



**HOUSTON EDUCATION RESEARCH CONSORTIUM
KINDER INSTITUTE FOR URBAN RESEARCH
RICE UNIVERSITY**

**Analytic Summary of Existing
Project GRAD Evaluations**

UPDATED

March 16, 2012

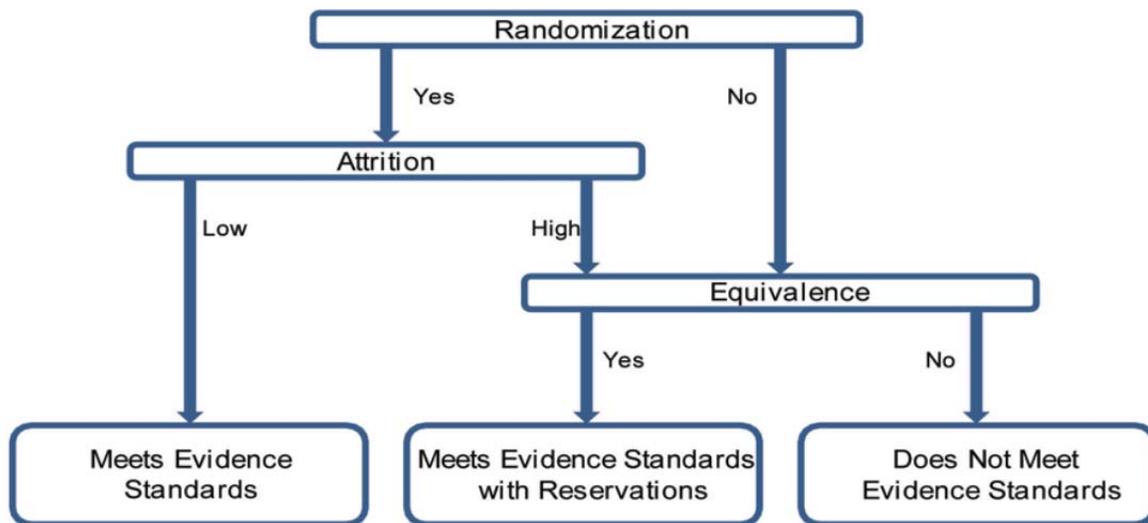
Summary

Project GRAD (PG) is a non-profit education reform model with the goal of increasing high school graduation and college attendance rates for low-income students. PG works in the feeder patterns for five disadvantaged high schools in Houston: Davis, Sam Houston, Reagan, Wheatley, and Yates. Through the supplemental GEAR UP (GU) program, PG also worked to boost student achievement in the 11 middle schools in the PG feeder patterns. This report is written to provide information regarding the quality and strength of the existing evidence related to the effectiveness of PG in raising middle school student achievement, and high school student graduation and college readiness. This report uses the *What Works Clearinghouse* (2010) standards for research to evaluate the quality of the evidence generated to date on the effectiveness of PG. A comprehensive analysis of each PG evaluation is listed in Table 1.

Only one study, released in 2006 and based on data as recent as 2004, met evidence standards with reservations. This study reported **no discernable effects** of PG on indicators of high school graduation and college readiness. Available studies on middle school outcomes did not meet evidence standards.

Method

For this report, HERC reviewed evaluations likely to be most relevant to the current PG program, which focuses on graduation and college readiness in high schools, and on academic achievement in 11 PG middle schools. We identified six independent evaluations that evaluated these outcomes.¹ Based on the *What Works Clearinghouse* guidelines, we identified each study as “Meets Evidence Standards,” “Meets Evidence Standards with Reservations,” or “Does Not Meet Evidence Standards.” Criteria for each designation are described below.



Meets Evidence Standards: An evaluation that meets evidence standards includes randomization between treatment and control groups and low attrition.

¹ HERC also found independent evaluations of test achievement in elementary schools (Opuni 1999, 2002). Since the current focus of Project GRAD is on high school students and middle-school students (through GEAR UP), HERC chose not to review detailed results for elementary schools in this report. A review of these results is available upon request; however, the overall quality of the evidence from these evaluations is similar to most of the evaluations: Does Not Meet Evidence Standards.

Meets Evidence Standards with Reservations: An evaluation that meets evidence standards with reservations either includes randomization, but with high attrition, or does not include randomization but equivalence between treatment and comparison groups is adequately established.

