic was similar in *Project GRAD* and comparison schools.
Setting	The initiative originated at Jefferson Davis High School in Houston, Texas, with the implementation of the model's components in the 1994/95 school year. *Project GRAD* was implemented next at Jack Yates High School in the 1996/97 school year, and at Phillis Wheatley High School in the 1997/98 school year. The baseline period for Davis was the two years prior to 1994/95, while the baseline for the other schools was the three years prior to implementation. The analysis focuses on outcomes at these high schools through the 2003/04 school year.

(continued)

Appendix A1 Study characteristics: Snipes, Holton, Doolittle, & Sztejnberg, 2006—Houston study (quasi-experimental design) *(continued)*

Characteristic	Description
Intervention condition	*Project GRAD* targets a high school and the middle and elementary schools that feed into it. It combines a number of reforms with a goal of increasing reading and math achievement test scores, improving classroom behavior, providing a safety net for students to help reduce dropout rates, and increasing rates of high school graduation and college enrollment. At the high school level, *Project GRAD* has two main components: 1. *Project GRAD* college scholarships are provided to students who have a cumulative grade point average of at least 2.5, graduate within four years, complete a recommended college preparatory curriculum, and participate in two summer institutes. Scholarship amounts and criteria vary by site, averaging $1,000 to $1,500 a year. Each participating school has a scholarship coordinator who provides counseling, tutoring, and college admission preparation. 2. Summer institutes consist of four to six hours a day of instruction and related activities for four to six weeks in the summers. Parental and community improvement components seek to engage parents and the community in the schools and support students, along with social services and academic enrichment programs. Additionally, classroom management programs attempt to produce orderly classrooms focused on learning and promote positive relationships among students, teachers, and other adults. *Project GRAD* works with the entire feeder system of elementary and middle schools that send students to *Project GRAD* high schools to address early problems that can affect high school completion. To help students arrive at middle and high school better prepared academically, *Project GRAD* elementary schools provide professional development and coaches for teachers of reading and math and also implement curricula such as *MOVE IT Math™, Everyday Math™,* or *Success For All™*. To improve classroom behavior, *Project GRAD* schools implement Consistency Management & Cooperative Discipline®, an instructional discipline management system in which the teacher acts as an instructional leader and students have leadership roles. It is based on five elements: prevention of disruptive behavior through classroom management, a caring environment, cooperation, classroom organization, and parental and community involvement activities. *Project GRAD* also provides staff who deliver school-based social services—guidance, counseling, community outreach, and family case-management services—and facilitate parent involvement. Some sites link with Communities in Schools (CIS), a dropout prevention and social service agency, to provide social service and parent involvement staff members. In sites where there is no local CIS organization, *Project GRAD* has established a variation of the CIS component called Campus Family Support (CFS), which customizes traditional CIS services to meet the needs within the feeder system. In addition to student services, staff organize activities to enhance communication between teachers and parents.
Comparison condition	Matched comparison schools were Houston high schools that did not implement *Project GRAD*. Specifically, the analysis identified a set of comparison schools from the same district that were similar in terms of average performance on standardized achievement tests in the years immediately preceding program implementation and the percentages of students in key demographic groups.
Primary outcomes and measurement	Outcomes in two of the domains are included in this study. Two measures related to progressing in school were included: credits earned in 9th grade and promotion from 9th to 10th grade. One measure in the completing school domain was included: ever graduated, looking ahead at least three years. All measures are from administrative records. The study also examined *Project GRAD*'s effects on attendance and standardized test scores. These outcomes do not fall within the three domains (staying in school, progressing in school, completing school) examined by the WWC'S review of dropout prevention interventions and are not included in this report.
Staff training	Teachers at *Project GRAD* high schools were regular teachers employed by the Houston Independent School District. Information on staff training was not available.

1. The authors did not report the number of students in the sample.

Appendix A2.1 Outcome measures in the progressing in school domain

Outcome measure	Description
9th grade credits	This measure represents the cumulative total of all the credits that a student earned over the course of the first year of high school. These data were collected from individual students' school records obtained from the district.
9th grade promotion	This measure represents whether a student was in 10th grade by the end of the following year. These data were collected from individual students' school records obtained from the district.

Appendix A2.2 Outcome measure in the completing school domain

Outcome measure	Description
Ever graduated, looking ahead at least three years	This measure represents whether a student ever graduated from a school in the district, looking ahead at least three years. Specifically, it is measured for the 1997/98, 1998/99, and 1999/2000 cohorts of 9th graders, using graduation data from the 2000/01, 2001/02, and 2002/03 school years. These data were collected from individual students' school records obtained from the district.

