Washington State Institute for Public Policy

110 Fifth Avenue SE, Suite 214 • PO Box 40999 • Olympia, WA 98504 • 360.586.2677 • www.wsipp.wa.gov

July 2016

Graduation, Reality, and Dual-role Skills (GRADS) Program for Pregnant and Parenting Teens: *Outcome Evaluation and Benefit-Cost Analysis*

Summary

Graduation, Reality, and Dual-role Skills (GRADS) is a program for pregnant and parenting teens in grades 9-12 in some public schools in Washington State. The focus of GRADS is to help students take on the "dual role" of student and parent, by providing classes related to employment and parenting skills and child care on-site or within walking distance.

In 2014, the Washington State Department of Health and Office of Superintendent of Public Instruction requested that the Washington State Institute for Public Policy conduct an outcome evaluation and benefit-cost analysis of the GRADS program.

In this evaluation, we compare teen mothers that participated in GRADS to a group of similar teen mothers from districts that did not offer the program.

Based on the results of our analysis, we estimate that GRADS participants have a 10.6 percentage point higher rate of high school graduation by age 22 and a 6.5 percentage point higher rate of postsecondary course enrollment by age 24.

Our benefit-cost analysis indicates that the benefits of GRADS outweigh its costs. The net per-student cost to provide GRADS is about $7,588. The per-student benefits total $22,839, for a benefit-cost ratio of $3 to $1. We tested the uncertainty in our estimate and find that benefits outweigh costs 93% of the time.

Suggested citation: Lemon, M., Fumia, D., & He, L. (2016). *Graduation, Reality, and Dual-role Skills (GRADS) program for pregnant and parenting teens: Outcome evaluation and benefit-cost analysis* (Document No. 16-07-2202). Olympia: Washington State Institute for Public Policy.

Graduation, Reality, and Dual-role Skills (GRADS)[1] is a program for pregnant and parenting teens in grades 9-12 in some public schools in Washington State. The focus of GRADS is to help students take on the "dual role" of student and parent and prepare them for the world of work. Participants take classes related to employment and parenting skills and receive child care on-site or within walking distance. The program is funded in part by the state's per-student Career and Technical Education allocation[2] and the Working Connections Child Care subsidy.[3]

In 2014, the Washington State Department of Health (DOH) and Office of Superintendent of Public Instruction (OSPI) received a federal grant to address teen pregnancy and requested that the Washington State Institute for Public Policy (WSIPP) conduct an outcome evaluation and benefit-cost analysis of the GRADS program with support from this funding source. WSIPP's Board of Directors approved the project in December 2014.

[1] Information on the GRADS program can be found at http://www.k12.wa.us/CareerTechEd/GRADSProgram.aspx.

[2] In school year 2013–14, state funding averaged $6,043 per CTE full-time equivalent (FTE) student compared to $5,297 per basic education FTE, a difference of $746 or 14%. http://www.k12.wa.us/LegisGov/2014documents/CTESkillCenterFunding.pdf

[3] Subsidies vary depending on household income and child's age. http://www.dshs.wa.gov/onlinecso/wccc.shtml

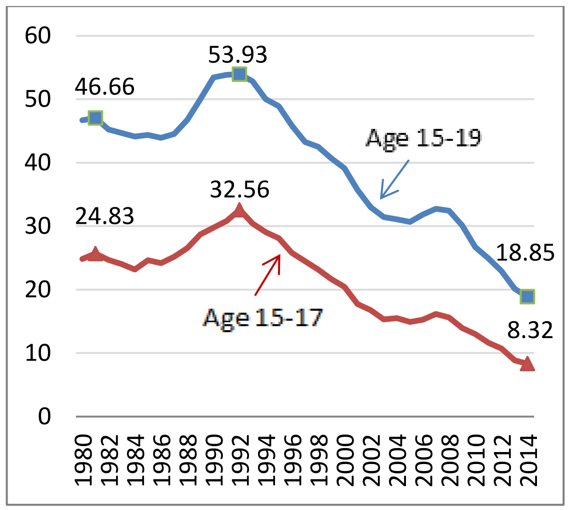
The report is organized as follows. Section I provides background on teen parents in Washington State and the GRADS program. Section II outlines our methodology, Section III summarizes the key findings from our outcome evaluation, and Section IV presents our benefit-cost analysis of GRADS. A Technical Appendix is provided for supplemental analysis and technical detail.

I. Background

The teen birth rate in Washington State has declined over the past few decades. In 2014, the birth rate for 15–17-year-olds was 8.32 per 1,000 females—the lowest rate since tracking began in 1980 (Exhibit 1). Even with this decline, there were still more than 4,000 births to mothers age 19 or below in 2014 (including more than 1,000 to 15-17-year-olds).[4]

Exhibit 1

Annual Teen Birth Rate per 1,000 Females, by Age Group



Source: WSIPP calculations using birth data from 1980-2014 retrieved from WA Department of Health at http://www.doh.wa.gov/DataandStatisticalReports/VitalStatisticsData/Birth/BirthTablesbyYear and intercensal and postcensal population estimates retrieved from the Office of Financial Management at http://www.ofm.wa.gov/pop/asr/default.asp. The calculation is based on the method used in Vollan, T., & Saffod, C. (2014). Teen pregnancy and childbearing (Health of Washington State). Olympia, WA: Department of Health. Retrieved from http://www.doh.wa.gov/Portals/1/Documents/5500/MCH-TPC2014.pdf

[4] Natality Table A3. Mother's Age Group by Child's Sex for Residents, 2014. Retrieved from http://www.doh.wa.gov/DataandStatisticalReports/VitalStatisticsData/Birth/BirthTablesbyYear.

Teen childbearing has been linked to negative outcomes for teen parents and their community. Nationally, an estimated 49% of young women who give birth before the age of 19 do not receive a high school diploma by age 22 and 30% of teen girls who have dropped out of high school cite pregnancy or parenthood as a reason.[5] Teen childbearing has also been linked to a reduction in years of postsecondary education completed, reduced likelihood of employment, reduced earnings, and an increased probability of receiving public assistance.[6] In addition, local communities and states can bear additional costs through lost tax revenue, increased use of public health care, and more.[7]

The GRADS program aims to mitigate these negative outcomes by providing students with a specialized curriculum and accessible child care to help them stay in school and learn parenting and work-related skills. This report is the first rigorous evaluation of the impact, benefits, and costs of GRADS in Washington State.

[5] Perper, K., Peterson, K., & Manlove, J. (2010). *Diploma attainment among teen mothers* (Child Trends Fact Sheet Publication #2010-01). Washington DC: Child Trends; National Campaign to Prevent Teen and Unplanned Pregnancy. (2012). *Why it matters: Teen childbearing, education, and economic wellbeing*. Washington D.C.: Author.

[6] Hoffman, S.D. (2008). Updated estimates of the consequences of teen childbearing for mothers. In S.D. Hoffman, & R.A. Maynard (Eds.), *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy* (2nd ed., pp. 74-118). Washington D.C.: The Urban Institute Press.; Fletcher, J.M., & Wolfe, B.L. (2008). *Education and labor market consequences of teenage childbearing: Evidence using the timing of pregnancy outcomes and community fixed effects* (NBER Working Paper 13847). Cambridge, MA: National Bureau of Economic Research; Ashcraft, A., Fernandez-Val, I., & Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal, 123*, 571, 875-905.

