CREMA

Center for Research in Economics, Management and the Arts

Impact Evaluation of an Incentive Program on Educational Achievement of Indigenous Students

Working Paper No. 2014-13

CREMA Südstrasse 11 CH - 8008 Zürich www.crema-research.ch

Impact Evaluation of an Incentive Program on Educational Achievement of Indigenous Students

Uwe Dulleck[a], Juliana Silva-Goncalves*[a] and Benno Torgler[a,b]

[a]School of Economics and Finance, Queensland University of Technology, Brisbane, QLD, Australia

[b]CREMA - Center for Research in Economics, Management and the Arts, Zürich, Switzerland

July 25, 2014

Abstract

This article introduces the Fogs Artie program that attempts to close the gap in educational attainment between Indigenous and non-Indigenous Australians, and provides an evaluation of its effectiveness. The program is of special interest as it uses in-kind incentives conditional on achievement of a specific target for academic grades, behaviour and attendance, coupled with information sessions on the importance of educational achievement. In 2012, all Indigenous students enrolled in 21 high schools in Queensland were invited to take part in the program. Using a differences-in-differences strategy, we find that the program improved behavioural and academic grades and reduced the number of unexplained absences for female students, but not for male students. In contrast, the program improved scores on a standardized national assessment test for male students. Moreover, we find that the program is only effective for students from intact families.

Keywords: Education, Incentives, Indigenous, Program Evaluation, Policy

JEL Classifications: I24, I25, I28

*Corresponding author at: School of Economics and Finance, Queensland University of Technology, Brisbane, QLD, Australia, Tel:+61731386669

Email address: j.silvagoncalves@qut.edu.au

1 Introduction

Underperformance of Indigenous populations with respect to educational attainment is common in developed countries (e.g. see Patrinos (1992) for Canada, Bradley et al. (2007) for Australia, Ladson-Billings (2006) for the US) as well as developing countries (see Patrinos (2004) for Bolivia, Ecuador, Guatemala, Mexico and Peru). It is well documented that education is an important factor in facilitating a country's economic growth, as well as at a regional or community level (Barro, 2001). Several programs have been discussed and evaluated in the literature using incentives (Angrist and Lavy, 2009; Bettinger, 2011), information (Avvisati et al., 2014), public recognition (Kremer et al., 2009) and learning support (Rodriguez-Planas, 2012), some showing significant impact while others being less successful (Fryer, 2011). This article introduces a program that attempts to close the gap in educational attainment between Indigenous (Aboriginals and Torres Strait Islanders) and non-Indigenous Australians, and provides an evaluation of its effectiveness. The program is of special interest as it combines encouragement, incentives and learning support, delivered by volunteers organized by a charity associated with a sports code popular with the target population.

Inequalities in educational achievement between Indigenous and non-Indigenous Australians begins very early in life and widens over time (De Schutter, 2008). Only 48 percent of Indigenous children attend pre-school versus 58 percent of non-Indigenous children, and only 36 percent of Indigenous Australians complete year 12 versus 75 percent of non-Indigenous Australians (Bath and Biddle, 2011; De Schutter, 2008). Besides this disparity in participation, there is also a large gap in cognitive and non-cognitive outcomes: Indigenous students perform systematically worse than non-Indigenous students on standardized tests and also tend to have worse behavioural indicators. In this study, we analyse the impact of the FOGS Artie program, which targets the educational achievement - behaviour, academic grades and attendance - of Indigenous students in 21 public high schools in Queensland. In 2012, the program offered in-kind incentives to all Indigenous students enrolled in the participating schools who reached a goal defined at the start of each school term, coupled with information sessions on the importance of educational achievement. Using individual level data, we identify the impact of the program by applying a differences-in-differences strategy, that enables comparisons of the pre and post-treatment outcomes progression for Indigenous students in program schools with that of Indigenous students in control schools over the same period. Our results indicate that the program improved both behaviour and academic grades and reduced the number of unexplained absences for female but not male students. The program did, however, improve scores on a standardized national assessment test for male Indigenous students.

Of the several factors that may contribute to this gap in educational outcomes, the most important is likely to be family environment (Heckman, 2008); specifically, the tendency for Indigenous children to come from more disadvantaged socio-economic backgrounds than non-Indigenous children. In particular, Indigenous parents, who themselves tend to have low levels of education and financial resources, may be less likely to encourage their children to perform well at school. Children from disadvantaged socio-economic backgrounds who do not benefit from a supportive family environment have lower levels of educational achievement even in their first years of formal education. A second potential explanation for the persistent educational disadvantage is the stereotype that Indigenous children perform worse at school than non-Indigenous children, which, especially when pervasive in the school community, may adversely influence the children's own beliefs about their abilities and chances of pursuing post-secondary education. Both this lack of confidence in their own abilities and the lack of a supportive family environment may lead Indigenous students to set lower goals for their educational achievement and be less motivated to perform well at school than non-Indigenous children. At the same time, because Indigenous people tend to be employed in the public sector or private firms that rely on government support or in organizations whose industrial relations practices are aimed at encouraging greater diversity within the firm (Rowse, 2002), they may perceive the economic returns from investment in education as being low. They may also see employment opportunities as very limited or non-existent in institutions without affirmative action rules (Paradies et al., 2008). Being unemployed, however, may increase social inclusion in their own group, especially if the alternative is employment in a place hostile to the Indigenous culture. On the other hand, in employment sectors or firms that reserve positions for Indigenous employees, these latter do not have to compete for these positions against non-Indigenous candidates. Hence, affirmative action policies can actually lower Indigenous people's incentives to invest in the education and training that facilitates entry into the labour market.

In light of the above observations, reducing disparities in educational achievement may be one of the most powerful instruments for diminishing overall inequalities between Indigenous and non-Indigenous Australians. Low educational achievement not only reduces employment opportunities, it directly influences individual behaviour, particularly health and consumption choices and ability to plan fertility (Conti et al., 2010). Moreover, at the societal level, early intervention aimed at narrowing the educational achievement gap is less costly than remedial policies like unemployment benefits, subsidies and increases in health care costs. Most important, the intergenerational transmission of skills is likely to lead to a sustained reduction in inequalities, meaning that there is no equity-efficiency trade-off in early intervention (Doyle et al., 2009). In fact, given the already observable gap in cognitive and non-cognitive abilities between 5-year-old children from different socio-economic backgrounds, childhood intervention is likely to be the most effective strategy. Evaluations of programs conducted with adolescent students strongly suggest that although the interventions tend to be effective in improving the educational achievement of girls, they seem to have no effect for boys (Angrist et al., 2002; Angrist and Lavy, 2009; Angrist et al., 2009; Rodriguez-Planas, 2012).

This greater malleability of female high school student behaviour relative to male behaviour and its amenability to short-term intervention can be at least partly explained by persistent gender differences in economic preferences or non-cognitive skills developed and reinforced as individuals progress through their developmental stages. The economics literature, for example, provides strong evidence for a gender bias in risk attitudes, time preferences, preferences for competitive environments and self-confidence (Croson and Gneezy, 2009). A major factor in explaining gender differences in educational achievement and behavioural responses to remedial intervention is the gender gap in patience; in particular, the ability to delay rewards (Shoda et al., 1990). Other key predictors of educational achievement include the ability to set goals for educational performance and establish and adhere to a work schedule for goal achievement. For instance, there is evidence that female primary school students are more patient and more able to delay rewards than male students (Bettinger and Slonim, 2007; Castillo et al., 2011) and that girls apply self-regulated learning strategies that involve goal setting and planning more frequently than boys (Zimmerman and Martinez-Pons, 1990). Education outcomes are also impacted by gender differences in classroom behaviour, which are partly associated with differences in maturity between boys and girls of similar age. Evidence that women react more strongly to emotions than men (Croson and Gneezy, 2009) also implies that female students might be more averse to negative feedback on their educational performance and thus exert more effort to avoid negative outcomes. The gender gap may also be widened by different teacher expectations for the educational performance of female versus male students, not only through subjective grading (Cornwell et al., 2013) but also via the impact on the students' own expectations for their achievement potential.

The remainder of the paper is structured as follows. Section 2 reviews the literature on the impact of programs designed to improve students' educational achievement. Section 3 describes the intervention, the analytical data and our empirical strategy. Section 4 presents our main findings, and Section 5 concludes the paper.

2 Evidence on incentive programs for educational achievement

The literature on incentive programs aimed at improving educational achievement tends to focus heavily on two aspects: cash transfers and in-kind rewards conditional on the achievement of a specific goal (Angrist et al., 2002; Angrist and Lavy, 2009; Bettinger, 2011; Schultz, 2004; Rodriguez-Planas, 2012) and sanctions in case of non-compliance with a minimum attendance rate (Dee, 2011; Jones et al., 2002). Extant research also devotes attention to the impact of providing information to students and parents on the importance of educational achievement (Avvisati et al., 2014) or the effect of combining different types of interventions (Angrist et al., 2009). Methodologically, many of these studies examine the impact of incentive programs in an experimental context, using random assignment of students to treatment and comparison groups. Table A.2 in appendix lists 11 of these studies. Students assigned to the treatment group are exposed to the intervention at test, while students assigned to the comparison group are exposed to the same conditions as the treatment group would have been exposed to in the absence of the treatment (counterfactual). The validity of the causal inferences made by comparing the post-intervention outcome means of treatment versus control groups, however, relies heavily on individuals from the subject pool being randomly assigned to the treatment and control groups. That is, if individuals are not self-selected into the treatment group and have similar relevant (i.e., potentially outcome correlated) pre-treatment observable characteristics as individuals in the comparison group, the post-treatment differences in the outcome of interest between the treatment and control group can be validly attributed to the intervention being tested.