Appendix A3.1 Summary of study findings included in the rating for the progressing in school domain[1]

			Authors' findings from the study					
			Mean outcome		WWC calculations			
Outcome measure	Study sample	Sample size (schools)[2]	*Project GRAD* group	Comparison group[3]	Mean difference[4]	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Snipes, Holton, Doolittle, & Sztejnberg, 2006—Houston study (quasi-experimental design)[8]								
9th grade credits	Cohorts 1–4	13	2.4	2.7	–0.3	–0.17	ns	–7
9th grade promotion (%)	Cohorts 1–4	13	44.9	46.9	–2.0	–0.05	ns	–2
Domain average for progressing in school[9]						–0.11	ns	–4

ns = not statistically significant

1. This appendix reports findings considered for the rating of effectiveness and the improvement index. These results were measured at the end of the fourth year of the intervention, when only data for cohorts 1 through 4 were available. Findings based on later cohorts were not included in these ratings, but are reported in Appendix A4.
2. Snipes, Holton, Doolittle, & Sztejnberg (2006) used individual student data, but did not report the number of students in the sample. In the study, each block of schools consists of a *Project GRAD* school matched with a group of between two and four comparison schools; there were three intervention and 10 comparison schools in total.
3. Snipes, Holton, Doolittle, & Sztejnberg (2006) reported baseline to follow-up changes for both the intervention and comparison groups. The WWC generated the adjusted comparison group means reported here using the following transformation: adjusted comparison group mean = follow-up comparison group mean + (baseline intervention group mean – baseline comparison group). Stated differently, the adjusted comparison group mean equals the follow-up intervention group mean minus the impact, since, under the comparative interrupted time-series technique using in the Snipes, Holton, Doolittle, & Sztejnberg (2006) study, impacts are calculated as follows: impact = (follow-up intervention group mean – baseline intervention group mean) – (follow-up comparison group mean – baseline comparison group mean).
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of effect size calculation, please see the Technical Details of WWC-Conducted Computations. For the credits earned measure, the student-level standard deviation of the outcome was not available for the intervention and comparison samples; however, the authors provided the WWC with the standard deviation of the measure for the entire school district (1.73), which was used for both groups in the effect size calculation. The effect size for the dichotomous variable "9th grade promotion" was computed using the Cox Index.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank 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.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of the Houston *Project GRAD* study, the study authors provided upon the WWC request details of their two-level analysis model, which adjusted for clustering within the school, and thus no additional corrections for clustering were necessary.
9. This row provides the study average, which in this instance, 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 A3.2 Summary of study findings included in the rating for the completing school domain[1]

			Authors' findings from the study					
			Mean outcome		WWC calculations			
Outcome measure	Study sample	Sample size (schools)[2]	*Project GRAD* group	Comparison group[3]	Mean difference[4]	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Snipes, Holton, Doolittle, & Sztejnberg, 2006—Houston study (quasi-experimental design)[8]								
Ever graduated, looking ahead at least three years	Cohorts 1–3	13	32.0	34.6	–2.5	–0.07	ns	–3
Domain average for completing school[9]						–0.07	ns	–3

ns = not statistically significant

1. This appendix reports overall findings that are considered for the rating of effectiveness and the improvement index.
2. Snipes, Holton, Doolittle, & Sztejnberg (2006) used individual student data, but did not report the number of students in the sample. In the study, each block of schools consists of a *Project GRAD* school matched with a group of between two and four comparison schools; there were three intervention and 10 comparison schools in total.
3. Snipes, Holton, Doolittle, & Sztejnberg (2006) reported baseline to follow-up changes for both the intervention and comparison groups. The WWC generated the adjusted comparison group means reported here using the following transformation: adjusted comparison group mean = follow-up comparison group mean + (baseline intervention group mean – baseline comparison group). Stated differently, the adjusted comparison group mean equals the follow-up intervention group mean minus the impact, since, under the comparative interrupted time-series technique using in the Snipes, Holton, Doolittle, & Sztejnberg (2006) study, impacts are calculated as follows: impact = (follow-up intervention group mean – baseline intervention group mean) – (follow-up comparison group mean – baseline comparison group mean).
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of effect size calculation, please see the Technical Details of WWC-Conducted Computations. The effect size for the dichotomous variable "ever graduated, looking ahead at least three years" was computed using the Cox Index.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank 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.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of the Houston *Project GRAD* study, the study authors provided upon the WWC request details of their two-level analysis model, which adjusted for clustering within the school, and thus no additional corrections for clustering were necessary.
9. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The improvement index for the domain is calculated from the domain average effect size.

