RUHR
ECONOMIC PAPERS

Alfredo R. Paloyo
Sally Rogan
Peter Siminski

The Effect of Supplemental Instruction on Academic Performance – An Encouragement Design Experiment

rwi

#603

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #603

Responsible Editor: Jochen Kluve

ISSN 1864-4872 (online) – ISBN 978-3-86788-700-7

Ruhr Economic Papers #603

Alfredo R. Paloyo, Sally Rogan, and Peter Siminski

The Effect of Supplemental Instruction on Academic Performance – An Encouragement Design Experiment

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788700
ISSN 1864-4872 (online)
ISBN 978-3-86788-700-7

Alfredo R. Paloyo, Sally Rogan, and Peter Siminski[1]

The Effect of Supplemental Instruction on Academic Performance – An Encouragement Design Experiment

Abstract

While randomized controlled trials (RCTs) are the "gold standard" for impact evaluation, they face numerous practical barriers to implementation. In some circumstances, a randomized-encouragement design (RED) is a viable alternative, but applications are surprisingly rare. We discuss the strengths and challenges of RED and apply it to evaluate a mature Supplemental Instruction (SI) or PASS (Peer Assisted Study Session) program at an Australian university. A randomly selected subgroup of students from first-year courses (N = 6954) was offered large incentives (worth AUD 55,000) to attend PASS, which increased attendance by an estimated 0.47 hours each. This first-stage (inducement) effect did not vary with the size of the incentive and was larger (0.89) for students from disadvantaged backgrounds. Instrumental variable estimates suggest that one hour of PASS improved grades by 0.065 standard deviations, which is consistent with the non-experimental literature. However, this estimate is not statistically significant, reflecting limited statistical power. The estimated effect is largest for students in their first semester at university.

JEL Classification: C93, I21, I23, I24

Keywords: Australia; randomized-encouragement design; student outcomes; peer-assisted study session; supplemental instruction; selection bias

February 2016

---

1 Alfredo R. Paloyo: Centre for Human and Social Capital Research (CHSCR), University of Wollongong (UOW), RWI, and IZA; Sally Rogan: UOW; Peter Siminski: CHSCR, UOW, and IZA. *– The authors are grateful for research assistance provided by Alex Cousley, James Passfield and Daniel Thomason. The idea for this evaluation was developed in a conversation between Peter Siminski and Daniel Thomason. We are also grateful for the comments we received across a number of seminars, workshops, and conferences in which this paper was presented. This study received funding from the Higher Education Participation and Partnerships Program (HEPPP) of the Australian Government and an Educational Inclusion Grant from the UOW Faculty of Business. Ethics approval was obtained from the Social Sciences Human Research Ethics Committee of the University of Wollongong (HE13/360). This trial was registered with the American Economic Association's randomized controlled trials registry with the title "The Causal Effects of the Peer Assisted Study Sessions (PASS) on Educational Outcomes" (AEARCTR-0000290).* *- All correspondence to: Peter Siminski, School of Accounting, Economics, and Finance, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia, e-mail: siminski@uow.edu.au*

1 Introduction

Selection bias is a pervasive challenge for evaluating the impact of any intervention where there is self-selection into participation. Participants and non-participants are likely to differ on observed and unobservable characteristics, and these differences can also influence the outcome being measured. The dominant view is that randomized controlled trials (RCTs) constitute the "gold standard" for impact evaluation because the random assignment of the policy or treatment eliminates the bias arising out of endogenous selection. Hence, evidence obtained from RCTs sits firmly atop most hierarchies of evidence. Nevertheless, the prevalence of RCTs in evaluating social programs is limited for a variety of reasons. Major barriers to their implementation include program cost, ethical and practical considerations, and political and social acceptance. Treatment randomization may not be operationally feasible when a program already exists, participation is voluntary, and program proponents—for ethical or political reasons—refuse to deny treatment to eligible participants.

Recognizing the difficulties associated with deploying an RCT in the field, we have two principal objectives in this paper. First, it is to argue that the randomized-encouragement design (RED) is a viable alternative evaluation strategy, especially considering that it survives many of the challenges that RCTs face. Second, it is to demonstrate the use of an RED in estimating the impact of supplemental instruction (SI) in tertiary education. With respect to the strategy, we show that genuine REDs have found limited application in general but also specifically in Economics. This is a peculiar state of research and practice since REDs have structural properties that are well-understood by practitioners, and the design could be used in contexts where an RCT would be operationally or ethically infeasible. In terms of our specific application of an RED, we provide suggestive experimental evidence based on a large and representative sample that SI improves the academic performance of participants, although there is a large degree of uncertainty associated with the estimated impact. Our application thus serves to highlight the difficulties associated with an RED-based evaluation of a program.