In the extant literature, randomized evaluations of cash incentive programs tend to find heterogeneous program effects across sub-groups. In particular, they show that incentives are more effective for female than male students, except among primary school children (Bettinger, 2011). For example, Angrist et al. (2002), show that Colombia's PACES program (*Programa de Ampliación de Cobertura de la Educación Secundaria*), which randomly assigns vouchers covering some costs of private secondary school to children from the lowest socio-economic strata, reduces the probability of grade repetition for female students more than male students, increasing test scores by 0.2 standard deviations from the mean. A similar positive effect for girls was observed in Angrist and Lavy's (2009) study of an experimental Israeli Achievement Awards program that provided cash awards for successful completion of high school exit exams to low-achieving high school students. The program allocation was randomized among 40 high schools selected by the Ministry of Education for their low success rates on high school exit exams. The cash incentives were high: a student who passed all tests received an amount equivalent to just under US$2,400. Although simple comparison of treatment and control group means identified no program effect on the probability of successful exit exam completion, once controlling for unobservable school characteristics and focussing on students close to the certification threshold, they found a 0.1 increase in certification probability for girls only.

Likewise, based on a randomized evaluation in 11 high schools of the 5-year US Quantum Opportunity Program, which offered mentoring, educational services and financial rewards to low-performing high school students, Rodriguez-Planas (2012) provides evidence of a heterogeneous intention-to-treat effect between male and female students, albeit with no adjustments made for ongoing participant involvement or program exposure intensity. Specifically, the demonstration program had large positive effects on educational outcomes for female students with persistent effects on employment outcomes. It increased female students' probability of graduating from high school by 15 percent and of pursuing post-secondary education by 20 percent. The author found no evidence, however, of a positive effect on male students' educational outcomes, although some evidence did emerge of increased engagement in risky behaviour by male students 5 years after the program ended. In addition, the program design prevented separate identification of each component's impact (mentoring, educational services and rewards) or any assessment of whether the effects were driven by the combination of different measures.

In terms of programs focused exclusively on females, Kremer et al. (2009) show that scholarships (covering 2 years of school fees and expenses) offered to grade 6 girls in the top 15 percent of achievers in 34 randomly selected Kenyan primary schools, coupled with public recognition at an assembly of students, parents, teachers and school officials, did increase test scores. They also identify positive externalities on low-achieving students unlikely to win the scholarship. Nevertheless, as in Rodriguez-Planas (2012), the experiment does not allow the scholarship effects to be disentangled from the impact of public recognition.

As regards school attendance, Schultz (2004) study of Mexico's Progresa program, which provides grants to poor mothers in 314 communities randomly selected from 495 eligible communities conditional on their children's school participation, shows a 0.66 years increase in school attendance. The impact of such financial incentives, together with academic support services, is also the subject of Angrist et al.'s (2009) investigation of the academic performance of first-year students at a large Canadian university. All first-year students, except those in the upper quartile of the grade point average distribution, were randomly assigned to a treatment group (which received financial incentives, academic support services or both) or a control group. The results suggest that providing academic support services has a significant positive impact on performance, but only for women. Providing the combined intervention in the first year produced higher second-year performance in female students than in the rest of the student population. Bettinger (2011), however, finds no such gender differences among primary school students in the low socio-economic area of Coshocton (Ohio) who were offered a cash incentive for academic performance. In this program, 8 of the 16 eligible grade-school combinations were selected by lottery in each academic year (from 2004-2005 to 2006-2007), with all children in the treatment grades in each school being eligible for the incentive program that year. Students in the treatment grades received US$15 for each test on which they scored proficiently and US$20 for each on which they received an advanced score. The results show positive program effects for math scores (an increase of 0.15 standard deviations from the mean), but only for students at the top of the test score distribution.

One shortcoming of the above studies suggesting that incentive programs improve educational outcomes for female students is that they target students from disadvantaged socio-economic backgrounds whose families are more likely to be cash constrained. These students may therefore be more likely than more advantaged students to positively respond to a cash incentive program. The more advantaged, in contrast, might already have benefitted from explicit or implicit reward mechanisms conditional on their academic achievement, meaning that cash incentive programs would have no impact on their behaviour and academic outcomes. In addition, because students from low socio-economic backgrounds tend to have poorer educational achievement than students from higher socio-economic backgrounds, they are more likely to have ex-ante achievement levels below the threshold fixed by the programs. These interventions are therefore designed to influence the behaviour of students who are below threshold levels and more socio-economically disadvantaged than those above the threshold. For example, Henry and Rubenstein (2002) find evidence that it was students from low socio-economic backgrounds that benefitted most from the HOPE scholarship program in Georgia, which provided merit-based financial aid to high school students independent of family income level. They also emphasize, however, that the program had a positive effect on student achievement and helped reduce inequalities between Whites and African Americans.

Team incentives like micro-credit designs in developing countries, have proven very effective in improving student learning, most probably because of peer effects and peer monitoring. That is, if students see their peers achieving high performance in school, they are likely to increase their own aspiration levels and set higher achievement goals. Groups can thus serve as a coordination device to break poverty traps (Ray, 2006). Specifically, when group members coordinate to reach a common goal, the group support and monitoring of each member's compliant behaviour provide all members greater incentives to work towards goal achievement. Such monitoring and ensuing social sanctions for non-compliance reduce incentives for free riding.

In the school context, peer effects may be strengthened by the presence of team incentives (Blimpo, 2010), although when coordination and monitoring costs are too high, such incentives can be ineffective. Hence, to assess the role of individual versus group monetary incentives in increasing schooling outcomes, Blimpo (2010) conducts an experiment in which 100 secondary schools in Benin were randomly assigned to three treatment groups and one control group. In the first treatment, students were offered a monetary prize based on their individual performance on the secondary school certification examination. In the second and third treatments, they were matched into groups of four and either received a monetary prize based on the group's average exam performance or participated in a tournament in which only the three teams with the highest average performance won a prize. Each of the three treatments significantly increased student test scores by 0.29, 0.27 and 0.34 standard deviations, respectively.

Methodologically, the best practice for ensuring validity of casual inferences is random assignment of the intervention at test between observational units, individuals or groups. However, for ethical and/or practical reasons, random treatment assignment is not always feasible. Moreover, many programs have been implemented without a design planification that allows ex-ante definition of a rigorous counterfactual. Hence, evaluating these programs requires the application of quasi-experimental methods that allow ex-post counterfactual construction, such as differences-in-differences and matching methods. Jackson (2010), for example, evaluates the Advanced Placement Incentive Program in Texas, in which participation, rather than being randomly assigned, was allocated to schools selected by funding donors from among all interested schools. There was also substantial variation in the date of program introduction in the different schools driven by donor availability and preferences. These private donors (who covered between 60 and 75 percent of the total program cost) also decided on the amount of the cash incentives awarded to both high school students (grades 11 and 12) and their high school teachers for each passing or above score in an eligible subject on the Advanced Placement (AP) exams. These incentives were very high: between US$100 and 500 for teachers and discretionary bonuses of up to US$5,000, and between US$100 and 500 for students together with reduced examination fees. Fifty-five schools adopted the program between 1995 and 2007, and an additional 6 had adopted it by 2008.

To exploit the phased-in implementation of the project in interested high schools, Jackson (2010) employs a differences-in-differences strategy in which he compares the difference in aggregate (school level) outcomes pre and post program between schools involved in the program and schools with the same pre-treatment test scores that adopted it at a later stage. Specifically, he estimates the intention-to-treat, defined as the program's impact on students enrolled in grade 10 in a school that would have been treated in the subsequent 2 years after being asked to participate in the program. The results indicate that the intervention not only increased the probability of a student enrolling in college and remaining in college beyond the first year but increased the first-year GPA. The author also identifies a stronger positive program effect for Black and Hispanic students.

Nevertheless, cash incentives conditional on achievement can also have discouraging effects on low-achieving students, as Leuven et al. (2010) show in an experiment that rewarded first-year university students conditional on their completion of all first-year subjects by the end of the following academic year. This experiment, conducted at the University of Amsterdam in the 2001-2002 academic year, randomly assigned 249 first-year student volunteers to two treatment groups (a large reward and a small reward) and a control group (no reward). Although the average treatment effects of the small and large rewards were not significantly different from zero, heterogeneous treatment effects emerged for high- versus low-ability students. Whereas high-ability students responded to the large reward by improving their performance in terms of pass rate and number of credit points earned, low-ability students responded by lowering their performance. The program also had lasting effects: the high-ability students in the treatment groups performed even better in their second and third years of study, whereas the low-ability students performed worse.