Does Not Meet Evidence Standards: An evaluation that does not meet evidence standards lacks both randomization and equivalence. Furthermore, evaluations characterized by randomization and high attrition (particularly with high differential attrition) that fail to establish and adequately account for non-equivalence do not meet evidence standards.

As is stated by the *What Works Clearinghouse*, decisions about education programs, policies, and practices should be based on relevant studies that reach sound conclusions about the effectiveness of particular interventions. When studies about a particular intervention—such as Project GRAD—meet evidence standards (with or without reservations), educators can exercise confidence in conclusions about the effectiveness of the intervention.

Of the six existing studies that evaluate the influence of Project GRAD, one study (Snipes, et al. 2006) met the evidence standards with reservations (see Table 1). HISD can have confidence in the findings associated with this study. The other studies we evaluated did not meet evidence standards. In an effort to provide HISD with thorough information about each study, we discuss results from studies that did not meet evidence standards; however, we caution that one cannot draw reliable conclusions from these results. A detailed explanation of the weaknesses in each of these studies is listed in Appendix A.

Existing Evidence on Project GRAD

Based on the one study that met the evidence standards with reservations (Snipes et al. 2006), PG is associated with the following outcomes:

- No significant or meaningful gains among PG high school students relative to comparison students in non-PG high schools in HISD, over four outcomes: high school graduation, graduating high school with a core academic curriculum (GPA above 75 in math, English, social studies, and science, and accumulating three math credits, four English credits, two social studies credits, and two science credits), 9th grade credits, 9th grade promotion to 10th grade.
- One exception was found in Davis high school: students at Davis were significantly more likely to graduate with the core academic curriculum than comparison students. Since Davis has the longest-running PG program, it is possible that successful implementation of the program takes several years. However, the lack of significant effects at Davis in other measured outcomes (overall high school graduation, 9th grade credits and promotion) weakens this hypothesis.
- The authors suggested two explanations for the lack of positive results: (1) the limited exposure to PG curriculum among students before reaching PG high schools, which may be related to duration of program implementation, “leaky” feeder patterns, and high student mobility; and (2) PG’s lack of direct involvement in the 9th grade curriculum, which might keep students on-track after the first year of high school and on a better path toward graduation.

While only one study met research standards with reservations, another study could potentially be considered credible, with additional information. Chan et al. (2006) was a preliminary study that only provided bar charts showing gains in graduation rates in three PG high schools: Davis, Yates, and Wheatley, relative to comparison schools. Without additional information on the comparison group match and statistical analyses, this study could not meet evidence standards. However, the charts do indicate that, after four years of implementation at each PG school, these schools do show significantly greater

gains in graduation rates than comparison schools. In addition, two studies by Opuni (1999, 2009) used minimally adequate research designs, but inappropriate statistical methods. These studies could meet standards with reservations with more rigorous and appropriate statistical tests.

Response from Project GRAD

HERC had several conversations with PG staff, in which they raised some concerns about the Snipes, et al. (2006) study. Their major concern was that this study was too dated to reflect substantive improvements made to their program in the years since, and that this review did not reflect PG as it currently stands. These changes were made as a consequence of adoption by HISD of their most successful interventions throughout the district, responding to emerging challenges, and cuts in funding. In particular, PG reports that it has changed its program in the following ways:

- In the elementary and middle schools, the PG program has shifted from a focus on reading and math curriculum, academics, and classroom management, to more direct intervention with parents, concentrating on college awareness and planning.
- PG has also expanded the fine arts program in the Davis feeder pattern, which provides direct arts instruction, teacher professional development, and after school and summer programming. Currently, this program includes five full-time and three part-time staff.
- Since 2006, PG has shifted more of its focus and resources to the high school programs, which until then operated with about two staff members at each high school, and two coordinators in the PG home office. Currently, the high school teams consist of 3-5 staff members.
- PG also reports that high school staff have provided direct services to 83% of juniors and seniors during the 2011-12 school year. HERC has no comparative information on the percentage of high school students receiving services in previous years.
- PG has also expanded access to the College Institutes, which are intended to provide academic instruction and added a focus on career awareness and college knowledge. Yearly student attendance increased from approximately 700 students before 2003, to about 1600 students currently. They have also expanded the college partners beyond University of Houston-Downtown to add five additional local colleges and universities.
- PG also increased support for the transition to college and college retention, with scholarship application support and on-campus mentoring at many Texas universities.