Dawson et al. [2014] conducted a systematic review of research on SI which indicated that SI was "effective". The degree of effectiveness was tentatively summarized as a 0.5-standard-deviation increase in grades associated with SI attendance. According to the authors, their review is consistent with correlational claims made by the US Department of Education between SI participation and outcomes such as mean grades and failure, withdrawal, retention, and graduation rates. Therefore, to deny students the opportunity to receive these putative benefits merely for the benefit of research would be an ethically tenuous proposition. Nevertheless, the authors noted the need for experimental evaluations, citing only one such study which had a sample of just 67 students, of which 24 were in the treatment group [Parkinson 2009]. Due to the likely presence of selection bias, experimental evidence in this context is necessary to be able to credibly estimate the program's impacts on the outcome measures of interest.

We present the results of a large-scale RED to evaluate a mature SI program: the Peer Assisted Study Sessions (PASS) program at the University of Wollongong (UOW), a regional Australian university located in the state of New South Wales. We are interested in estimating the impact of PASS on academic performance. PASS is well-suited to be evaluated within the context of an RED since students can voluntarily attend the study sessions and denying access or compelling participation would be ethically questionable. The experiment demonstrates the usefulness of REDs as a tool to evaluate similar programs, but we also document the associated challenges that researchers face when developing and implementing an encouragement design. While REDs should constitute part of the policy evaluator's toolkit together with RCTs and other quasi-experimental designs and associated estimation techniques, potential difficulties with this approach must also be taken into consideration.

The estimation results suggest that an hour of PASS increases the standardized final grade by 0.065 standard deviations. However, the estimated impact is not statistically significant. If the impact of PASS is constant for each hour of PASS, then an average PASS attendance of 6.25 hours over a session would increase marks by 0.41 standard deviations, although the confidence interval is fairly wide. We provide further suggestive evidence that the size of the impact is contingent on whether the student is on his or her first semester in the university. In particular, we note that the impact magnitude could be larger for those in their first semester.

In conducting an RED, we encountered a specific difficulty associated with this evaluation approach. The precision of the estimates largely depends on the strength of the incentive to induce people to participate in the program. If there is a weak relationship between the incentive and the probability of participation, then the resulting estimate of the impact of the program would be associated with a high degree of uncertainty, putting into question the credibility of the estimate itself. This prevents researchers such as ourselves from further analyzing treatment-effect heterogeneity across subgroups, as these would necessarily involve smaller sample sizes, thereby exacerbating the problem of statistical inference.

The remainder of the paper is structured as follows. We describe the randomized-encouragement design, and then discuss its advantages and disadvantages. In Section 3, we describe SI, specifically the PASS program at UOW. In Section 4, we outline the experimental design and estimation approach. We discuss the results in Section 5, including the challenges we faced in using an RED for this evaluation. We conclude in Section 6.

2 Randomized-Encouragement Design

When a randomized controlled trial is infeasible or unjustifiable, the researcher who is tasked with evaluating a program's impact will instead usually follow one of two alternative approaches: (1) conducting observational studies with no experimental features or (2) exploiting natural or quasi-experiments. In the former, one compares outcomes between participants and non-participants while usually controlling for observed differences using regression or matching techniques. While typically easier to implement, this approach cannot account for unobserved differences, which are often likely to be important as well.

As for natural experiments, researchers often exploit policy changes or discontinuities, and sometimes even lotteries (this is usually the case when the program is oversubscribed, and the implementing agency randomizes eligibility or participation in the interest of fairness and transparency). These are features that are typically inherent in the program and were not designed with evaluating the program in mind. While these quasi-experimental estimation strategies allow the researcher to obtain a reasonable estimate of the program impact, the absence of a natural experiment precludes the evaluation of programs where such a phenomenon has not occurred. From the perspective of a policymaker and evaluator, the serendipitous manifestation of a natural experiment is not a particularly promising strategy to pursue, especially if there is a desire to make independent evaluations an integral part of a transparent and accountable governance strategy.