Other studies, however, find no effect of cash incentives on academic achievement. For example, in a randomized controlled experiment conducted in schools in three different US districts - Dallas, New York and Chicago - Fryer (2011) provided varying cash incentives to students conditional on achievement. In Dallas, second grade students from 21 public schools randomly selected from those that applied to be part of the program received monetary incentives to read books (US$2 per book). In New York, students from 63 similarly selected schools were rewarded for their performance on reading and math exams, with fourth grade students receiving US$5 for each test completed and US$25 for each perfect score. The incentives were doubled for seventh grade students. In Chicago, students from 40 schools were given incentives for their grades in five core courses: US$50 for an A, US$35 for a B and US$20 for a C. Half of the students were rewarded immediately and the other half at graduation. The participating schools in these three states were characterized by a very high share of Black and Hispanic students, as well as students from low socio-economic backgrounds. Overall, the results indicate that the incentives had no impact on either the direct outcomes for which students received the incentives or on self-reported effort. A similar lack of impact was reported by Angrist et al. (2010) for the Opportunity Knocks program, which randomly offered financial rewards and peer mentoring for first- and second-year financial aid applicants at a Canadian university in Ontario conditional on grades.

Other randomized evaluations look particularly at the impact of in-kind transfers conditional on schooling outcomes. For instance, Vermeersch and Kremer (2005) assessment of the impact of providing school meals in 25 randomly selected Kenyan pre-schools shows large positive program effects on school enrolment and attendance (a 30 percent increase in participation in treated schools). Another study by Berry (2009) not only tests whether monetary and in-kind rewards generate different incentive effects but also whether the incentive effects vary with reward recipient (child or parents). This experiment, conducted in 8 primary schools in India, randomly allocated five different treatments at the student level. In the first and second treatments, the children were rewarded for reaching a target test score level with either a toy or money, respectively. In the third, the money reward was given to the parents. In the fourth and fifth treatments, the parents were given the choice either ex-ante or ex-post, to reward their child with a toy or keep the money themselves. The value of the reward, either toy or money, was identical across treatments. The results show that although the incentives did increase the children's test scores overall, the treatments resulted in no significant differences. Evidence did emerge, however, of heterogeneous treatment effects based on pre-test scores. Specifically, relative to high achievers, children with lower achievement levels benefitted more from the toy treatment than from the money treatment. On the other hand, children with better socio-economic conditions benefitted more from the money treatment than the toy treatment.

There is also evidence that establishing sanctions for non-compliance with a minimum rate of school attendance is effective in increasing attendance rates. Jones et al. (2002), for instance, find that providing orientation and assistance programs to secondary school students while imposing sanctions (lost eligibility for social benefits) for non-compliance with an 80 percent school attendance target increases the probability that students will meet the attendance target. Again, however, the experimental design allows no disentanglement of the sanction effects versus the learning and orientation effect. Nevertheless, according to Dee (2011), the Wisconsin Learnfare program, a six semester fee waiver reform in nine counties that sanctioned a family's welfare grant when covered teens (between 13 and 19 years) failed to meet school attendance and completion targets, increased school enrolment by 3.5 percent and school attendance by 4.5 percent.

With only a few exceptions (Angrist et al., 2010; Fryer, 2011; Leuven et al., 2010), the literature shows cash and in-kind rewards conditional on educational attainment to be effective. This finding, however, may be driven partly by publication bias (i.e., studies showing significant program effects are more likely to be published than those finding no impacts). Moreover, no evidence exists on the long-term effects of such interventions. If the most important factor in explaining educational inequalities between children from different socio-economic backgrounds is family environment (Doyle et al., 2009), it is questionable whether short-term interventions providing monetary or in-kind incentives will have persistent effects on the students' educational trajectories. This type of intervention, particularly when targeted at parents, assumes that the poor educational achievement of children from disadvantaged socio-economic backgrounds can be primarily explained by financial constraints. Yet the information asymmetries experienced by families from different socio-economic backgrounds regarding the benefits of education are also likely to play a major role in explaining inequalities in educational achievement. Indeed, some studies clearly show that providing information on the benefits of academic achievement has a positive impact on educational outcomes. Nguyen (2008), for instance, assesses the impact on school achievement of providing the parents of children in 640 Madagascan primary schools with information on returns to schooling. Specifically, the parents either received statistical information on school returns, met with role models who gave information about their own backgrounds, educational achievement and current employment situation or were exposed to a combination of both. Although the results show no impact of the role model intervention on school attendance and test scores, they identify a positive impact of providing statistical information on both outcome variables. Similarly, Avvisati et al. (2014) evaluated the effect of providing information sessions for interested parents of sixth grade students in 34 volunteer middle schools in deprived Paris suburbs on assisting and encouraging their children to expend effort in school and practical guidance for doing so. Not only did this intervention have a positive effect on school attendance, it improved behavioural outcomes for the students whose parents took part in the program. There was, however, no effect on test scores, although there were positive spill-over effects (of lower magnitude) for the students whose parents did not participate in the program but whose classmates' parents were involved.

3 Program design, empirical methodology and data

3.1 The FOGS ARTIE program

The Former Origin Greats (FOGS) is a non-profit organization created in 1997 in the state of Queensland by former elite players from the rugby leagues. One of its main programs is the FOGS ARTIE program (Achieving Results Through Indigenous Education), funded partly by the Australian government, which focuses on the educational outcomes of Indigenous students. It began in 2010 as a pilot project in 8 selected public high schools located in both metropolitan and provincial areas in Queensland. In 2011, the program added another 13 high schools for a total of 21 schools, 13 in metropolitan areas, 6 in regional areas and 2 in areas designated as very remote. Although FOGS originally invited many more than 13 schools to participate in the program based on their large Indigenous enrolment, budget restrictions limited the selection. In 2010, because of lack of funding and clear planification, the program's scope was very limited and offered mostly individual support to Indigenous students identified by their teachers as at risk of repeating grades. The very limited number of available tutors, however, meant that such help could not be offered to all Indigenous students in need of it. In 2011, FOGS expanded the tutoring program to 13 additional schools, in which tutoring was again offered to students at risk of grade repetition but with more regular sessions and five times the number of tutors. In 2012, FOGS introduced a broader program that specifically targeted school attendance, grades in math and English and the behaviour of all Indigenous students enrolled in grades 8 to 12 in the participating schools. Specifically, the program provided information on the benefits of educational achievement, established specific goals and awarded prizes at the end of each term for students who reached these goals. In 2012, at the start of each term in each participating school, all Indigenous students were invited to a launch during which FOGS administrative staff members and former elite rugby players held an information session about the benefits of education and invited students to take part in specific challenges for the upcoming school term. These challenges took the form of an incentive mechanism involving end-of-term in-kind rewards conditional on students' attendance rates, academic grades and behaviour. The objectives fixed for each term differed across the participating schools (see Table 3.1). In 2012, the program objectives and prizes offered were identical in 11 out of the 21 participating schools. In term 1, conditional on having obtained at least a passing score for effort and behaviour in both math and English classes, students could choose between two tickets to the cinema or a night out at football. In term 2, conditional on having obtained at least a passing score in math and English, students were offered sports clothes with the Artie logo. In terms 3 and 4, conditional on having reached an attendance rate of at least 90 percent, students were again offered sports clothes with the Artie logo. In 8 other schools, students defined individual goals at the beginning of each term and received prizes at the end of the term based on whether they had achieved their goal. In the 2 remaining schools, the goal for each term was to achieve an attendance level of at least 90 percent. Although we have no information on program take-up (i.e., how many students took part in the launches and were informed about the FOGS program), all Indigenous students were invited to participate, so it is likely that even those who did not attend were aware of the program through their classmates. We also have no

Table 1: Distribution of schools between different goals by term in 2012

	Term			
Goal	1	2	3	4
Passing score for behaviour in math and English	11			
Passing score in math and English		11		
At least 90% of attendance	2	2	13	13
Individuals goals	8	8	8	8
No. of schools	21	21	21	21

data on the students' level of engagement with the program and whether they were fully aware of the challenges specified for each term.

3.2 Empirical strategy

Program placement in the participating schools, rather than being random, depended on several factors, not only the number of Indigenous students enrolled in the school but also whether the school principal knew about the program and was willing to take part in it. This non-random assignment inherently prevents us from constructing a valid ex-ante counterfactual to estimate causal effects. Rather, we build a control group by matching each school involved in the program with non-participant schools that have a very close probability of program involvement. We estimate this latter for all Queensland high schools using a set of school-level relevant characteristics observed prior to the 2009 start of the intervention: geographic location, total number of students, total number of Indigenous students, proportion of Indigenous students, socio-economic indicators and attendance rates.[1] This matching method identifies a control group similar in its characteristics to the treatment group and thus more likely to respond in kind to underlying trends or contemporaneous shocks (Eissa and Liebman, 1996). Appendix Table A.1, which presents the results of the corresponding probit regression, identifies the number of Indigenous students enrolled and the school's geographic location as the relevant variables for propensity for program involvement, which is consistent with FOGS's use of Indigenous student enrolment as its main selection criterion.