[7] For additional information on the impact of teen birth, see our Technical Documentation at http://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf

GRADS Program Description

The GRADS program began in the early 1980s in both comprehensive and alternative high schools in Washington.[8] Currently, the program operates in 23 school districts across the state (including locations on the east and west side of the state and in urban and non-urban areas) and serves 400-500 students per year.[9]

The program provides students the opportunity to earn high school credit in a series of courses developed at the local level based on the Work and Family Foundations areas of the National Standards for Family and Consumer Sciences Education (FACSE). The courses cover topics including nutrition and wellness; human development; parenting; and career connections. GRADS teachers must have FACSE certification, complete GRADS training provided by OSPI at least once every five years, and complete additional locally determined training.[10]

Child care (including infant care) is available on-site or in an easily accessible location that meets the state's licensing requirements.[11] Some local GRADS programs choose (but are not required) to provide additional support to students, including financial assistance, case management, service coordination, and more.[12]

[8] http://www.k12.wa.us/CareerTechEd/GRADS/GRADSNewsletterSpring2014.pdf.

[9] http://www.k12.wa.us/CareerTechEd/GRADSProgramMap.aspx.

[10] http://www.k12.wa.us/CareerTechEd/pubdocs/GRADSProgramPamphlet.pdf.

[11] For additional information on licensing requirements, please see WAC 170-290-0125 and WAC Chapter 170-295.

[12] Personal communication with Denise Mileson, GRADS program specialist at OSPI, April 2016.

II. Evaluation Methodology

In this evaluation, we assess the impact of participation in GRADS on high school graduation and postsecondary course enrollment. In order to estimate the effects of GRADS on education outcomes, we must compare outcomes of GRADS participants to a similar group of students who met eligibility requirements but did not participate.

Program evaluations often exhibit "selection bias" when participation is not random (e.g. when individuals choose whether or not to participate), which means that the characteristics of participants may vary systematically from non-participants and this may, in turn, affect observed differences in outcomes. For example, students who qualify for free- or reduced-priced meals may be less (or more) likely to participate in GRADS and less likely to graduate, while students who are more motivated may be more (or less) likely to participate and to graduate.

Ideally, we would test the impact of the GRADS program using a randomized controlled trial—the "gold standard" experimental approach to estimating treatment effects. Random assignment to treatment allows for a direct, unbiased comparison of outcomes between participants and non-participants and eliminates selection bias because participation is not confounded with either observable characteristics like socioeconomic status or unobservable characteristics like intrinsic motivation.[13]

Since students choose whether or not to participate in GRADS, we are unable to use this approach. Instead, we use an advanced statistical technique called propensity score matching. This technique allows us to compare outcomes of participants and non-participants after matching on observable baseline characteristics and is used as a way to approximate the covariate balance and lack of selection bias found in randomized controlled trials.[14] However, we recognize that propensity score matching may not eliminate all selection bias due to unobservable characteristics that may affect participation and outcomes.

We use administrative data obtained from DOH and the Education Research and Data Center (ERDC)[15] to evaluate the program. All data was de-identified prior to receipt by WSIPP, which means that it did not contain any information that could directly identify a participant (such as names, addresses, or Social Security Numbers).

[13] Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424.

[14] Ibid.

[15] For additional information on the ERDC, please see http://erdc.wa.gov/. The ERDC states that "The research presented here utilizes confidential data from the Education Research and Data Center (ERDC), located within the Washington Office of Financial Management (OFM). Committed to accuracy, ERDC's objective, high-quality data helps shape Washington's education system. ERDC works collaboratively with educators, policymakers and other partners to provide trustworthy information and analysis. ERDC's data system is a statewide longitudinal data system that includes de-identified data about people's preschool, educational, and workforce experiences. The views expressed here are those of the authors and do not necessarily represent those of the OFM or other data contributors. Any errors are attributable to the authors."

The project was approved by the Washington State Institutional Review Board and DOH in spring 2016.[16] The following subsections provide additional detail about our data, the selection of the study groups in the analytical sample, outcome measures, and methods.

Data and Study Groups

To test the impact of GRADS, we measured outcomes for students who participated in the program at any time between 2007 and 2013. This time period allows us to observe a portion of students for a sufficient amount of time after program participation to capture impacts on postsecondary outcomes while also including a large enough sample to improve the accuracy of the estimation of impacts on high school graduation.

Our first task was to identify students who participated in the program (the "treatment group") and a sufficiently similar group of students who met eligibility requirements but did not participate (the "comparison group"). In order to participate in GRADS, a student must be either pregnant or parenting and enrolled in a high school with access to the program. Thus, the comparison group must be drawn from students who are enrolled in high school and are either pregnant or parenting.

OSPI does not collect information on the pregnancy or parenting status of students, and we were therefore unable to use school records to identify a sample.

[16] https://www.dshs.wa.gov/sesa/research-and-data-analysis/human-research-review-section.

Instead, we use birth record data maintained by DOH to identify individuals (both female and male) who became a teen parent in Washington from 2007-2013 and linked those individuals to educational records maintained at the ERDC. Since birth record data contain more detailed information for mothers, the analysis and results presented in the following sections are for females only.[17]

To construct the initial dataset, DOH identified every live birth from 2007-2013 in which at least one parent was age 19 or younger.[18] These individual-level data were sent directly to the ERDC where they were linked to records from the K–12 and higher education sectors including information on students' demographic characteristics; state assessment results; program eligibility and participation including free- and reduced-price meals, special education, and the Transitional Bilingual Instruction Program; and postsecondary enrollment.

Using these linked data, our initial sample contains 18,076 teen mothers with a live birth in high school between 2007 and 2013.[19] This sample excludes students who participated in the Teen Parenting

[17] In our main analysis sample, approximately 93% of GRADS participants are female.

[18] Many teen parents remain in school past their expected graduation year (i.e. for more than four years). Thus, we requested individuals up to 19 years old in order to capture information on teen parents who were still in or eligible to be in school after age 18.

[19] The linked sample obtained from ERDC includes data on over 52,000 individuals (female and male) with a live birth in which at least one parent was 19 or under. We excluded more than 25,000 individuals whose pregnancy occurred after leaving high school (e.g. the pregnancy occurred at age 18 after the student had graduated) as they would not have been eligible for GRADS.

program[20] or who participated in GRADS outside of the analysis period.

We use data relative to an "index" year and birth in our analyses. The index year and birth is the first time in which a pregnancy was observed in high school for both the treatment and comparison students.[21] Students without valid academic or pregnancy data in the relevant years are excluded from the sample.[22]

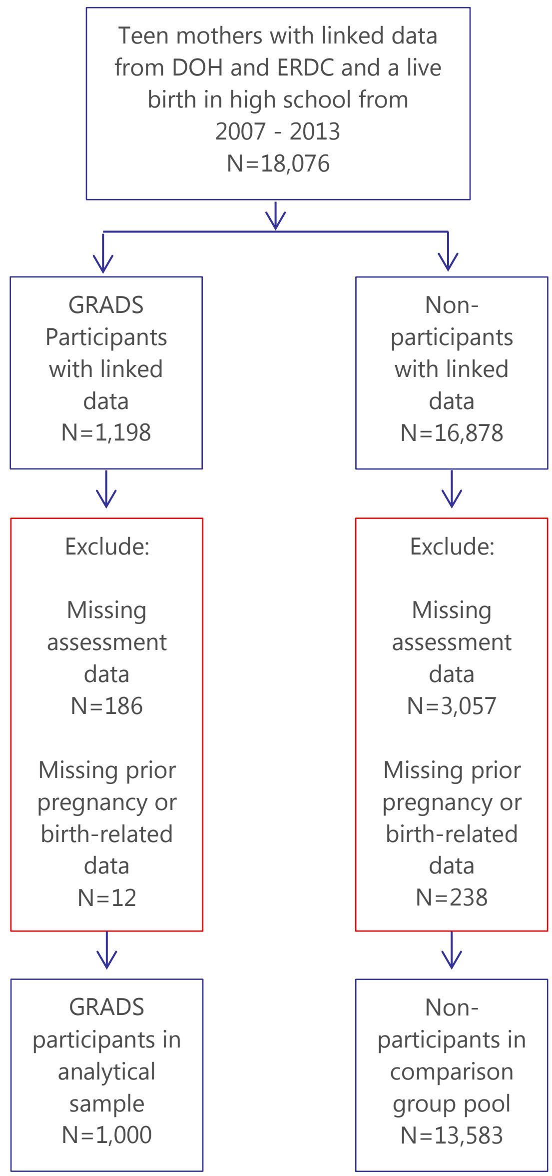
We identified the treatment group using annual course enrollment records collected by OSPI. Students were categorized as a participant if they were enrolled in GRADS at any time from 2007-2013 in the annual Career and Technical Education Student Enrollment File.[23] Any student in the sample that did not participate in GRADS was included in the comparison pool.