Appendix A4 Summary of longer-term findings for the progressing in school domain[1]

			Authors' findings from the study					
			Mean outcome		WWC calculations			
Outcome measure	Study sample	Sample size (schools)[2]	*Project GRAD* group	Comparison group[3]	Mean difference[4]	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Snipes, Holton, Doolittle, & Sztejnberg, 2006—Houston study (quasi-experimental design)[8]								
9th grade credits	Cohorts 5–7	13	3.0	3.1	–0.1	–0.06	ns	–2
9th grade promotion (%)	Cohorts 5–6	13	50.7	46.1	4.6	0.11	ns	+4

ns = not statistically significant

1. This appendix presents findings for later cohorts of 9th graders for measures that fall in the progressing in school domain. The WWC rated the effectiveness of *Project GRAD* in the progressing in school domain based on results for cohorts 1 through 4. These results are presented in Appendix A3.1.
2. Snipes, Holton, Doolittle, & Sztejnberg (2006) used individual student data, but did not report the number of students in the sample. In the study, each block of schools consists of a *Project GRAD* school matched with a group of between two and four comparison schools; there were three intervention and 10 comparison schools in total.
3. Snipes, Holton, Doolittle, & Sztejnberg (2006) reported baseline to follow-up changes for both the intervention and comparison groups. The WWC generated the adjusted comparison group means reported here using the following transformation: adjusted comparison group mean = follow-up comparison group mean + (baseline intervention group mean – baseline comparison group). Stated differently, the adjusted comparison group mean equals the follow-up intervention group mean minus the impact, since, under the comparative interrupted time-series technique using in the Snipes, Holton, Doolittle, & Sztejnberg (2006) study, impacts are calculated as follows: impact = (follow-up intervention group mean – baseline intervention group mean) – (follow-up comparison group mean – baseline comparison group mean).
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of effect size calculation, please see the Technical Details of WWC-Conducted Computations. For the credits earned measure, the standard deviation of the outcome was not available for the intervention and comparison samples; however, the authors provided the WWC with the standard deviation of the measure for the entire school district (1.73), which was used for both groups in the effect size calculation. The effect size for the dichotomous variable "9th grade promotion" was computed using the Cox Index.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank 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.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of the Houston *Project GRAD* study, the study authors provided upon the WWC request details of their two-level analysis model, which adjusted for clustering within the school, and thus no additional corrections for clustering were necessary.

Appendix A5.1 *Project GRAD* rating for the progressing in school domain

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

For the outcome domain of progressing in school, the WWC rated *Project GRAD* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because it only had one study, and that study showed no statistically significant or substantively important outcomes, either positive or negative, in this domain.

Rating received
No discernible effects: No affirmative evidence of effects.
• Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*
Met. The single study of *Project GRAD* showed no statistically significant or substantively important 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. No studies of *Project GRAD* showed a statistically significant positive effect in this domain.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. No studies of *Project GRAD* showed statistically significant or substantively important negative effects in this domain.
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.
Not met. No studies of *Project GRAD* showed a statistically significant or 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.
Not met. No studies of *Project GRAD* showed statistically significant or substantively important negative effects in this domain, and one study showed an indeterminate effect in this domain.

(continued)

Appendix A5.1 *Project GRAD* rating for the progressing in school domain *(continued)*

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No studies of *Project GRAD* showed a statistically significant or substantively important effect, either positive or negative, in this domain.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No studies of *Project GRAD* showed a statistically significant or substantively important effect in this domain.

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

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

  Not met. No studies of *Project GRAD* showed a statistically significant or substantively important negative effect in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies of *Project GRAD* showed statistically significant or substantively important positive effects in this domain.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

  Not met. No studies of *Project GRAD* showed a statistically significant negative effect in this domain.

AND

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

  Met. No studies of *Project GRAD* showed statistically significant or substantively important positive 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. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Project GRAD* rating for the completing school domain

The WWC rates an intervention's effects in 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 *Project GRAD* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because it only had one study, and that study showed no statistically significant or substantively important outcomes, either positive or negative, in this domain.

Rating received
No discernible effects: No affirmative evidence of effects.
• Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*
Met. The single study of *Project GRAD* showed no statistically significant or substantively important 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. No studies of *Project GRAD* showed a statistically significant positive effect in this domain.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. No studies of *Project GRAD* showed statistically significant or substantively important negative effects in this domain.
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.
Not met. No studies of *Project GRAD* showed a statistically significant or 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.
Not met. No studies of *Project GRAD* showed statistically significant or substantively important negative effects in this domain, and one study showed an indeterminate effect in this domain.

(continued)

Appendix A5.2 *Project GRAD* rating for the completing school domain *(continued)*

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No studies of *Project GRAD* showed a statistically significant or substantively important effect, either positive or negative, in this domain.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No studies of *Project GRAD* showed a statistically significant or substantively important effect in this domain.

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

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

  Not met. No studies of *Project GRAD* showed a statistically significant or substantively important negative effect in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies of *Project GRAD* showed statistically significant or substantively important positive effects in this domain.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

  Not met. No studies of *Project GRAD* showed a statistically significant negative effect in this domain.

AND

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

  Met. No studies of *Project GRAD* showed statistically significant or substantively important positive 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. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Staying in school	0	0	0	na
Progressing in school	1	13	nr	Small
Completing school	1	13	nr	Small

na = not applicable/not studied
nr = not reported

1. A rating of "moderate 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."