Matching on observable characteristics, however, does not account for differences in unobservable characteristics between control schools and those that received the intervention (treatment schools). That is, if schools whose administrative staff were more concerned about Indigenous student outcomes were more likely to be part of the pro-

[1] We estimate a probit model for a school's propensity for treatment group assignment using data for all Queensland high schools. We regress an indicator variable for whether the school was part of the program on a set of observable relevant characteristics and then match each treatment school with a close probability non-treatment school. Details of our matching method are given in the appendix.

gram, the estimates would confound not only the program effect but also unobservable school characteristics capable of impacting educational outcomes. We therefore attenuate the selection on unobservable characteristics by including in our control group schools willing to be part of the program but ineligible because of insufficient Indigenous enrolment or funding restrictions whose probability of program involvement is closest to that of the treatment schools. This criterion applies to 14 of the 24 schools in the control group. We also believe it plausible that these schools may have similar unobservable characteristics to those involved in the intervention.

To account for unobservable heterogeneity between treatment and control schools that could influence students' academic achievement and behaviour, we apply a differences-in-differences strategy that enables comparisons of the pre and post-treatment outcomes progression for Indigenous students in program schools with that of Indigenous students in control schools over the same period. Because of data limitations, we estimate the intention-to-treat (ITT) - that is, the impact of offering the program on educational outcomes - without taking into account whether students attended the launches or were aware of the project or their level of involvement if they participated. We do so using the following model:

$$Y_{ift} = \alpha + \gamma_f + \lambda_t + \beta(f.t) + X'_{ift}\delta + \epsilon_{ift} \tag{1}$$

where γ_f is an indicator variable equal to 1 if the student is enrolled in a program school and 0 if enrolled in a control school, λ_t is a time period dummy equal to 1 if the outcomes are observed post-intervention (in 2012) and 0 if observed pre-intervention (in 2009), and X represents a vector of individual and school level control variables. The standard errors are clustered at the school level, allowing for outcome correlations between students enrolled in the same school, a likely occurrence given that they share the same school environment. We estimate equation (1) for the pooled sample of Indigenous students and also separately by gender. The causal impact of the program is given by parameter β, the coefficient of the interaction term between the treatment and year dummy variables:

$$\begin{aligned}\beta = [E(y_{ift} \mid f = 1, t = 1) - E(y_{ift} \mid f = 1, t = 0)] - \\ [E(y_{ift} \mid f = 0, t = 1) - E(y_{ift} \mid f = 0, t = 0)]\end{aligned}$$

Within each sub-sample (treatment and control), the difference in outcomes post and pre-intervention controls for specific observed and unobserved school effects, thereby eliminating common time effects across treatment and comparison schools. It accounts, for instance, for potential changes in national or state level education policies, which could impact student outcomes and be confounded with program effects.

The main identification assumption of the differences-in-differences model is that time trends are common across treatment and control groups (Angrist and Pischke, 2008). Hence, in the absence of the treatment, educational outcomes in the treated schools would have evolved just as in control schools. This hypothesis can be tested by analysing time trends over the same period for different groups or by looking at the time trends for control and treated groups over a larger number of pre-intervention periods. Given that the program was offered exclusively to Indigenous students in the treatment schools, we investigate whether there is a common time trend in the progression of educational outcomes for non-Indigenous students in both the treatment and control schools in the 2009 and 2012 periods. The main limitation is that non-Indigenous students might also be considered treated if there are externalities from the intervention. Given that in most schools Indigenous students represent less than 8 percent of the student population and not all Indigenous students were engaged in the program, however, we can plausibly assume that any potential program externalities affecting non-Indigenous students will be very small.

3.3 Data

3.3.1 Data and descriptive statistics

We use data from the Queensland Department of Education Training and Employment (DETE) on all students, Indigenous and non-Indigenous, enrolled in schools that took part in the FOGS Artie program and schools identified as control schools. The data cover the period from 2009 until 2012 for both treatment and control schools and include individual level data on whether students are identified as Indigenous (determined by parental declarations at the start of the academic year[2]), as well as gender, year level, date of birth, parental education and postal code. We base student academic outcomes on individual level data by semester on students' academic scores for math, English and science. We supplement this information with ninth grade student scores on the Naplan test, a nationwide assessment of core literacy and numeracy skills for students in years 3, 5, 7 and 9 conducted annually for purposes of national comparison. We also have individual level data on student absences and whether they were explained or unexplained. Finally, we have data on students' behavioural scores by subject and

[2]The Australian Bureau of Statistics defines an Indigenous person as a person of Aboriginal or Torres Strait Islander descent who self-identifies as an Indigenous and is accepted as such by the community in which he or she lives. No proof of aboriginality is required, however, to declare a child Indigenous in schools, although it might be requested if the family applies for special financial support or assistance programs reserved for Indigenous people.

semester derived from the end-of-semester grades awarded for overall performance and behaviour and effort in each subject. Our sample includes a total of 3,059 Indigenous and 37,493 non-Indigenous students distributed across 21 treatment and 24 control schools, whose 2009 socio-demographic statistics are given in Table 2. As the table shows, the number of Indigenous students decreases with the progression in high school year level, only 10 percent of the Indigenous students in our sample are enrolled in grade 12 versus 15 percent of non-Indigenous students. Additionally, approximately one third of Indigenous students in the sample are from a single-parent family versus 18 percent of non-Indigenous students, and the parents of the former have lower education levels than parents of the latter.

3.3.2 Pre-treatment characteristics of treatment and control schools

We first check for possible differences in 2009 pre-treatment characteristics between treatment and control schools but find no pre-treatment differences in school level or socio-demographic characteristics (see Table 3). This outcome was expected given that we selected the control schools based on a set of school-level characteristics.

Table 3: Differences between treatment and control groups in pre-treatment characteristics

	% of Indigenous	Total enrolment	Indigenous enrolment	Male	Mother's education	Father's education
Treatment	-0.001	76.454	5.223	0.009	0.102	0.077
	(0.011)	(96.100)	(11.518)	(0.006)	(0.137)	(0.134)
Constant	0.076***	1040.647***	73.537***	0.502***	2.712***	2.597***
	(0.009)	(68.752)	(9.528)	(0.003)	(0.102)	(0.104)
Observations	40552	40552	40552	40552	40541	32630
Schools	45	45	45	45	45	45

Standard errors clustered at the school level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We then test for pre-treatment differences between treatment and control schools in behaviour and academic scores (Tables 4 and 5, respectively). The results in Table 4 indicate no statistically significant differences in grades for behaviour for either the whole sample (panel A) or the sample of Indigenous students only (panel B). Likewise, we find no differences in pre-treatment academic outcomes for either Naplan scores or grades awarded by teachers (Tables 5 and 6). Finally, we identify no significant pre-treatment differences in either attendance rates or number of unexplained absences (Table 7).

Table 2: Socio-demographic characteristics of Indigenous students in 2009, by school treatment status

	Indigenous				Non-Indigenous			
	Control		Treatment		Control		Treatment	
	No.	Col %	No.	Col %	No.	Col %	No.	Col %
Year Level								
8	399	25	377	26	3974	21	3948	22
9	390	25	358	24	4190	22	3870	21
10	337	21	334	23	4347	23	4017	22
11	295	19	251	17	3746	19	3666	20
12	161	10	157	11	3008	16	2727	15
Total	1582	100	1477	100	19265	100	18228	100
Female	813	51	735	50	9559	50	8886	49
Male	769	49	742	50	9706	50	9342	51
Total	1582	100	1477	100	19265	100	18228	100
Biparental family	1067	67	938	64	15907	83	14718	81
Monoparental family	515	33	539	36	3358	17	3510	19
Total	1582	100	1477	100	19265	100	18228	100
Mother's education								
Not stated/unknown	549	35	519	35	7504	39	6194	34
Year 9 or below	151	10	176	12	977	5	1072	6
Year 10	429	27	409	28	4711	24	4949	27
Year 11	190	12	145	10	1631	8	1610	9
Year 12	263	17	227	15	4440	23	4395	24
Total	1582	100	1476	100	19263	100	18220	100
Father's education								
Not stated/unknown	413	39	366	39	6708	42	5467	37
Year 9 or below	105	10	104	11	853	5	1036	7
Year 10	292	27	268	29	3855	24	3996	27
Year 11	57	5	72	8	1078	7	1057	7
Year 12	200	19	128	14	3413	21	3162	21
Total	1067	100	938	100	15907	100	14718	100
Schools	24		21		24		21	

Table 4: Differences between treatment and control groups in pre-treatment scores for behaviour