Our analytical sample (before matching) includes 1,000 students in the treatment group who participated in GRADS between 2007 and 2013 and 13,583 students in the comparison group who had a live birth while in high school but did not participate in GRADS during the same period.

The flow of participants and non-participants from the initial sample to the analytical sample, including the reasons various groups were removed from the analysis, are presented in Exhibit 2.

Exhibit 2

Creating the analytical sample



[20] The Teen Parenting program offers students similar coursework to GRADS but does not include other components like child care. We exclude 969 students (female and male) that participated in Teen Parenting courses to avoid confounding the estimate of the impact of GRADS.

[21] GRADS participants vary in the amount of time from the index year to participation. In our data, approximately 27% participated in GRADS during their index year while 51% participated the following year; a small percentage of students waited up to five years before participation.

[22] We tested alternative models that retained the dropped participants by excluding variables with missing data from the matching and outcome models and found substantively similar results.

[23] Participants are students with a reported Classification of Instructional Program (CIP) code of 190726 (GRADS) in any year between 2007 and 2013 in the P210 Voc file. Please see http://www.k12.wa.us/CareerTechEd/CodeChart.aspx for additional information on CIP codes.

Outcome Measures

Our analysis focuses on high school graduation and postsecondary enrollment.[24] Each measure is described in the following subsections.

High school graduation

Students' high school graduation status is derived from their last enrollment status in our data. In order to account for variation in the amount of time it takes students to graduate, our main outcome is high school graduation by the age of 22. We also present results for four- ("on time"), five- ("extended"), and six-year graduation rates.[25]

A student is categorized as a high school graduate when her last enrollment status at each time point is any of the following:

- received a regular high school diploma,
- completed an Individualized Education Program,[26] or
- received an adult high school diploma.

Students are classified as non-graduates at each time point when their last enrollment status is any of the following:

- continuing,[27]
- received a GED,
- confirmed dropout, or
- unknown.

Postsecondary Course Enrollment

A student is considered to have enrolled in a postsecondary course if they enroll in any credit-bearing, college-level course at a public institution of higher education in Washington.[28] Students who attend a postsecondary institution but only enroll in non-college level courses such as adult basic education or developmental courses (i.e. courses designed to remediate basic skills prior to enrolling in credit-bearing courses) are not considered to have enrolled. We present results separately for postsecondary course enrollment by the age of 22 and 24.

Methods

We use propensity score matching to select the matched comparison group from the pool of teen mothers that did not participate in GRADS. Propensity score matching has three steps.[29]

---

[24] While our analysis focuses on educational outcomes, there are a variety of other potential outcomes that could be affected by participation in GRADS including changes in the probability of employment, wages, receipt of public assistance, subsequent teen births, and outcomes of the children of GRADS participants. While these outcomes are important and could help provide a more complete picture of the impact, benefits, and costs of GRADS, they are outside the scope of this analysis.

[25] For each student in our sample, we estimate an "expected graduation year" based on typical grade progression from the year and grade we first observe the student in high school. For example, if a student is first observed in 9th grade in the 2007-08 school year, her expected graduation year would be 2010-11. The four-, five-, and six-year measures of high school graduation are relative to the expected graduation year.

[26] An Individualized Education Program or IEP is a written plan for students with a disability who receive special education. The IEP describes services the student receives, academic and behavioral goals and expectations, and more. For additional information, please see http://www.k12.wa.us/SpecialEd/Families/IEP.aspx.

[27] Students who are still continuing in their high school education at each time point are included as non-graduates, while students who have not been observed for the requisite amount of time and were continuing at their last enrollment status are excluded from the analysis.

[28] We are unable to observe students who enroll in either private or out-of-state postsecondary institutions.

[29] More detailed methods for this evaluation are described in the Technical Appendix.

First, we estimate a propensity score (the predicted probability of participating in GRADS) for each subject in the sample. We use a statistical model that includes a variety of factors that may affect the probability that a student would enroll in GRADS. We use school year, grade level, demographics, academic measures, and birth-related characteristics (see Exhibit 3 on the next page for the list of variables).

Second, we randomly sort the individuals and match treatment students to comparison group students with a sufficiently close propensity score. Our preferred model matches each treatment student to the comparison group student with the closest propensity score within a predefined range (this ensures that treatment students are not matched to highly dissimilar students in the comparison group).[30] Prior to matching, we restrict comparison group students to districts without a GRADS program in order to reduce the possibility of selection bias.[31]

After matching, the final sample for our main outcome (high school graduation by age 22) is 920 GRADS participants and 986 comparison group students.[32] Exhibit 3 (next page) reports the means and percentages for all variables used in our analysis before and after matching.

Before matching, the treatment group is significantly different from the comparison group across several variables. GRADS participants are less likely to be white, are younger at first birth, and have more children and prior pregnancies by the index year. Treatment students are also more likely to be eligible for free- and reduced-price meals, in the Transitional Bilingual Instructional Program, transfer between school districts, and have lower assessment scores. After matching, the two groups balance across all variables (Exhibit 3, next page).[33]

Third, we perform an outcome analysis using this matched sample. We employ logistic regression on the matched sample to estimate the impact of GRADS participation on education outcomes.[34] The results of this analysis are presented in the next section (page 10).

[30] We use nearest neighbor caliper matching without replacement and match on the logit of the propensity score as recommended in Austin, P.C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine, 33*, 1057-1069. We use a caliper equal to 0.1 times the standard deviation of the logit of the propensity score and include ties (comparison group students with identical propensity scores). We tested other matching models including nearest neighbor with replacement, 1:5, and more. Each model showed sufficient balance across the covariates and reduced overall bias. We select our preferred model based on recommendations in the literature, parsimony, and ease of interpretation. We present results using the alternative matching models in the Technical Appendix.

[31] Restricting the comparison group to students in non-GRADS districts may help reduce selection bias because the sample does not include students who actively chose not to participate. We also tested matching and outcome models using comparison group students from GRADS districts only and students from any district in the state. Additional details and results for these approaches are presented in the Technical Appendix.

[32] Since we observed students for varying amounts of time, matching is conducted separately based on the available sample for each outcome. The matched number in the comparison group exceeds the number in the treatment group because we include ties in our model. Outcome analyses are weighted to account for these students.

[33] We assess balance using a variety of diagnostics. Please see the Technical Appendix for additional detail.

[34] We tested several models to estimate outcomes. We present results using the unmatched sample, using alternative comparison group restrictions, and using alternative matching models in the Technical Appendix.