It is beyond the scope of this review to verify these accounts or to evaluate their effectiveness. However, the relevance of any program changes are dependent on how well they are aligned with the recommendations of previous evaluations (see Snipes et al., 2006), as well as with measurable outcomes. While PG highlights changes made to the high school programs since 2004, based on information they have reported, these are better described as expansions of existing services, rather than substantive modifications to the overall program. Existing studies do not examine effectiveness of these particular services on measured outcomes; as a result, we have no evidence that expanding these services will significantly change the effectiveness of PG. However, it seems clear that these program changes do not specifically address the recommendations made by Snipes, et al.: particularly, a specific intervention in 9th grade instruction, in order to improve the number of students who are on-track to graduate after their freshman year. Implementation of program elements is another important consideration, particularly specific dosage indicators (e.g., number of contacts with each student, types of contacts). These types of implementation data have not been made available to HISD in the past, nor have they been examined in a published, research-based evaluation that connects implementation to outcomes. Analysis of such data would greatly improve the extent and quality of evidence used to assess the PG program.

Second, PG staff were concerned with the method used to calculate graduation rates. Snipes et al. used the percentage of first-time 9th grade students who graduate on-time from any HISD school, not just the

school of origin. This method masks any transfers in or out of schools before 12th grade. It is unclear whether this method biases graduation rates in ways that may advantage or disadvantage PG schools, but the amount of bias is likely to be quite small. It should be noted that Snipes et al. also found no positive effects of PG on early high school outcomes (such as 9th grade promotion) that are not affected by transfers.

Third, PG staff were concerned that Snipes et al. (2006) were unable, due to data limitations, to use a measure of economic disadvantage to help select the comparison schools, suggesting that this led to a less-than-equivalent match. However, matching on economic disadvantage likely would result in a more economically advantaged set of comparison schools; as a consequence, the authors suggest that “if there is a bias, it is in favor of finding effects for Project GRAD” (footnote 8, p. 32).

Finally, PG had some concerns with the quality of the data that Snipes et al. (2006) used for their study, which was obtained from HISD. PG staff felt that more valid data come directly from the Texas Education Agency, which they use for their own analyses. However, HISD reports that TEA data come directly from HISD, so they are effectively the same data. While HERC cannot verify the specifics of how Snipes et al. (2006) created their measures, there is no evidence that their data are any different than what would have been obtained from TEA.

Lessons from Project GRAD Evaluations

One reason why PG does not seem to have any discernable effect on high school outcomes, as evidenced by Snipes et al. (2006), could be that it may not have a substantive impact on sufficient numbers of students within PG schools. PG could not provide specific details on the number of 9th grade students who receive direct program interventions, relative to the number of students who complete the program (i.e., qualify for the scholarship) and who go on to college. However one can calculate a rough estimate based on evidence from Stiles (2011)², which suggests that 66% of 9th grade students at PG schools graduate from high school, of whom 45% qualify for the scholarship, of whom 70% enroll in college within one year of graduation. HERC therefore estimates that approximately 21% of students who start in PG high schools are likely to graduate and attend college. While PG may make a significant and substantial impact on this 21%, it is possible that this subgroup is selective of the strongest students who are most likely to attend college, with or without assistance.

PG faces several challenges that restrict its ability to successfully impact more students. The “leaky” HISD feeder school system and high student mobility across schools limit the number of students who have had exposure to PG curriculum and services prior to high school. Snipes (2006) found that, after 10 years of program implementation, 9th graders at Davis high school reported an average of 4.5 years of exposure to PG. Also, magnet programs in PG high schools (such as Davis) may attract students from non-PG feeder patterns; these students are selective of those who are most likely to attend college, even without PG, and yet will have the lowest exposure to PG programs before high school. In addition, the substantial program requirements may necessarily limit the number of students who eventually complete the program, which may also limit its overall impact.

Detailed information on implementation of the core aspects of the PG program, and recruitment and retention of students, would allow for a better picture of the impact of PG on students. Key questions include: which aspects of the PG program are most likely to keep students active and engaged throughout

² Stiles (2011) was a study of the likelihood of enrolling in college, among the PG students who successfully completed the high school program. Since it was not a formal evaluation of the PG program (that is, it did not examine outcomes for all students in PG schools and did not include a comparison group), it was not reviewed for this study. However, Stiles (2011) did provide useful background information on the program.

high school? Which aspects of the program are most likely to predict which students graduate from high school and go on to college? These questions can only be answered with detailed, student-level data on all entering freshmen at PG schools, as well as program information on the number of students who are contacted, who participate in specific portions of the program, and who graduate and go on to college.