	Behaviour					
	Semester 1			Semester 2		
	Math	English	Science	Math	English	Science
Panel A: Pooled sample						
Treatment	0.022	0.023	0.037	0.034	0.003	0.060
	(0.052)	(0.057)	(0.058)	(0.051)	(0.065)	(0.057)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27617	27244	18352	24231	23997	17595
Schools	45	45	45	45	45	45
Panel B: Indigenous						
Treatment	-0.072	0.026	-0.068	0.061	-0.130	0.040
	(0.108)	(0.083)	(0.095)	(0.101)	(0.089)	(0.105)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1616	1585	997	1439	1434	1010
Schools	45	45	45	45	45	45

Standard errors clustered at the school level; grades standardized to mean 0 and standard deviation 1. The control variables include gender, parents' education, proportion of Indigenous students and dummy variables for year level, plus a dummy for Indigenous students in the Panel A specification. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 5: Differences between treatment and control groups in pre-treatment academic grades

	Academic grades					
	Semester 1			Semester 2		
	Math	English	Science	Math	English	Science
Panel A: Pooled sample						
Treatment	-0.011	0.029	0.002	0.006	0.017	0.067
	(0.044)	(0.048)	(0.055)	(0.041)	(0.066)	(0.054)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28768	28209	19567	26105	25716	19048
Schools	45	45	45	45	45	45
Panel B: Indigenous						
Treatment	0.046	0.105	-0.009	0.071	0.018	0.052
	(0.073)	(0.069)	(0.087)	(0.072)	(0.091)	(0.078)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1678	1610	1077	1530	1496	1088
Schools	45	45	45	45	45	45

Standard errors clustered at the school level; grades standardized to mean 0 and standard deviation 1. The control variables include gender, parents' education, proportion of Indigenous students, and dummy variables for year level, plus a dummy for Indigenous students in the Panel A specification. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 6: Differences between treatment and control groups in Naplan scores

	Naplan scores			
	Writing	Spelling	Grammar	Numeracy
Panel A: Pooled sample				
Treatment	-0.004	0.003	-0.001	0.024
	(0.051)	(0.045)	(0.050)	(0.049)
Controls	Yes	Yes	Yes	Yes
Observations	6945	6945	6945	6940
Schools	45	45	45	45
Panel B: Indigenous				
Treatment	-0.149	-0.076	-0.028	-0.058
	(0.130)	(0.123)	(0.113)	(0.134)
Controls	Yes	Yes	Yes	Yes
Observations	476	476	476	478
Schools	45	45	45	45

Standard errors clustered at the school level; grades standardized to mean 0 and standard deviation 1. The control variables include gender, parents' education, proportion of indigenous students, and dummy variables for year level, plus a dummy for Indigenous students in the Panel A specification. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 7: Differences between treatment and control groups in pre-treatment attendance outcomes

	Attendance			
	Semester 1		Semester 2	
	Attendance rate	Unexplained absences	Attendance rate	Unexplained absences
Panel A: Pooled sample				
Treatment	0.169	0.286	0.363	0.381
	(0.752)	(0.831)	(0.870)	(0.774)
Controls	Yes	Yes	Yes	Yes
Observations	32629	32629	32629	32629
Schools	45	45	45	45
Panel B: Indigenous				
Treatment	1.026	0.127	-0.186	0.656
	(1.662)	(1.340)	(1.370)	(1.154)
Controls	Yes	Yes	Yes	Yes
Observations	2005	2005	2005	2005
Schools	45	45	45	45

Standard errors clustered at the school level. The control variables include gender, parents' education, proportion of indigenous students, and dummy variables for year level, plus a dummy for Indigenous students in the Panel A specification. *p < 0.10, **p < 0.05, ***p < 0.01.

4 Results

We arrange our discussion of the results for FOGS' impact on Indigenous students' educational outcomes around three aspects: behaviour scores (section 4.1), academic grades (section 4.2) and attendance (section 4.3). We also present separate results for male and female students because, like the experimental literature on incentives for educational achievement, we find that incentives are more effective for female than for male students. In addition, because our results suggest that the program is only effective for students from intact families, we limit our interpretation of the results to students in the sample who live with both parents,[3] which accounts for 66 percent of the total sample of Indigenous students and 82 percent of the total sample of non-Indigenous students. This decision is in line with the research evidence that family structure is a crucial variable in explaining educational outcomes. Conti et al. (2010), for instance, show that the family environment is the strongest predictor of students' educational achievement from the earliest years of school. One reason may be that single parents tend to be single mothers with low levels of education and more financial constraints than intact families. They

[3]The results for single parent households are available from the authors upon request.

may thus be less able or likely to offer a supportive environment for their children's educational achievement. When the family environment is highly detrimental to schooling outcomes, any external intervention during high school years will have very little power to enhance educational achievement by improving behaviour or changing study habits.

To check for similar time trends in outcomes between Indigenous and non-Indigenous students, we also report the treatment effects for non-Indigenous non-participants in the program, which also allows us to check whether Indigenous students in treatment schools have the same outcome progressions as those in control schools. As previously discussed, testing for time trends in non-Indigenous students is problematic because the effect of possible program externalities on non-Indigenous student outcomes could be confounded with program impact. However, because Indigenous students constitute such a small fraction of the student population (7 percent on average across schools), finding no significant treatment effect on non-Indigenous student outcomes would be evidence that any significant treatment effect likely captures a program effect.

It should also be noted that because of reporting errors, the sample size is not constant across either subjects (math, English and science) or the outcomes considered (behaviour and academic grades and attendance). These reporting errors, however, are likely to be random and so do not bias our estimates. In addition, the sample size differences across subjects (math, English and science) result primarily from the fact that high school students can choose their own curriculum, so not all students are enrolled in all three subjects. In particular, many students substitute electives for science classes, making students enrolled in science subjects a specific sample that tends to include high achievers. Hence, although we report the treatment effects for science, students were actually given no incentives targeting science so any positive impact on science grades should not be interpreted as an average effect of the program.

Next, in section 4.4, we describe the robustness checks conducted to assess the stability of our results, in particular, the checks that include school fixed effects. We also report the results obtained when we analyse the program impact on each outcome for all 21 schools involved in the program, not simply the schools in which students were incentivised to improve an outcome. We also discuss the treatment effect estimates obtained after restricting the comparison group to the sample of 14 schools that expressed their willingness to take part in the program. Finally, in section 5, we discuss potential behavioural mechanisms for the heterogeneity in treatment effects between male and female students.

4.1 Behavioural outcomes

In the first term of 2012, Indigenous students from 11 out of the 21 treatment schools were offered a prize at the end of the term conditional on obtaining at least a passing score for behaviour in both math and English classes. As discussed in the previous section, although the students were not incentivized to improve their effort and behaviour in science classes, we also report the treatment effect for this component. Our data include student grades by semester but not by term. The regression results for the pooled sample of Indigenous students, considering the 11 schools with behaviour and effort incentives as the treatment group, are given in Table 8, panel I.A, and separately by gender in panels I.B and I.C. These results show a positive program effect on student behaviour in math classes, significant at the 10 percent level, as well as mild evidence (of lower magnitude) for a positive effect on student behaviour in math classes in semester 2. This positive effect is driven by female students. For male and female students separately (panels I.B and I.C), we find no significant treatment effects on behaviour for male students but a significant positive effect for female students in both math and science in semester 1 (but not in semester 2). Relative to the 2009 scores, the program is expected to increase female students' behaviour scores in math by 0.28 standard deviations from the mean (significant at the 10 percent level) and result in female students in treatment schools scoring 0.27 standard deviations from the mean higher on behaviour in science classes than female students in control schools (also significant at the 10 percent level). Table 8 also reports the treatment effects for non-Indigenous students, both for the pooled sample (panel II.A) and separately by gender (panels II.B and Panel II.C). The results indicate no significant treatment effect on these students' behavioural outcomes, which supports that we are measuring the program effect on the behavioural outcomes of Indigenous students.