Exhibit 3

Study Group Characteristics

High School Graduation by Age 22 Sample

Variable	Before matching		After matching	
	GRADS (n=939)	Comparison (n=8,575)	GRADS (n=920)	Comparison (n=986)[1]
Percent 2006 school year (index)	0.029	0.034	0.029	0.030
Percent 2007 school year (index)	0.141	0.146	0.142	0.149
Percent 2008 school year (index)	0.162	0.163	0.162	0.157
Percent 2009 school year (index)	0.183	0.169	0.179	0.172
Percent 2010 school year (index)	0.142	0.164	0.143	0.146
Percent 2011 school year (index)	0.160	0.147	0.157	0.163
Percent 2012 school year (index)	0.144	0.121*	0.146	0.147
Percent 2013 school year (index)	0.040	0.057*	0.041	0.037
Percent Hispanic	0.447	0.329*	0.441	0.443
Percent White	0.430	0.537*	0.434	0.418
Percent Black	0.063	0.046*	0.064	0.077
Percent Native	0.026	0.052*	0.026	0.026
Percent other race/ethnicity	0.034	0.036	0.035	0.035
Age (index year)	17.2	17.6*	17.2	17.1
Percent in 9th grade (index year)	0.240	0.122*	0.233	0.248
Percent in 10th grade (index year)	0.306	0.208*	0.305	0.313
Percent in 11th grade (index year)	0.297	0.327	0.302	0.279
Percent in 12th grade (index year)	0.158	0.343*	0.160	0.160
State reading assessment (z-score)	-0.008	0.177*	0.001	0.013
Percent homeless (prior year)	0.045	0.037	0.046	0.042
Percent free- and reduced-price meals (prior year)	0.759	0.691*	0.755	0.767
Percent transitional bilingual program (prior year)	0.151	0.092*	0.147	0.162
Percent special education (prior year)	0.121	0.097*	0.122	0.117
Percent gifted (prior year)	0.009	0.002*	0.005	0.007
Num. of school transfers within district (prior year)	0.327	0.339	0.327	0.326
Num. of school transfers outside district (prior year)	0.348	0.208*	0.315	0.276
Estimated age at first birth	016.9	17.5*	16.9	16.9
Number of children by index birth	0.121	0.029*	0.108	0.098
Num. of prior pregnancies by index birth	0.244	0.147*	0.230	0.220
Percent smoked prior to pregnancy (index birth)	0.154	0.183*	0.153	0.158

*Indicates a significant difference between treatment and comparison groups at $p < 0.05$

[1] We use ties (comparison group members with identical propensity scores) in our analyses and weight accordingly.

Notes:

"Index" year and birth is the first time in which a pregnancy was observed in high school. "Prior" is the most recent available data in any year before the index event. Reading assessment scores were standardized within grade and year with a mean of 0 and a standard deviation of 1.

III. Evaluation Findings

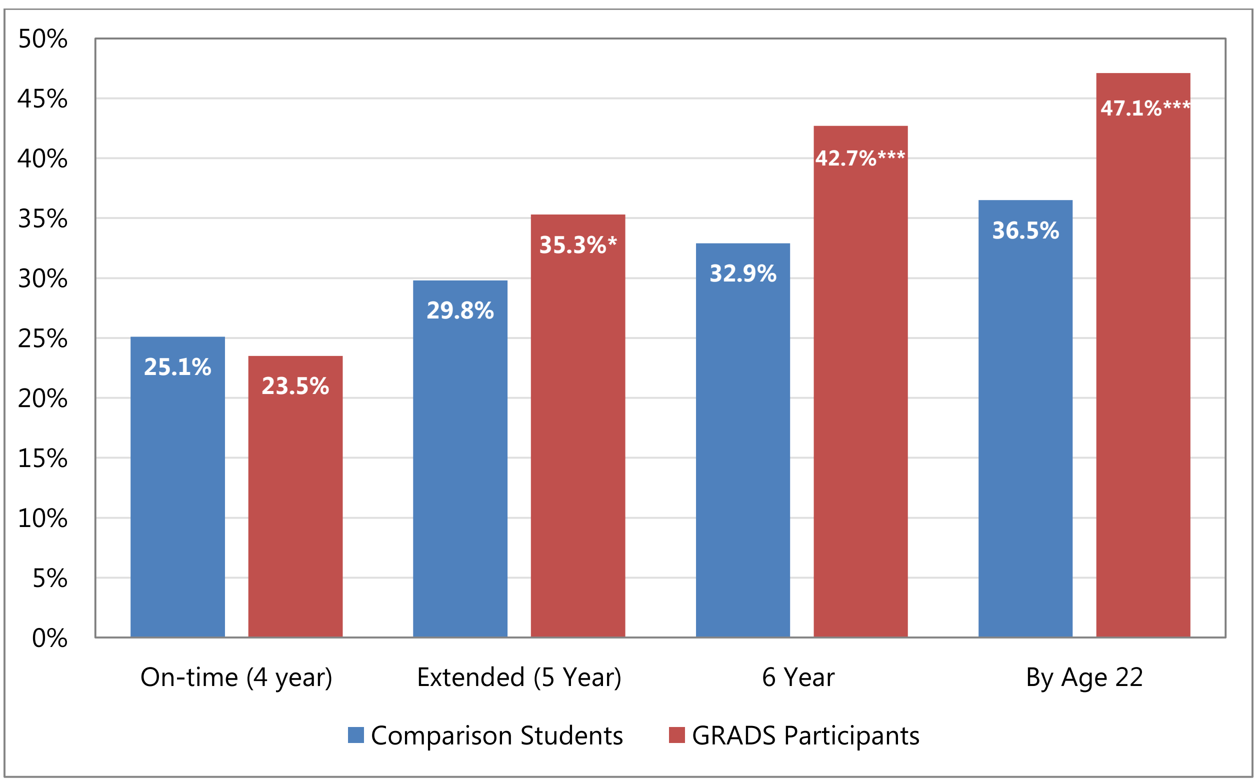
We analyze the effect of GRADS participation by teen mothers on high school graduation and postsecondary course enrollment and present results for each outcome in the following subsections.

High School Graduation

We present our regression-adjusted estimates for each high school graduation measure in Exhibit 4. We find that GRADS is associated with slightly lower, though non-significant, on-time graduation rates with 23.5% of GRADS participants and 25.1% of comparison group students graduating on time. The graduation rate rises for both groups as students are given more time to complete. GRADS participation is associated with positive, significant, and increasingly larger effects at each additional measure. By age 22, our main outcome and the estimate used in our benefit-cost analysis, 47.1% of GRADS participants and 36.5% of the comparison group graduate, a difference of 10.6%.

Exhibit 4

Regression-Adjusted High School Graduation Rates



* Significant at $p < 0.05$, *** Significant at $p < 0.001$

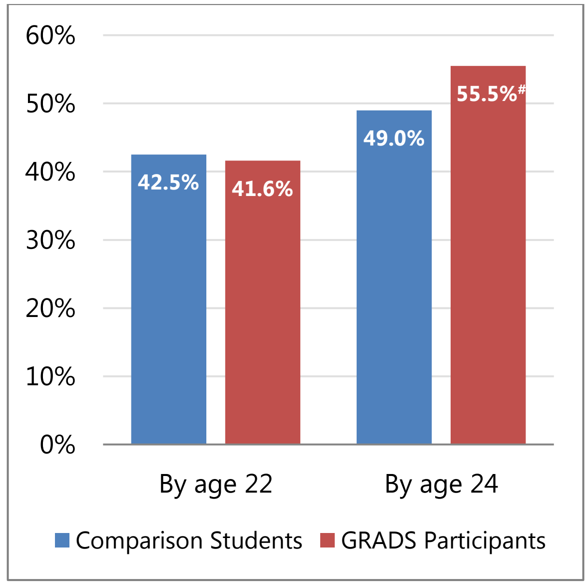
Postsecondary Course Enrollment

The pattern of results for enrollment in any postsecondary course is similar to the pattern observed for high school graduation.