HERC Recommendations for Future Research

One conclusion HERC has drawn from this review process is that HISD has access to a great deal of research evaluating their programs, including both internal reports produced by HISD research staff and external reports produced by researchers outside of HISD. However, this research varies widely in quality, and thus in its ability to predict the effectiveness of the program. This variation is particularly consequential because these research reports are used to assist HISD leaders in making decisions regarding funding and program continuation. High-quality, unbiased research is most likely to provide an accurate representation of a program's effectiveness, uncover areas for improvement, and guide HISD leaders when setting budget priorities. In contrast, lower-quality research often leads to incorrect conclusions about the effectiveness of a program. As a consequence, all reports used for decision-making should be subject to a review process.

Conclusion

Based on the one 2006 study that meets evidence standards with reservations, we find no evidence that PG has increased high school student attainment or graduation, relative to students from non-PG schools. It is possible that additional rigorous studies will find evidence of PG's impact, or that a more detailed study linking implementation to student outcomes will reach different conclusions. Given the popularity of PG among successful completers of the program and their parents, it is likely that PG is successful in helping this subgroup of all students in PG schools. However, if one evaluates PG as a school-wide intervention intended to "saturate" the entire school with its program, we find no evidence of a school-level positive effect on boosting the graduation and college attendance rates of disadvantaged students.

TABLE 1. SUMMARY OF EVALUATIONS OF PROJECT GRAD PROGRAM.

Study	Source of Evaluation	Outcome Studied	Quality of Evidence			Effectiveness			Applicability to HISD Pop
			Overall Rating	Strength of Research Design	Analytic Method	Impact ^a & Statistical Significance ^b	Effect Size ^c & Substantive Significance ^d	Effect on Achievement Gaps	
Chan et al. (2006) [AIR] ^e	Independent Third Party	Graduation Rates	Does Not Meet Evidence Standards	Meets Standards with Reservations	Does Not Meet Evidence Standards	Significant gains in Davis, Yates, and Wheatley schools ^e	Not Rated	N/A	HISD Population
Opuni (1999)	Independent Third Party	8 th G Stanford NCE Scores, Yates Feeder	Does Not Meet Evidence Standards	Meets Standards with Reservations	Does Not Meet Evidence Standards	Significant gains in reading and math scores. ^f	Math- .34** Reading- .19	N/A	HISD Population
Opuni (2002)	Independent Third Party	College Attendance Rate	Does Not Meet Evidence Standards	Does Not Meet Evidence Standards	Does Not Meet Evidence Standards	Not Rated	Not Rated	N/A	HISD Population
Opuni (2007)	Independent Third Party	7 th G Stanford NCE scores	Does Not Meet Evidence Standards	Does Not Meet Evidence Standards	Does Not Meet Evidence Standards	Reading: -.91* Math: -2.03*	Reading: .04 Math: .09	N/A	HISD Population
	Independent Third Party	6 th G Stanford NCE scores	Does Not Meet Evidence Standards	Does Not Meet Evidence Standards	Does Not Meet Evidence Standards	Reading: -1.22 Math: .66*	Reading: -.03 Math: .05	N/A	HISD Population
Opuni (2009)	Independent Third Party	9 th G TAKS scale scores	Does Not Meet Evidence Standards	Meets Standards with Reservations	Does Not Meet Evidence Standards	Reading: 3* Math: -55*	Reading: .14 Math: .17	N/A	HISD: Sub-Group of At-Risk Students
	Independent Third Party	9 th G Stanford NCE scores	Does Not Meet Evidence Standards	Meets Standards with Reservations	Does Not Meet Evidence Standards	Reading: 1.13* Math: -2.2* Science: .4* Soc. Studies: -3.72	Reading: .13 Math: .16 Science: .29 Soc. Studies: .05	N/A	HISD: Sub-Group of At-Risk Students