Table 8: OLS estimates of the incentive effects on behaviour

	Behaviour					
	Semester 1			Semester 2		
	Math	English	Science	Math	English	Science
			I. Indigenous			
Panel I.A: Pooled sample						
Program x Year 2012	0.188*	0.076	0.225	0.131*	0.168	0.192
	(0.105)	(0.125)	(0.146)	(0.077)	(0.138)	(0.120)
Observations	2592	2561	1734	2280	2287	1690
Panel I.B: Male						
Program x Year 2012	0.086	0.024	0.178	0.082	0.158	0.217
	(0.150)	(0.194)	(0.213)	(0.138)	(0.189)	(0.169)
Observations	1283	1275	864	1153	1160	859
Panel I.C: Female						
Program x Year 2012	0.284*	0.141	0.272*	0.182	0.176	0.176
	(0.149)	(0.100)	(0.160)	(0.165)	(0.154)	(0.182)
Observations	1309	1286	870	1127	1127	831
			II. Non-Indigenous			
Panel II.A: Pooled sample						
Program x Year 2012	-0.014	-0.037	-0.043	0.022	-0.061	-0.016
	(0.055)	(0.079)	(0.068)	(0.052)	(0.088)	(0.068)
Observations	41171	40901	28639	35695	35431	26992
Panel II.B: Male						
Program x Year 2012	0.009	0.007	-0.044	0.037	-0.031	-0.009
	(0.063)	(0.084)	(0.088)	(0.061)	(0.084)	(0.076)
Observations	21005	20854	14534	18252	18111	13750
Panel II.C: Female						
Program x Year 2012	-0.037	-0.083	-0.039	0.007	-0.094	-0.023
	(0.053)	(0.079)	(0.062)	(0.052)	(0.098)	(0.070)
Observations	20166	20047	14105	17443	17320	13242
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Schools	35	35	35	35	35	35

Standard errors clustered at the school level; grades standardized to mean 0 and standard deviation 1. The control variables include a year fixed effect, an indicator variable for whether the student is enrolled in a treated school, parents' education, the proportion of indigenous students enrolled at the school, and year level fixed effects. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

4.2 Academic grades

In the second term, Indigenous students were offered a prize conditional on obtaining at least a passing grade in both math and English. Table 9 shows the regression results for semesters 1 and 2. Although the results in panel I.A for the pooled sample of Indigenous students indicate no incentive effect on grades for semester 1, they do show a positive effect in semester 2 on both math and English scores. According to panels I.B and I.C, this positive impact on academic grades is driven by female students. Specifically, the incentive increased female students' semester 1 grades in English by 0.23 standard deviations from the mean (significant at the 5 percent level) and their semester 2 grades in both math and English by 0.28 and 0.34 standard deviations from the mean, respectively. The effect for science is also positive for both female and male students, but not statistically significant. Panel II of Table 9, however, shows no impact of the program on the outcomes of non-Indigenous students.

We also analyse whether the program impacts student outcomes on the Naplan test of the four core skills of writing, spelling, grammar and numeracy, which is conducted at the schools but externally marked. Although the students in grades 3, 5, 7 and 9 who took the test at the end of the first semester were offered no rewards conditional on their test achievement and their test performance did not impact their academic grades, it seems plausible to assume that an improvement in behaviour and academic grades would also positively influence the ninth graders' Naplan scores. In fact, as Table 10 shows, male Indigenous students enrolled in schools involved in the program performed significantly better in all three literacy components of the test (panel I.B), improving their test scores by [0.35; 0.64] standard deviations from the mean. However, we find no positive effect on female students' test performance (panel I.C) and no evidence of differences between treatment and control schools in non-Indigenous students' score progression (panel II).

Table 9: OLS estimates of the inventive effect on academic grades

	Academic grades					
	Semester 1			Semester 2		
	Math	English	Science	Math	English	Science
			I. Indigenous			
Panel I.A: Pooled sample						
Program x Year 2012	0.065	0.042	0.149	0.178*	0.251**	0.180
	(0.112)	(0.113)	(0.111)	(0.088)	(0.112)	(0.122)
Observations	2718	2656	1889	2438	2420	1845
Panel I.B: Male						
Program x Year 2012	0.030	-0.142	0.075	0.063	0.165	0.220
	(0.140)	(0.146)	(0.131)	(0.116)	(0.147)	(0.146)
Observations	1337	1309	928	1219	1215	925
Panel I.C: Female						
Program x Year 2012	0.090	0.231**	0.220	0.280*	0.340***	0.140
	(0.167)	(0.113)	(0.132)	(0.141)	(0.123)	(0.147)
Observations	1381	1347	961	1219	1205	920
			II. Non-Indigenous			
Panel II.A: Pooled sample						
Program x Year 2012	0.005	-0.022	-0.030	0.008	-0.049	-0.064
	(0.064)	(0.052)	(0.053)	(0.062)	(0.070)	(0.069)
Observations	43564	43091	31066	38669	38346	29655
Panel II.B: Male						
Program x Year 2012	0.017	-0.016	-0.038	0.035	-0.026	-0.063
	(0.067)	(0.064)	(0.064)	(0.068)	(0.077)	(0.072)
Observations	22201	21933	15762	19709	19526	15070
Panel II.C: Female						
Program x Year 2012	-0.007	-0.028	-0.021	-0.020	-0.072	-0.067
	(0.073)	(0.049)	(0.065)	(0.065)	(0.071)	(0.079)
Observations	21363	21158	15304	18960	18820	14585
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Schools	35	35	35	35	35	35

Standard errors clustered at the school level; grades standardized to mean 0 and standard deviation 1. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 10: OLS estimates of effects of the program on Naplan scores

	Naplan scores			
	Writing	Spelling	Grammar	Numeracy
	I. Indigenous			
Panel I.A: Pooled sample				
Program x Year 2012	0.302*	0.113	0.130	0.155
	(0.160)	(0.137)	(0.132)	(0.178)
Observations	972	972	972	974
Panel I.B: Male				
Program x Year 2012	0.635***	0.419*	0.351*	0.357
	(0.227)	(0.219)	(0.202)	(0.250)
Observations	527	527	527	528
Panel I.C: Female				
Program x Year 2012	-0.083	-0.241	-0.125	-0.093
	(0.184)	(0.182)	(0.186)	(0.247)
Observations	445	445	445	446
	II. Non-Indigenous			
Panel II.A: Pooled sample				
Program x Year 2012	0.020	0.005	0.036	-0.013
	(0.066)	(0.060)	(0.063)	(0.067)
Observations	12367	12367	12367	12360
Panel II.B: Male				
Program x Year 2012	-0.000	-0.007	0.021	-0.030
	(0.063)	(0.063)	(0.065)	(0.073)
Observations	6496	6496	6496	6491
Panel II.C: Female				
Program x Year 2012	0.042	0.019	0.053	0.006
	(0.080)	(0.071)	(0.076)	(0.084)
Observations	5871	5871	5871	5869
Controls	Yes	Yes	Yes	Yes
Schools	45	45	45	45

Standard errors clustered at the school level; Naplan scores standardized to mean 0 and standard deviation 1. *p < 0.10, **p < 0.05, ***p < 0.01.

4.3 Attendance

In the second semester of 2012, Indigenous students from 13 out of the 21 schools involved in the program were given incentives for regular school attendance, with Indigenous students receiving a prize at the end of each term conditional on reaching an attendance rate of at least 90 percent. As shown in Table 11, panel I, the program was most effective for female participants, being expected to decrease the number of their unexplained absences by 4.25 days during semester 2 (panel I.C). However, we find no such significant program effect for male students, although the results for the sample of non-Indigenous students (panel II) provide mild evidence (significant at 10 percent) that in semester 1 only, male non-Indigenous students have a lower number of unexplained absences in treated schools than in control schools.

Table 11: OLS estimates of the program effects on attendance

	Attendance			
	Semester 1		Semester 2	
	Attendance rate	Unexplained absences	Attendance rate	Unexplained absences
	I. Indigenous			
Panel I.A: Pooled sample				
Program x Year 2012	-0.662	-2.532	0.308	-2.469
	(2.435)	(1.950)	(2.017)	(1.557)
Observations	3390	3390	3390	3390
Panel I.B: Male				
Program x Year 2012	-0.005	-2.740	-0.017	-0.907
	(2.643)	(2.267)	(2.523)	(2.014)
Observations	1713	1713	1713	1713
Panel I.C: Female				
Program x Year 2012	-1.291	-2.344	0.766	-4.250**
	(2.790)	(1.865)	(2.167)	(1.627)
Observations	1677	1677	1677	1677
	II. Non-Indigenous			
Panel II.A: Pooled sample				
Program x Year 2012	0.070	-2.209*	-1.046	-0.736
	(1.377)	(1.264)	(1.361)	(1.247)
Observations	50365	50365	50365	50365
Panel II.B: Male				
Program x Year 2012	0.467	-2.488*	-0.697	-1.183
	(1.357)	(1.326)	(1.297)	(1.173)
Observations	25684	25684	25684	25684
Panel II.C: Female				
Program x Year 2012	-0.345	-1.911	-1.405	-0.260
	(1.458)	(1.233)	(1.492)	(1.362)
Observations	24681	24681	24681	24681
Controls	Yes	Yes	Yes	Yes
Schools	37	37	37	37

Standard errors clustered at the school level; *p < 0.10, **p < 0.05, ***p < 0.01.

4.4 Robustness analysis

In the first of our three different robustness checks, we estimate all the regressions using all 21 of the schools involved in the FOGS Artie program as treatment schools (i.e., not simply those in which Indigenous students were given incentives to improve the outcomes under study). These results indicate a smaller program effect that in most cases loses its statistical significance. Only the positive treatment effect on behaviour scores in English classes in semester 1 and the negative treatment effect on the number of unexplained absences in semester 2 for female students remain significant (at the 10 and 5 percent levels, respectively). These findings clearly suggest that (female) students adapt their behaviour in response to the incentive.

We also estimate the program effects using school fixed effects, showing overall that the program effects reported in the previous section remain robust. In particular, the positive impact of the incentives on female students' behaviour scores in math remains significant for both semesters, and the magnitude of the effect is larger. In terms of academic grades, the positive effect on math scores for female students is significant at the 5 percent level and larger; however, the positive impact on English grades is smaller and loses its statistical significance. The positive effect of the incentives in reducing the number of unexplained absences for female students, however, remains significant at the 5 percent level (although slightly smaller in magnitude). With respect to the Naplan scores, we observe a positive significant effect for male Indigenous students in literacy.