We find that participation in GRADS is associated with slightly lower, though non-significant, enrollment in a postsecondary course by age 22. However, we find that GRADS has a positive and marginally significant effect by age 24. Our regression-adjusted results are presented in Exhibit 5.

Exhibit 5

Regression-Adjusted Proportion Enrolling in a Postsecondary Course



Significant at $p < 0.1$

IV. Benefit-Cost Analysis

In addition to the impact of a program on participants' outcomes, WSIPP analyzes the benefits and costs associated with implementing a program.

For example, a program that produces an increase in the probability of high school graduation can lead to benefits for program participants, taxpayers, and other people in society through increased employment, greater tax revenue, and "spillover" effects. An increase in the rate of high school graduation can also lead to reductions in the probability of crime, reductions in the use of publicly-provided healthcare, and more. These benefits can then be compared to the cost to implement a program in order to estimate an overall return on investment.

Our model provides an internally consistent estimate of the benefits and costs of programs so that options can be compared on an apples-to-apples basis.[35] We present our results using standard financial summary statistics including net present values and benefit-cost ratios. We also provide an estimate of risk that accounts for the uncertainty present in any statistical or benefit-cost estimate using a "Monte Carlo simulation" that varies the key factors in our calculations. Additional detail on our benefit-cost methods can be found in the Technical Documentation.[36]

The following describes our cost estimate and the results of our benefit-cost analysis.

[35] Washington State Institute for Public Policy (2016). *Benefit-cost technical documentation*. Olympia, WA: Author.
[36] http://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf

Cost Estimates

Our model estimates benefits and costs on a per-participant basis. Thus, the first step is to estimate the annual per-participant cost to provide GRADS and any annual costs that can be attributed to comparison group students.

The basic components of GRADS (i.e. classes related to work and parenting skills and accessible child care) are similar across sites, but actual implementation of the program varies considerably. For example, sites differ in the number of students served each year and the length of the GRADS class period each day. Sites also differ on the model used to provide child care. Many sites provide child care on-site, using staff hired by the district, while other sites contract with outside providers.[37]

While we recognize that implementation costs vary across sites, we estimate the annual per-student cost to implement a "typical" or average program. We do, however, account for this variation in our risk analysis.[38]

Our annual per-student cost estimate for a typical program is based on the cost to implement the program at one school.[39] The estimate assumes that the school serves 15 students per year and offers one class period (one hour) of GRADS each day

[37] Personal communication with Denise Mileson, GRADS program specialist at OSPI, April 2016.
[38] In this analysis, we allow the cost to vary within +/- 20% of our mean estimate.
[39] Our estimate is based on program implementation information provided by the GRADS program specialist at OSPI.

led by a certificated teacher.[40] The estimate assumes that the school provides an on-site child care center staffed by three district employees.[41] Finally, we include costs for training,[42] supplies for the child care center,[43] and additional materials, supplies, and operating costs to provide the courses.[44] Using these factors, we estimate an annual per-student cost to provide GRADS of $11,207 (2014 dollars).

To estimate the cost for the typical comparison group student, we assume that the student is in school pursuing a high school diploma and thus qualifies for the state's Working Connections Child Care subsidy.[45] To estimate the per-student annual cost of the subsidy, we use the average daily subsidy rate for licensed or certified child care centers[46] and assume that students are in school for 180 days per year. Using these factors, we estimate an annual per-student comparison cost of $6,045 (2014 dollars).

Thus, the net cost to provide GRADS is $5,162 per-student per year. The net cost primarily reflects the additional costs of offering the GRADS courses and providing child care on-site.

Benefit-cost Results

Exhibit 7 (next page) shows the results of our benefit-cost analysis. Our estimate is based on the effect of GRADS on high school graduation by age 22.

In our data, teen mothers participate in GRADS for approximately 1.5 years on average, so the net per-student cost to provide GRADS is $7,588 (2015 dollars). GRADS participation results in total benefits of $22,839 per participant due to changes in labor market, health-care, crime, and education-related factors (Exhibit 7).

Thus, we estimate net benefits (benefits minus costs) of $15,251 per participant and a benefit-cost ratio of $3.01. Finally, our estimate of risk shows that GRADS produces positive net benefits 93% of the time.

[40] We estimate the cost of teacher time using 2013-14 average hourly compensation (including all benefits) for teachers in grades 9-12 in Washington State.

[41] The estimate assumes that the salary and benefits for these employees is equivalent to the average hourly compensation (including all benefits) for an instructional aide in Washington State.

[42] In addition to accounting for teacher time to attend four days of training, we include $1,600 per teacher to cover travel, lodging, and substitute costs.

[43] We use an annual cost of $5,000 in a typical center to cover costs such as diapers, toys, and other supplies.

[44] We use the difference ($56) between the per-student Basic Education allocation for materials, supplies, and operating costs (MSOC) for students in grades 9-12 ($1,376) and the per-student allocation for students in approved career and technical education programs ($1,432) from the 2015-16 school year and adjust for inflation to 2014 dollars using the implicit price deflator. The per-student MSOC allocations are from Washington State's 2015-17 supplemental operating budget (Second Engrossed Substitute House Bill 2376, Chapter 36, Sec. 502, Laws of 2016).

[45] Students may qualify for the subsidy when engaged in basic education activities under certain circumstances. Please see WACs 170-290-0005; 170-290-0040; and 388-310-0900 for more information.

[46] To calculate the average daily rate, we include the subsidy for infants (one to 11-months old) and toddlers (12-29 months old) across all regions in the state for 2015 and adjust for inflation to 2014 dollars using the implicit price deflator. Please see WAC 170-290-0200 for subsidy rates and information.

Exhibit 7

Benefits and Costs per Participant for GRADS vs. Comparison Group in 2015 Dollars

Program cost		
GRADS participants		
Cost per participant including GRADS courses, on-site child care, and other factors		($16,473)
Comparison group costs		($8,885)
	(1) Net GRADS cost	**($7,588)**
Labor market effects		
Increased income to participants due to increased labor market participation		$15,433
Increased tax revenue to taxpayers due to increased labor market participation		$7,009
Positive externalities ("spillover effects") to society due to greater number of high school graduates		$7,093
Education effects		
Increased postsecondary costs for participants due to increased probability of attending college		($1,099)
Increased cost to taxpayers due to increased probability of attending college		($1,544)
Increased cost to others due to increased probability of attending college		($411)
Health care-related effects		
Increased insurance costs for participants due to shift from public to private or employer provided insurance		($1,134)
Decreased cost to taxpayers due to shift from public to private or employer provided insurance		$4,199
Increased costs to private or employer provided insurance programs		($4,561)
Crime-related effects		
Decreased cost to taxpayers due to reduced probability of crime		$83
Decreased crime victim costs due to reduced probability of crime		$197
Deadweight cost of taxation		($2,427)
	(2) Total benefits	**$22,839**
Bottom line:		
Net benefits (cost) per participant	(3) Net (benefits – costs)	**$15,251**
Benefit-to-cost ratio		**$3.01**
Probability of positive net benefits (risk analysis)		**93%**

Technical Appendix

Graduation, Reality, and Dual-role Skills (GRADS) Program for Pregnant and Parenting Teens:
Outcome Evaluation and Benefit-Cost Analysis

Appendix

Ideally, we would test the impact of GRADS using a random assignment approach to assign eligible students to GRADS or a comparison group. Successful random assignment to treatment allows for an unbiased comparison of outcomes between participants and non-participants that is not confounded by observable characteristics (like socioeconomic status) or unobservable characteristics (like intrinsic motivation); thus, any differences in outcomes can be attributed to the effect of the treatment. However, since GRADS is only available in some districts and students choose to participate (and thus assignment is not random), we are unable to use this approach.