Study	Source of Evaluation	Outcome Studied	Quality of Evidence			Effectiveness		Applicability to HISD Pop	
			Overall Rating	Strength of Research Design	Analytic Method	Impact ^a & Statistical Significance ^b	Effect Size ^c & Substantive Significance ^d		Effect on Achievement Gaps
	Independent Third Party	8 th G TAKS scale scores	Does Not Meet Evidence Standards	Meets Standards with Reservations	Does Not Meet Evidence Standards	Reading: -.3* Math: -.63	Reading: .14 Math: -.04	N/A	HISD: Sub-Group of At-Risk Students
	Independent Third Party	8 th G Stanford NCE scores	Does Not Meet Evidence Standards	Meets Standards with Reservations	Does Not Meet Evidence Standards	Reading: -1.74 Math: -3.65 Science: -4.74* Soc. Stud.: 2.03*	Reading: -.02 Math: -.07 Science: -.17 Soc. Studies: .18	N/A	HISD: Sub-Group of At-Risk Students
Snipes et al. (2006) [MDRC]	Independent Third Party	Graduation rate	Meets Standards with Reservations	Meets Standards with Reservations	Meets Standards with Reservations	Davis: .5 All Schools: -1.8	Davis: Not Rated All Schools: -.07	N/A	HISD Population
	Independent Third Party	Graduation with core curriculum	Meets Standards with Reservations	Meets Standards with Reservations	Meets Standards with Reservations	Davis: 6.6* All Schools: 1.7	Not Rated	N/A	HISD Population
	Independent Third Party	9 th G credits	Meets Standards with Reservations	Meets Standards with Reservations	Meets Standards with Reservations	Davis: -.05 All Schools: -.3	Davis: Not Rated All Schools: -.17	N/A	HISD Population
	Independent Third Party	9 th G promotion	Meets Standards with Reservations	Meets Standards with Reservations	Meets Standards with Reservations	Davis: -.7 All Schools: -2.0	Davis: Not Rated All Schools: -.05	N/A	HISD Population

^a Calculated by HERC (for Snipes [2006], calculated by authors). Not calculated for Chan (2006), see below. Impact= (follow-up intervention group mean – baseline intervention group mean) – (follow-up comparison group mean – baseline comparison group mean), following standards of What Works Clearinghouse (2011).

^b Calculated by study author.

^c Calculated by study author (for Snipes [2006], calculated by What Works Clearinghouse, and not calculated for Davis feeder pattern).

^d Assessed by HERC. Effect sizes that are substantively significant (large enough to be educationally meaningful) are indicated by ** (see below).

^e Chan (2006) did not provide specific group means, so impacts and effect sizes were not calculated. Charts show that gains in graduation rates were significantly greater than comparison schools for cohorts in all PG schools: Davis (1998-99, 1999-00, 2000-01, 2001-02), Yates (1999-00, 2000-01, 2001-02), and Wheatley (2001-02).

^f Could not calculate impact, because study used different pre-test (1995-96 TAAS) and post-test (1998-99 Stanford).

*Significant $p < .05$. **Substantively significant effect size $\geq .33$ (Cohen 1988).

APPENDIX A: METHODOLOGICAL SUMMARY OF PROJECT GRAD EVALUATIONS THAT DO NOT MEET EVIDENCE STANDARDS

Study

Chan et al. (2006) [AIR]³

Outcome

Graduation Rate

Randomly Assigned

No.

Equivalence

Adequate Baseline Data: Yes.

Adequate Control Variables: No information.

Appropriate Matching: No information.

Attrition

Overall Attrition: No information.

Evenness between groups: No information.

Generalizability

Treated sample size & generalizable population: Not specified.

Representative of HISD population: Yes.

Analytic Method

No information on method. If study provided more detail on research design, comparison group matching, and analytic method, Overall Rating could be improved to Meets Standards with Reservations.

Study

Opuni (1999)

Outcome

8th Grade Stanford NCE Scores

Randomly Assigned

No.

Equivalence

Adequate Baseline Data: No- used 6th Grade TAAS to predict 8th Grade Stanford NCE.

Adequate Control Variables: None.

Appropriate Matching: Minimal- PG schools matched with other HISD schools with similar student ethnicity, gender, test scores, promotion rates, and teacher characteristics.

Attrition

Overall Attrition: Not specified.

Evenness between groups: Not specified.

Generalizability

Treated sample size & generalizable population: 250+ students, HISD middle schools.

Representative of HISD population: Yes

Analytic Method

No, used ANCOVA to adjust for pre-test score differences. This method does not correct for correlations between pre-test scores and grouping variable. This study would be improved with matched comparison schools, regression models to confirm match, and control for student characteristics.