Finally, we estimate the program effects using as our control group only the 14 (out of 21) schools that expressed interest in being part of the program, which are likely to have similar unobservable characteristics. This similarity might be important in that the teaching and administrative staff's motivation to reduce disparities in Indigenous versus non-Indigenous students' educational outcomes could lead to improved outcomes for Indigenous students even in the absence of the intervention. These results are consistent with those in the previous section, although the magnitude and significance level of the treatment effect estimates are slightly larger. As before, however, we find no significant effects for non-Indigenous students other than on the number of unexplained absences of male students in semester 1.

4.5 Potential mechanisms for the heterogeneity of the program effect

The findings presented previously highlight heterogeneous program effects for male and female students; specifically, an improvement in female Indigenous students' academic grades and behaviour and a reduction in the number of their unexplained absences. They show no positive impact, however, on these students' standardized test scores (Naplan). In contrast, although the program seems to improve male Indigenous students' scores on (objective) standardized tests, it does not improve their teacher-assessed grades for academic achievement and behaviour. Nor do we find any significant impact on male students' attendance rates or unexplained absences.

Of particular interest is that female students' behavioural responses to the program appear to be driven by the specific incentive or goal; for instance, the intervention has a positive impact on their behaviour scores when the challenge specifically targets behaviour outcomes. Similarly, female students have fewer unexplained absences when the specific challenge relates to attendance rate. Although we do not know exactly which mechanisms explain these gender differences, conclusions from other studies offer several possible explanations. First, such gender differences may be explained by the nature of the reward: those offered to students conditional on their achieving a specific goal were low-value symbolic rewards, which an experimental study of Swedish sixth graders suggests motivate girls but not boys (Jalava et al., 2013). Girls also tend to be more self-disciplined than boys, an advantage more relevant to teacher-assessed report card grades than to standardized achievement tests (Duckworth and Seligman, 2006). On the other hand, Jacob (2002) finds that 90 percent of the gender gap in higher education is due to differences in non-cognitive skills, such as the inability to pay attention in class, seek help from the teachers, or set and self-commit to goals like studying regularly and doing homework. The fact that girls are better at self-monitoring and achieving goals than boys (Zimmerman and Martinez-Pons, 1990) might at least partly explain the observed gender differences. As to our finding of a positive program impact on girls' behaviour scores but not boys', the behaviour progression of female students is also likely to be reflected in the overall teacher-assessed academic scores. Another potential contributing factor for the male-female disparity is that teachers may have different expectations for female and male reactions to the program, which might also be reflected in the outcomes. For example, Cornwell et al. (2013) find that female students benefit from (subjective) teacher grading because it is subject to a gender gap that exceeds the predicted gender differences between male and female students' test scores. Two plausible explanations for the different incentive effects for male and female students on the Naplan test are differences in male and female preferences for performing in competitive environments, as well as differences in self-confidence (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009). The Naplan, for instance, is competitive, with students receiving feedback not only on their absolute performance but also their performance relative to the entire nation. Moreover, Niederle and Vesterlund (2007), in an experimental setting, find that even in the absence of performance differences, female students are less likely than male students to choose to perform in a competitive environment. They also find that this reticence is only partly explainable by the fact that female students are less over-confident and more risk averse than male students. Rather, a large share of this male-female difference results from a female *disutility* for performing in a competitive setting. The program may thus be less effective in improving female students' test scores in competitive settings but more effective for male students.

5 Conclusion

In this study, we evaluate the impact of the FOGS Artie program, which not only provides information sessions stressing the importance of educational achievement but gives symbolic rewards to Indigenous high-school students conditional on their achievement of a specific educational goal. At the start of each term of 2012, all Indigenous students enrolled in the 21 schools involved in the program were invited to a launch organized by FOGS, during which former elite rugby players and administrative staff encouraged the students to perform well in school and announced a specific goal for the upcoming term. We provide evidence that the program did have a positive impact on a set of indicators of educational performance - including behaviour, academic grades and attendance - but that this positive impact was driven mainly by female students. Specifically, the program improved female Indigenous students' scores in both behaviour and academic grades and reduced the number of unexplained absences. Our results further indicate that symbolic rewards and specific challenges matter to female students because their positive responses to the program are larger for the indicators defined as semester targets. For the standardized national test (the Naplan), in contrast, the program improved scores only for male Indigenous students, which raises the question of what leads to this gender difference between female and male students. One possible explanation is that males outperform females in competitive settings because of such factors as stronger feelings of confidence or competence (Gneezy et al., 2003). Because individual Naplan test scores are evaluated and compared at the national level, the test can be seen as a competitive setting in which students can see their individual performance against the national average and the range of achievement for the middle 60% of students in Australia.

To the best of our knowledge, this study is the first to examine the impact of incentives on the schooling outcomes of Indigenous students in Australia. Nevertheless, we acknowledge several limitations to our study, not least the fact that participating schools were self-selected into the FOGS program, which raises concerns about unobserved characteristics that could be crucial for program success. That is, even in the absence of the program, these schools could have had better outcomes than schools that were not exposed to any intervention. This selection effect, however, although a fair concern, is somewhat attenuated by considering as our treatment group sample schools willing to participate in the program that have the closest probability of program involvement (based on observed characteristics). In addition, all participating schools were selected based on an objective criterion, the number of Indigenous students enrolled. Moreover, we find no evidence of pre-treatment differences in observed outcomes between treatment and control schools prior to the 2009 intervention, and the 2009 to 2012 outcome progression for non-Indigenous students not involved in the program provides convincing evidence of no differences between Indigenous students in treatment and control schools. Admittedly, a second major limitation is the lack of information on whether the Indigenous students were actually treated; that is, whether they were informed about the program and at what level they engaged in it. As a result, rather than estimating the average treatment effect of the program, we necessarily test for an intention-to-treat effect.

One important contribution of our study is that it shows that symbolic rewards combined with strong encouragement and support are effective in improving Indigenous students' educational achievement and attitudes towards education. Another important contribution of this study is the clear support it provides for family environment being a determinant variable not only for educational outcomes but also for the success of remedial programs in the later years of compulsory schooling. We have evidence, for example, that the program is effective in improving the educational outcomes of students from intact families but no indication of any positive impact on those of students from single-parent families. Hence, the program does not seem effective for students who are most disadvantaged, possibly because, as both Conti et al. (2010) and Heckman (2008) point out, interventions aimed at improving the educational trajectories of children from disadvantaged socio-economic backgrounds are more likely to be effective if undertaken early in life.

At this point, therefore, we can only estimate the short-term effects of the intervention, even though it is the long-term effects that are most relevant from a public policy perspective. In particular, it is crucial to know whether the intervention has persistent effects on Indigenous students' educational outcomes by changing their study habits or class behaviour, as well as their aspirations and attitudes towards education. We also cannot identify which of the initiatives - encouragement, incentives and learning support - is the most effective, or whether it is their combination which is driving the positive outcomes. Therefore, trying to disentangle the impact of each of these measures is an interesting avenue for future research. In addition, factors that are known to play an especially important role in educational outcomes include non-cognitive skills like time preferences, the capacity for self-commitment to achieving goals and self-confidence (Shoda et al., 1990). Hence, future research might also examine whether Indigenous children and non-Indigenous children exhibit different levels of such non-cognitive skills or preferences in their first years of formal education and whether early intervention programs can successfully address these differences.

6 Acknowledgements

We are particularly grateful to the Former of Origin Greats ARTIE program manager Matt Martin, director Gene Miles and project services officer Hayley Camman. We are also deeply grateful to the Queensland Department of Education, Training and Employment for their great collaboration, in particular to Leon Schwerin, Boyd Paties, Craig Blair and Tina Chau.

References

Angrist, J., Bettinger, E., Bloom, E., Kremer, M., and King, E. (2002). The effects of school vouchers on students: evidence from colombia. *The American Economic Review*, 92(5):1535–1558.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–163.

Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *The American Economic Review*, 92(4):1384–1414.

Angrist, J., Oreopoulos, P., and Williams, T. (2010). When opportunity knocks, who answers? new evidence on college achievement awards. National Bureau of Economic Research. Working Paper 16643.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Avvisati, F., Gurgand, M., Guyon, N., and Maurin, E. (2014). Getting parents involved: A field experiment in deprived schools. *Review of Economic Studies*, 81(1):57–83.

Barro, R. J. (2001). Human capital and growth. *The American Economic Review*, 91(2):12–17.

Bath, M. J. and Biddle, N. (2011). Measures of indigenous wellbeing and their determinants across the lifecourse. *CAEPR Lecture Series, The Australian National University*.

Berry, J. (2009). Child control in education decisions: an evaluation of targeted incentives to learn in india. Mimeo, MIT.

Bettinger, E. and Slonim, R. (2007). Patience among children. *Journal of Public Economics*, 91(1):343–363.