Instead, we use administrative data and a propensity score matching approach. This technique allows us to compare outcomes of participants and non-participants after matching on observable baseline characteristics and is used as a way to approximate the covariate balance and lack of selection bias found in randomized controlled trials.[47] However, we recognize that propensity score matching may not eliminate all selection bias due to unobservable characteristics that may affect outcomes. Thus, we present results using a variety of matching models and sample constructions to test the sensitivity of our estimates.

In Section A.I. of this Technical Appendix, we discuss sample selection and propensity score matching procedures (including alternative samples and matching models). In Section A.II. we discuss the results of our outcome analyses under a variety of approaches.

A.I. Study Group Selection & Matching Procedures

Study groups

To test the impact of GRADS, we measure outcomes for students who participated in the program at any time between 2007 and 2013. We identify participants using annual course enrollment records collected by the Office of Superintendent of Public Instruction (OSPI). Students were categorized as a participant if they were enrolled in a course with a Classification of Instructional Program (CIP) code of 190726 (GRADS) in any year between 2007 and 2013 in the annual Career and Technical Education Student Enrollment File (P210 Voc).

[47] Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424.

Comparison group students are teen parents who had a live birth in high school over the same period. Any teen parent that did not participate in GRADS (or the similar Teen Parenting[48] program) was included in the comparison pool. Due to the availability of data from birth records specific to females and a smaller sample size of male participants, we focus our main analyses on teen mothers.

For our main analysis, we restrict comparison group students to districts that do not have a GRADS program. Restricting the comparison group to non-GRADS districts may help reduce selection bias because the sample does not include students who actively chose not to participate. That is, we can assume that a portion of the reason students in non-GRADS district did not participate was due to program availability rather than unobservable characteristics like motivation.

However, we test this assumption by using alternative approaches. We present results for matching and outcome models using our main or “between districts” sample that restricts the comparison group to non-GRADS districts, a “within districts” sample that restricts comparison group students to districts with a GRADS program, and an “across districts” sample in which students may be drawn from any district in the state.

We recognize that the between district restriction does not eliminate selection bias. Indeed, there may be a variety of reasons that students choose to participate in GRADS that are not fully addressed by this restriction. Thus, we employ propensity score matching to balance the treatment and comparison students on observable characteristics and further mitigate the influence of selection bias. The following subsection discusses our matching approach.

Propensity Score Matching

As previously discussed, propensity score matching allows us to match GRADS participants to similar students that did not participate in GRADS in order to balance the two groups on observable variables. This procedure helps to reduce selection bias and ensures that we are not comparing GRADS students to highly dissimilar students who are not likely participate in the program. In addition, we conduct regression on the matched sample using the variables included in the matching model to further control for residual bias.

We first estimate a propensity score (the predicted probability of participating in GRADS) for each subject in the sample. We use a statistical model that includes a variety of factors that may affect the probability that a student would enroll in GRADS and/or students’ outcomes. We match on school year, grade level, demographics, academic measures, and birth-related characteristics.[49] Exhibit A1 (next page) reports the results from this first stage model estimating the likelihood of GRADS participation.[50]

[48] The Teen Parenting program offers students similar coursework to GRADS, but does not include other components like child care. We exclude 969 students (female and male) that participated in Teen Parenting courses to avoid confounding the estimate of the impact of GRADS.
[49] Since we observed students for varying amounts of time, matching is conducted separately based on the available sample for each outcome.
[50] The propensity score estimation and matching were conducted in STATA using version 4.0.11of the following: Leuven, E. & Sianesi, B. (2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing". From http://ideas.repec.org/c/boc/bocode/s432001.html.

Exhibit A1

Logit Model Estimating the Likelihood of GRADS Participation
High School Graduation by Age 22 Sample

Covariate	Coefficient		SE
Year 2006 (index)	-0.121		0.272
Year 2007 (index)	0.087		0.199
Year 2008 (index)	0.144		0.195
Year 2009 (index)	0.266		0.193
Year 2010 (index)	0.012		0.198
Year 2011 (index)	0.359	#	0.196
Year 2012 (index)	0.468	*	0.197
Grade 9 (index)	1.024	***	0.171
Grade 10 (index)	0.963	***	0.137
Grade 11 (index)	0.583	***	0.116
Free- and reduced-priced meals (prior)	0.016		0.088
Special education (prior)	0.211	#	0.118
Transitional bilingual program (prior)	0.163		0.118
Gifted program (prior)	1.489	**	0.449
Homeless (prior)	0.110		0.180
Number of transfers within district (prior)	-0.034		0.047
Number of transfers outside district (prior)	0.227	***	0.050
State reading assessment (z score)	-0.080	#	0.042
Hispanic	0.239	**	0.090
Black	0.429	**	0.156
Native	-0.695	**	0.222
Other race/ethnicity	0.032		0.201
Estimated age at first birth	-0.027		0.059
Number of prior pregnancies by index birth	0.052		0.085
Number of prior children by index birth	1.490	***	0.187
Age (index)	-0.126	#	0.067
Mother smoked prior to index pregnancy	0.018		0.105
Constant	-0.694		0.933
N	9,514		
Pseudo R2	0.08		
AUC	0.70		

# p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001

After estimating the probability of treatment, we randomly sort the individuals and match treatment students to comparison group students. In our preferred model, we use nearest neighbor caliper matching without replacement and match on the logit of the propensity score.[51] We use a caliper equal to 0.1 times the standard deviation of the logit of the propensity score[52] and include ties (comparison group students with identical propensity scores). We tested other matching models including nearest neighbor with replacement, one to many matching (e.g. one treatment student to five comparison group students), and more. We present matching and outcome results using these alternative methods below and in Section A.II.

We used several approaches to assess how well our models improved balance and reduced bias between the two groups. Exhibit A2 (next page) presents results before and after matching employing a frequently used method to assess balance called the absolute standardized difference (bias), which is the difference in the mean or proportion for each covariate for the treated and comparison groups divided by the pooled standard deviation prior to matching.[53] An absolute standardized difference above 25 indicates substantial imbalance between the two groups and recommendations indicate that the difference should be below ten to consider the covariate balanced.[54] As shown in Exhibit A2, our preferred method successfully balanced the treatment and control groups with an absolute standardized difference well below ten for each covariate after matching.

In addition to absolute standardized bias for each covariate, we assessed our approach using measures based on the balance of the overall model. In Exhibit A3, we present results for our preferred model and for alternative matching models and samples including the "between," "within," and "across" districts approaches discussed previously. We assessed each model using Rubin's B (the standardized difference in the means of the linear prediction of the propensity score), Rubin's R (the ratio of variance in the treated and comparison group for the linear prediction of the propensity score), and the mean and median standardized difference (bias) across all of the covariates included in the model. Mean and median bias below 25 indicates sufficient balance, while Rubin's B should be less than 25 and Rubin's R should be between 0.5 and 2.[55]

As shown in Exhibit A3, the treatment and comparison groups are imbalanced prior to matching in each of the district-based samples. After matching, balance measures improve for every model in each of the district-based samples, with all measures within the recommended ranges. Given that each of our models show sufficient balance across a range of diagnostics, we select our preferred model (nearest neighbor caliper matching on the between district sample) to reduce selection bias and based on recommendations in the literature, ease of interpretation, and parsimony. In Section A.II., we discuss the methods used in the outcome analysis on the matched sample and present results for our preferred and alternative models.