³ Chan (2006) did not provide specific group means, so impacts and effect sizes were not calculated. Charts show that gains in graduation rates were significantly greater than comparison schools for cohorts in all PG schools: Davis (1998-99, 1999-00, 2000-01, 2001-02), Yates (1999-00, 2000-01, 2001-02), and Wheatley (2001-02).

Study

Opuni (2002)

Outcome

College Attendance Rate

Randomly Assigned

No.

Equivalence

Adequate Baseline Data: No, only one year pre-Project GRAD.

Adequate Control Variables: No.

Appropriate Matching: No comparison group.

Attrition

Overall Attrition: No information.

Evenness between groups: No comparison group.

Generalizability

Treated sample size & generalizable population: N/A- no comparison group, school-level analyses.

Representative of HISD population: Yes.

Analytic Method

Not appropriate. Descriptive trends, but no statistical analysis of size of change in college attendance rates over time, no significance tests of size of effect. Would be improved with matched comparison groups, sufficient data on pre-and post-test outcomes, and regression models to confirm match and control for student characteristics.

Study

Opuni (2007)

Outcome

6th & 7th Grade Stanford NCE Scores

Randomly Assigned

No.

Equivalence

Adequate Baseline Data: Yes.

Adequate Control Variables: No.

Appropriate Matching: No, matched to students with similar ethnicity, gender, and test scores in cohorts in same school in pre-PG years. Better to match to comparable students in comparable schools in same year, to account for general increase in test scores.

Attrition

Overall Attrition: No information.

Evenness between groups: No information.

Generalizability

Treated sample size & generalizable population: 2800 students in HISD middle schools.

Representative of HISD population: Yes.

Analytic Method

Not appropriate, used ANCOVA to adjust for pre-test score differences. This method does not correct for correlations between pre-test scores and grouping variable. Would be improved with matched comparison schools, and regression models to confirm match and control for student characteristics.

Study

Opuni (2009)

Outcome

8th & 9th Grade TAKS Scale and Stanford NCE Scores

Randomly Assigned

No.

Equivalence

Adequate Baseline Data: No, new norms for Stanford NCE in follow-up year (08-09).

Adequate Control Variables: None.

Appropriate Matching: Minimal- PG students were matched with students at non-PG schools with similar ethnicity, gender, absenteeism, disciplinary rates, and test scores, but limited by one-to-one matching; better to use multiple matches per PG student. Optimal method would be propensity score matching with multiple matches.

Attrition

Overall Attrition: 8% (minimal).

Evenness between groups: Even (PG 3%, non-PG 1.3%).

Generalizability

Treated sample size & generalizable population: 1100 at-risk middle school students (20% of all PG middle school students)

Representative of HISD population: No.

Analytic Method

Not appropriate. Used ANCOVA to adjust for pre-test score differences. This method does not correct for correlations between pre-test scores and grouping variable. Would be improved with matched comparison schools, and regression models to confirm match and control for student characteristics.

REFERENCES

- Chan, Tsze, Carlos Rodriguez, Charles Story, and Erin Hoppe. 2006. *High School Graduation Rates: Preliminary Results. A Pilot Study of Project GRAD in Houston ISD*. Washington, DC: American Institutes for Research.
- Opuni, Kwame. 1999. *Project GRAD: Graduation Really Achieves Dreams. Program Evaluation Report*. Houston, TX: Houston Independent School District.
- Opuni, Kwame and Mary L. Ochoa. 2002. *Project GRAD: A Comprehensive School Reform Model. Project GRAD-Houston: 2001-2002 Program Evaluation Report*. Center for Research on School Reform. Houston, TX: University of St. Thomas.
- Opuni, Kwame. 2007. *College for All (CFA): GEAR UP Program. First Biennial Evaluation Report*. Houston, TX: University of Houston-Downtown.
- Opuni, Kwame. 2009. *College for All (CFA): GEAR UP Program. Second Biennial Evaluation Report*. Houston, TX: University of Houston-Downtown.
- Snipes, Jason C., Glee Ivory Holton, Fred Doolittle, Laura Szejnberg. 2006. *Striving for Student Success: The Effect of Project GRAD on High School Student Outcomes in Three Urban School Districts*. New York, NY: MDRC.
- Stiles, Ann B. 2011. *College-Bound or College-Abandoned: Studies Stories from the Place Between*. Doctoral Thesis, College of Education. Houston, TX: University of Houston.
- What Works Clearinghouse. 2011. *Procedures and Standards Handbook. Version 2.1*. Washington, DC: Institute for Education Sciences.