Bettinger, E. P. (2011). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics*, 94(3):686–698.

Blimpo, M. P. (2010). Team incentives for education in developing countries a randomized field experiment in benin. Mimeo, Stanford University.

Bradley, S., Draca, M., Green, C., and Leeves, G. (2007). The magnitude of educational disadvantage of indigenous minority groups in australia. *Journal of Population Economics*, 20(3):547–569.

Castillo, M., Ferraro, P. J., Jordan, J. L., and Petrie, R. (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, 95(11):1377–1385.

Conti, G., Heckman, J., and Urzua, S. (2010). The education-health gradient. *The American Economic Review*, 100(2):234–238.

Cornwell, C., Mustard, D. B., and Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1):236–264.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.

De Schutter, O. (2008). *Promotion and Protection of All Human Rights, Civil, Political, Economic, Social and Cultural Rights, Including the Right to Development: Report of the Special Rapporteur on the Right to Food, Olivier De Schutter: Building Resilience, a Human Rights Framework for World Food and Nutrition Security.* United Nations General Assembly.

Dee, T. S. (2011). Conditional cash penalties in education: Evidence from the learnfare experiment. *Economics of Education Review*, 30(5):924–937.

Doyle, O., Harmon, C. P., Heckman, J. J., and Tremblay, R. E. (2009). Investing in early human development: timing and economic efficiency. *Economics & Human Biology*, 7(1):1–6.

Duckworth, A. L. and Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, 98(1):198–208.

Eissa, N. and Liebman, J. B. (1996). Labor supply response to the earned income tax credit. *The Quarterly Journal of Economics*, 111(2):605–637.

Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126(4):1755–1798.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic inquiry*, 46(3):289–324.

Henry, G. T. and Rubenstein, R. (2002). Paying for grades: Impact of merit-based financial aid on educational quality. *Journal of Policy Analysis and Management*, 21(1):93–109.

Jackson, C. K. (2010). A little now for a lot later: A look at a texas advanced placement incentive program. *The Journal of Human Resources*, 45(3):591–639.

Jacob, B. A. (2002). Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, 21(6):589–598.

Jalava, N., Joensen, J. S., and Pellas, E. M. (2013). Grades and rank: Impacts of non-financial incentives on test performance. Available at SSRN 2362002.

Jones, L. P., Harris, R., and Finnegan, D. (2002). School attendance demonstration project: An evaluation of a program to motivate public assistance teens to attend and complete school in an urban school district. *Research on Social Work Practice*, 12(2):222–237.

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *Review of Economics and Statistics*, 91(3):437–456.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in us schools. *Educational researcher*, 35(7):3–12.

Leuven, E., Oosterbeek, H., and Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6):1243–1265.

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. MIT Working paper.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Paradies, Y., Harris, R., and Anderson, I. (2008). *The impact of racism on Indigenous health in Australia and Aotearoa: Towards a research agenda*. Darwin: Cooperative Research Centre for Aboriginal Health.

Patrinos, Harry Anthony, H. G. (2004). *Indigenous Peoples, Poverty and Human Development in Latin America*. The World Bank.

Patrinos, Harry Anthony, S. C. N. (1992). North american indians in the canadian labour market: A decomposition of wage differentials. *Economics of Education Review*, 11(3):257–266.

Ray, D. (2006). *Aspirations, poverty, and economic change.* Understanding poverty, Oxford University Press, Oxford.

Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the united states. *American Economic Journal: Applied Economics*, 4(4):121–139.

Rowse, T. (2002). *Indigenous futures: Choice and development for Aboriginal and Islander Australia.* Sydney: University of New South Wales Press.

Schultz, T. P. (2004). School subsidies for the poor: evaluating the mexican progresa poverty program. *Journal of development Economics*, 74(1):199–250.

Shoda, Y., Mischel, W., and Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental psychology*, 26(6):978–986.

Vermeersch, C. and Kremer, M. (2005). School meals, educational achievement, and school competition: evidence from a randomized evaluation. World Bank Policy Research. Working Paper No. 3523.

Zimmerman, B. J. and Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of educational Psychology*, 82(1):51–59.

A Appendix

A.1 Propensity score matching methodology

We estimate a probit model for the propensity of each school in queensland to be part of the treatment group. We regress an indicator variable on whether the school was part of the FOGS program on a set of school level characteristics observed in 2009 (pre intervention), which might affect the likelihood of being part of the program. We then match (with replacement) each treatment school with the schools which had a very close conditional probability of being part of the program, i.e. the ones with the smallest absolute difference in the propensity score, given the pre-intervention characteristics. Our sample includes all secondary or combined (primary and secondary) schools in queensland, excluding those located in two regions, darling downs south west and far north queensland, as none of the schools located in these two regions took part in the FOGS Artie program.

We estimate the following probit model:

$$Pr(f_s = 1|X) = \Phi(X'\beta)$$

where X is a vector of school level observable characteristics: school category (secondary or combined school), region (central queensland, metropolitan, north coast, north queensland, south east), zone (metropolitan, provincial, rural, remote), icsea (index of community socio-educational advantage), total number of non Indigenous students enrolled at the school, total number of Indigenous students enrolled at the school, average attendance rate of non Indigenous students and average attendance rate of Indigenous students.

The probit regression results in Table 1 show that schools with larger number of Indigenous students have a higher probability to be part of the program. This is consistent with the fact that the main selection criteria applied by the FOGS was the number of Indigenous students enrolled at the schools.

Table A.1: Probit regression for the propensity to be part of the FOGS program

	$Pr(f_s = 1 \vert X)$
Combined	-0.846
	(0.605)
Icsea	-0.005
	(0.005)
No. of Indigenous students	0.027***
	(0.008)
No. of non Indigenous students	-0.000
	(0.001)
Attendance rate of Indigenous	6.514
	(4.998)
Attendance rate of non Indigenous	-7.837
	(9.023)
Region (ref.= Central QLD)	
Metropolitan region	-1.440*
	(0.821)
North Coast region	-0.026
	(0.609)
North QLD region	-7.568**
	(3.538)
South East region	-0.729
	(0.723)
EQ zone (ref.= Metropolitan)	
Provincial	-1.932**
	(0.851)
Remote	-0.573
	(0.544)
Rural	0.486
	(0.868)
Observations	189
Prob > χ^2	0.000
LR χ^2 (13)	46.91

Marginal effects; Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2 Related literature

Table A.2: Summary of studies on incentive programs for educational achievement

Author(s)	Intervention	Sample	Results
Angrist et al. (2002)	PACES programs in Colombia: vouchers covering part of private high school fees	125,000 students randomly chosen among applicants	Reduced the probability of grade repetition for female students. Increased test scores by 0.2 standard deviations from the mean.
Angrist and Lavy (2009)	Achievement Awards demonstration in Israel: high cash awards for successful completion of high school exit exams	Random selection of 20 high schools among schools characterised by low success rates	Increased by 0.1 the probability of graduation for girls who were close to the graduation threshold.
Rodriguez-Planas (2012)	Quantum Opportunity Program in the US: mentoring, educational services and financial rewards	Low-performing students from 11 high schools randomly selected	Increased the probability of graduation for female students by 15 percent. It also increased females' probability of pursuing postsecondary education by 20 percent.
Schultz (2004)	Progresa program in Mexico: cash transfers to poor mothers conditional on school participation of their children	Families living in 324 randomly selected poor communities, among 495 eligible localities	Increased average number of years of formal education by 0.66.
Kremer et al. (2009)	Scholarships coupled with public recognition at a school assembly award conditional on achievement in Kenya	Grade 6 female students from 34 randomly selected primary schools	Increased test scores for high but also low-achieving female students who were unlikely to win the scholarship.
Angrist et al. (2009)	Educational support services and financial incentives, and a combination of both at a Canadian university	First year university students were randomly assigned to one of the treatments and control groups	Academic support services and the combined intervention increased performance of female students.

cont'd

Author(s)	Intervention	Sample	Results
Bettinger (2011)	Cash incentive program in Ohio conditional on students' academic scores	Randomly selected grades in 4 primary schools	Increased students' math scores by 0.15 standard deviations for students at the top of the test scores distribution.
Leuven et al. (2010)	Cash awards conditional on passing all first year subjects at the university of Amsterdam	249 first year students randomly assigned to two treatment (large and small reward) and a control group	Increased achievement of high ability students and decreased achievement of low ability students.
Fryer (2011)	Cash incentives conditional on achievement in three US districts	In one district, students from 22 randomly selected schools were paid to read book; in the other two districts students from 80 schools were paid for their test scores performance	No effects of the program.
Dee (2011)	Wisconsin Learnfare program: sanctioned welfare grants when teens failed to meet regular school attendance	Schools in 9 counties randomly selected	Increased school enrolement by 3.5 percent and attendance by 4.5 percent.
Avvisati et al. (2014)	Information sessions to parents on the benefits of education in deprived areas in Paris among those who were willing to take part in the program	Parents of grade 6th students from 34 middle schools. Parents who received the intervention were selected	Positive effect on attendance and behaviour, but no effect on test scores.