---

[51] We employ this method for our main analysis based in part on recommendations in Austin, P.C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33, 1057-1069 and Rosenbaum, P.R., & Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33-38.
[52] We tested several calipers including a caliper equal to 0.2 times the standard deviation of the logit of the propensity score as suggested in Austin (2011).
[53] All calculations used to assess balance were conducted using pstest in STATA.
[54] Austin, P.C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083-3107.; Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science : A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1–21.
[55] Stuart (2010) Rubin, D.B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2(*3-4), 169-188.

Exhibit A2

Covariate Balance Before and After Matching

High School Graduation by Age 22 Sample

	After matching			Absolute standardized difference	
Variable	**GRADS**	**Comparison group**	**p-value**	**Before matching**	**After matching**
Year 2006 (index)	0.029	0.030	0.891	-3.2	-0.6
Year 2007 (index)	0.142	0.149	0.692	-1.6	-1.9
Year 2008 (index)	0.162	0.157	0.750	-0.2	1.5
Year 2009 (index)	0.179	0.172	0.668	3.9	2.0
Year 2010 (index)	0.143	0.146	0.895	-6.2	-0.6
Year 2011 (index)	0.157	0.163	0.703	3.6	-1.8
Year 2012 (index)	0.146	0.147	0.947	6.8	-0.3
Grade 9 (index)	0.233	0.248	0.445	**30.9**	-4.0
Grade 10 (index)	0.305	0.313	0.724	**22.5**	-1.8
Grade 11 (index)	0.302	0.279	0.281	-6.5	4.9
Free- or reduced-priced meals (prior)	0.755	0.767	0.548	**15.3**	-2.7
Special education (prior)	0.122	0.117	0.774	7.8	1.4
Transitional bilingual program (prior)	0.147	0.162	0.367	**18.0**	-4.7
Gifted program (prior)	0.005	0.007	0.762	8.8	-1.5
Homeless (prior)	0.046	0.042	0.733	3.9	1.6
Number of transfers within district (prior)	0.327	0.326	0.978	-1.5	0.1
Number of transfers outside district (prior)	0.315	0.276	0.186	**20.4**	5.7
State reading assessment (z score)	0.001	0.013	0.788	**-19.1**	-1.3
Hispanic	0.441	0.443	0.925	**24.4**	-0.4
Black	0.064	0.077	0.275	7.4	-5.8
Native	0.026	0.026	1.000	**-13.7**	0.0
Other race/ethnicity	0.035	0.035	1.000	-1.3	0.0
Estimated age at first birth	16.93	16.88	0.432	**-51.3**	3.8
Number of prior pregnancies by index birth	0.230	0.220	0.672	-7.7	-1.2
Number of prior children by index birth	0.108	0.098	0.516	**19.0**	2.1
Age (index)	17.17	17.13	0.399	**33.3**	3.5
Mother smoked prior to index pregnancy	0.153	0.158	0.797	**-40.8**	4.0

Note: **Bolded text indicates imbalance**

Exhibit A3

Overall Model Balance across Different Matching Methods
High School Graduation by Age 22 Sample

Matching method	Treated N	Comparison group N	Rubin's R	Rubin's B	Median bias	Mean bias
Between Districts						
Unmatched	939	8,575	1.31	74.70	7.80	14.00
1:1 Nearest neighbor w/ replacement	937	884	0.92	15.40	2.40	2.50
1:1 Nearest neighbor w/out replacement	**920**	**986**	**1.00**	**13.60**	**1.80**	**2.20**
1:5 Nearest neighbor	937	3,154	1.06	9.40	1.10	1.40
Within Districts						
Unmatched	939	4,415	1.25	75.30	5.20	11.30
1:1 Nearest neighbor w/ replacement	932	741	1.19	23.60	3.50	3.60
1:1 Nearest neighbor w/out replacement	914	932	0.88	15.40	2.40	2.20
1:5 Nearest neighbor	932	2,339	1.00	11.20	1.80	1.90
Across Districts						
Unmatched	939	12,990	1.28	71.70	6.60	12.70
1:1 Nearest neighbor w/ replacement	938	956	1.22	17.00	2.20	2.40
1:1 Nearest neighbor w/out replacement	937	1,030	1.13	15.20	2.20	2.20
1:5 Nearest neighbor	938	3,570	1.14	8.80	0.90	1.30

Notes:

Bolded text identifies chosen matching method.

The number of matched students in the comparison group includes individuals with a tie on the propensity score.

We use a caliper of 0.1 standard deviations of the logit of the propensity score in each matching model.

A.II. Outcome Analysis Methodology and Sensitivity

To estimate the effect of GRADS on students' education outcomes we conduct multivariate logistic regression on the matched sample. We include each of the covariates used in the matching model to account for residual imbalance between the groups and, since we include ties in our preferred model, we weight using a normalized weight proportional to the number of comparison group students matched to a particular treatment student.[56]

A frequent area of debate regarding the use of propensity score matching is how to correctly estimate standard errors. A particular area of concern is whether the uncertainty in the estimation of the propensity scores used in matching process should be taken into account using methods such as bootstrapping. However, research has indicated using estimated propensity scores may be more conservative in some cases and may be more efficient since using the estimated score removes both systematic and random differences in baseline characteristics.[57] In addition, bootstrapping may be unnecessary when the covariates used in the matching model are included in the outcome model, as they are in our approach.[58] Finally, bootstrapping is designed for population inference rather than within-sample estimation and is not appropriate for use in matching models with replacement.[59] Thus, since we are estimating a within sample effect, include the covariates used to match in our outcome model, and present models with replacement in our sensitivity analyses, we present analytical standard errors clustered at the district level in our models.

Exhibit A4 reports the results of our logistic regression estimating the effect of GRADS on high school graduation by age 22 using our preferred method (nearest neighbor matching without replacement, a caliper of 0.1 standard deviations of the logit of the propensity score, and restricting comparison group students to non-GRADS districts). Exhibit A5 presents the regression-adjusted results for each outcome.

---

[56] Stuart (2010).

[57] Stuart (2010); Austin, P.C. & Small, D.S. (2014). The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Statistics in Medicine, 33*, 4306-4319.

[58] Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*, 199-236.

[59] Austin & Small (2014); Abadie, A. & Imbens, G.W. (2006). On the failure of the bootstrap for matching estimators (NBER Technical Working Paper 325). Cambridge, MA: National Bureau of Economic Research.

Exhibit A4

Logistic Regression Estimating the Effect of GRADS on High School Graduation by Age 22

Covariate	Odds Ratio		SE
GRADS (treatment indicator)	1.686	***	0.219
Year 2006 (index)	1.325		0.599
Year 2007 (index)	1.072		0.339
Year 2008 (index)	0.925		0.290
Year 2009 (index)	1.300		0.445
Year 2010 (index)	1.170		0.367
Year 2011 (index)	1.145		0.328
Year 2012 (index)	1.069		0.342
Grade 9 (index)	0.065	***	0.020
Grade 10 (index)	0.182	***	0.041
Grade 11 (index)	0.401	***	0.078
Free Reduced Priced Meals (prior)	0.973		0.117
Special Education (prior)	1.616	**	0.272
Transitional Bilingual Program (prior)	1.014		0.155
Gifted program (prior)	1.174		0.832
Homeless (prior)	1.174		0.287
Number of transfers within district (prior)	0.717	***	0.063
Number of transfers outside district (prior)	0.828	*	0.076
State reading assessment (z score)	1.708	***	0.102
Hispanic	1.331	*	0.160
Black	0.828		0.143
Native	0.665		0.203
Other race/ethnicity	0.491	#	0.190
Estimated age at first birth	0.966		0.070
Number of prior pregnancies by index birth	0.844		0.153
Number of prior children by index birth	0.822		0.225
Age (index)	0.589	***	0.060
Mother smoked prior to index pregnancy	0.427	***	0.068
Constant	1.585		0.609
N	1906		
Pseudo R2	0.136		

# p< 0.1; * p < 0.05; ** p < 0.01;*** p < 0.001

Note:

Standard errors are clustered at the district. Continuous variables are mean-centered prior to estimation.

Exhibit A5

Regression-Adjusted Effect of GRADS Participation on Education Outcomes

Outcome	GRADS	Comparison group	Percentage point difference	SE
High school graduation—4 year (on time)	23.5%	25.1%	-1.6%	0.139
High school graduation—5 year (extended)	35.3%	29.8%	5.5%*	0.174
High school graduation—6 year	42.7%	32.9%	9.8%***	0.206
High school graduation—By Age 22	47.1%	36.5%	10.6%***	0.219
Postsecondary course enrollment—By Age 22	41.6%	42.5%	-0.9%	0.175
Postsecondary course enrollment—By Age 24	55.5%	49.0%	6.5%#	0.223

Notes:
# $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Standard errors are clustered at the district
Unweighted sample sizes are as follows: 4 year (Treatment 973, Comparison 1,031), 5 year (Treatment 965, Comparison 1,029), 6 year (Treatment 947, Comparison 1,003), HS Grad by Age 22 (Treatment 920, Comparison 986), Postsecondary by Age 22 (Treatment 650, Comparison 681), Postsecondary by Age 24 (Treatment 520, Comparison 543)

Sensitivity Analyses

We tested the sensitivity of our estimates of the effect of GRADS to different matching procedures and different restrictions on the district from which we draw comparison group students. Exhibit A6 presents the estimates of the effect of GRADS on high school graduation by the age of 22 for each of the alternative approaches. We also present results from regressions on the full sample (i.e. before matching). The following subsections discuss the sensitivity of our results under each of these conditions.

Regression on the Unmatched Sample

We first estimate the association between GRADS participation and student outcomes using multivariate logistic regression on the full sample prior to matching. While regression allows for an estimate of the effect of GRADS, it has several limitations that may bias the estimate. For example, standard regression is unable to control for unobservable factors and does not accurately adjust for differences in observed covariates when the distribution of those variables between two groups are considerably different.[60] However, estimates using a standard regression approach are widely reported and can serve as a useful benchmark. Thus, we present results from regression on the unmatched sample in Exhibit A6 under our preferred and alternative models.

The results using the unmatched between districts sample indicate that GRADS participation is associated with an increase in high school graduation by the age of 22, with an odds ratio of 1.811 (se = 0.215, $p < 0.001$, 95% confidence interval 1.435 – 2.285), a somewhat larger point estimate than our preferred matching model. Overall, the results using the unmatched sample are substantively similar to results from matching models within each comparison group structure.

[60] Rubin (2001).

Exhibit A6

Effect of GRADS on High School Graduation by Age 22 Using Various Matching Models and Samples

Matching method and sample type	OR	SE	Treated N	Comp. N
Unmatched				
Between districts	1.811	0.215	939	8,575
Within districts	2.343	0.354	939	4.415
Across districts	1.941	0.231	939	12,990
Preferred model - 1:1 nearest neighbor w/out replacement				
Between districts	**1.686**	**0.219**	**920**	**986**
Within districts	2.320	0.329	914	932
Across districts	1.871	0.259	938	956
1:1 nearest neighbor with replacement				
Between districts	1.627	0.216	937	884
Within districts	2.317	0.297	932	741
Across districts	1.866	0.261	938	956
1:5 nearest neighbor				
Between districts	1.757	0.214	937	3,154
Within districts	2.393	0.358	932	2,339
Across districts	2.008	0.249	938	3,570

Notes:
Bolded text identifies preferred method.
Standard errors are clustered on the district
All results significant at $p < .001$
We use a caliper of .01 standard deviations of the logit of the propensity score in each matching model.

Alternative Matching Models

As previously discussed, we found that several matching approaches showed sufficient balance across a range of diagnostics. Thus, in Exhibit A6 we present the sensitivity of our estimate of the impact of GRADS on high school graduation by age 22 using 1:1 nearest neighbor matching with replacement and 1:5 matching. We find that estimates are relatively consistent regardless of the matching method used, with between district estimates ranging from an odds ratio of 1.627 (se = 0.216, $p < 0.001$, 95% confidence interval 1.255 – 2.109) using nearest neighbor with replacement to an odds ratio of 1.757 (se = 0.214., $p < 0.001$, 95% confidence interval 1.383 – 2.231) using 1:5 matching.

Comparison Group Restrictions

In our preferred model, we restrict comparison group students to districts without GRADS under the assumption that the restriction may help reduce selection bias since a portion of the reason students in non-GRADS district did not participate may be due to program availability rather than unobservable characteristics like motivation. However, we also tested the sensitivity of our results to one approach that restricted comparison group students to the same districts as GRADS participants (within districts) and another that allowed comparison group students to come from any district in the state (across districts).

As shown in Exhibit A6, we find that our results seem sensitive to the choice of which districts we use to draw our comparison group students. Relative to our preferred between district model, the estimate of the effect is consistently largest using the within district sample regardless of the matching method used. In many ways, using students from within the same districts is an appealing approach because each student would be subject to the same unobserved district-wide factors that may influence outcomes.

However, we cannot fully explain why students from within the same districts did not participate in GRADS despite having the opportunity. Thus, those students may be quite different from GRADS participants on characteristics that we cannot observe in the data (such as motivation) and therefore selection bias is of increased concern. The larger effects observed from the within district sample is consistent with this reasoning and the estimate may be confounded by unobservable differences between the groups.

The estimates for the across district sample are larger than in our preferred approach, but smaller than in the within district approach. While this approach is appealing because it allows a treatment student to be matched to the most similar comparison student regardless of which district that student is in, we believe including comparison group students from within GRADS districts raises the same concerns regarding selection bias and unobservable factors that may influence participation and outcomes.

Thus, while results are sensitive to the decision to restrict comparison group students to non-GRADS districts, we believe the between district approach is more effective at dealing with selection bias and provides a more accurate estimate of the effect of GRADS.

Finally, because we use a comparison group consisting of non-GRADS districts, we are unable to determine the extent to which the estimate is partially due to characteristics that differ between districts that do and do not offer GRADS. Ideally, we would include district fixed effects in our models to account for those unobserved characteristics that differ between districts. However, we are unable to use this approach because district is perfectly correlated with treatment when using the between district sample and due to lack of sample size and outcome variability in some districts. We do use standard errors clustered at the district in our outcome models to account for the non-independence of students within the same district, but we cannot account for unobserved district characteristics and acknowledge that those characteristics may at least partially explain the estimated effect of GRADS.

Acknowledgements

The authors would like to thank the Department of Health, the Education Research and Data Center, and the Office of the Superintendent of Public Instruction for the data and expertise necessary to conduct this evaluation. In particular, we would like to thank Denise Mileson and Mary Nagel from OSPI, Katie Hutchinson from DOH, and Melissa Beard from the ERDC for this assistance throughout this project.

For further information, contact:
Matt Lemon at 360.586.2744, matt.lemon@wsipp.wa.gov

Document No. 16-07-2202