The randomized-encouragement design is an alternative impact evaluation strategy which can be used to evaluate existing voluntary programs that have partial take-up. In a RED evaluation, a randomly selected subgroup is offered an incentive or encouragement to participate in a given program. The incentive or encouragement could take a number of forms, such as a direct financial transfer or the provision of more information about a program. RED is akin to RCTs with partial or imperfect compliance, where consistent estimates of treatment effects can be recovered by standard instrumental-variable regression techniques [Angrist et al. 1996; Bloom 1984]. While participation is voluntary, one can expect that it is higher in the incentivized group. Treatment status can thus be instrumented by the randomly assigned incentive.

RED is suitable for evaluating programs that have already been rolled out or where it is impractical or unethical to deny the program to any willing and eligible participant. Job-training programs, preventive-healthcare initiatives, educational programs, and various other similar social policies are potential applications. For instance, many states in the US have reemployment initiatives—such as Texas's Rapid Reemployment Services or Pennsylvania's Rapid Reemployment Program—designed to quickly match unemployment-benefit claimants to new employers. While the detailed implementation of these programs are different, they share characteristics that make them amenable to an RED evaluation, such as partial take-up and voluntary participation. Unlike RCTs, or other experiments which randomly assign *eligibility*, REDs allow the researcher to estimate causal impacts without debilitating program delivery because participation is not denied for people in the control group. To the extent that the encouragement or promotion increases program participation, an RED can increase take-up rates, which may itself be a policy goal.

Despite their advantages, REDs have not seen mass appeal in the impact-evaluation literature, including in the Economics discipline.[2] This will probably come as a surprise to many readers because REDs can appear very similar to related approaches. We propose that a distinguishing characteristic of a genuine RED is that encouragement is deliberately randomized to estimate the impact of a program. Randomization for some other purpose or randomization that occurs by chance represents a natural experiment. While natural experiments are often just as convincing, the researcher can only exploit these opportunistically, when and where they arise. In contrast, in an RED, the researcher deliberately and purposefully imposes exogenous variation in the probability of program participation in order to specifically estimate the program impact.[3]

We also distinguish REDs from studies which primarily aim to evaluate the effectiveness of an incentive or encouragement. The latter are “regular” experiments, where the treatment is an incentive or information.[4] Unlike REDs, such experiments do not aim to evaluate an existing program. Finally, REDs are also similar to experiments in which program eligibility is randomly assigned. As discussed above, randomized eligibility involves denying access to the control group, which is unlikely to be a feasible strategy for evaluating a mature, existing program or where it would be unethical to deny access to potential beneficiaries.

---

[2] A cursory search (excluding citations) in Google Scholar for “randomized encouragement design” turned up 112 hits while “randomized controlled trial” had 1.61 million (1 June 2014). Note that the count for REDs is overstated, since a number of studies are mistakenly classified as an RED just because the treatment is an incentive, and the incentive is randomized. However, these studies should properly be viewed as an RCT where the treatment is an incentive. We discuss this further in the succeeding paragraph.

[3] Many natural experiment studies exploit lotteries that were conducted for other reasons. These lotteries are usually motivated by equity considerations, where limited places are rationed. High-profile examples include studies exploiting school voucher lotteries [Angrist et al. 2006], conscription lotteries [Siminski and Ville 2012], public health insurance lotteries [Baicker and Finkelstein 2011], and migration lotteries [Gibson et al. 2011].

[4] For example, there is a large literature on direct incentives tied to academic outcomes [Angrist and Lavy 2009; Fryer 2011; Kremer at al. 2009; Angrist et al. 2009; Angrist et al. 2014; Barrow et al. 2012; Cha and Patel 2010; Leuven et al. 2010]. Other examples of evaluations of incentives or information interventions are Just and Price [2013], Duflo and Saez [2003], and Card et al. [2010].

Limited access is a feature for oversubscribed programs, but access is also denied for other reasons. For instance, pension programs, merit-based scholarships, or anti-poverty programs are usually allocated based on whether the value of a continuous measure falls on either side of a predefined cutoff (age for pension eligibility and receipt, grades for merit-based scholarships, and an index measure of poverty for anti-poverty measures). This situation allows for an evaluation based on the regression-discontinuity design, a quasi-experimental strategy. Such program features, however, were the result of the inherent design of the program and did not come about because a future evaluation was expected.

Even in the absence of program-inherent cutoffs or, say, policy discontinuities along geopolitical borders, randomly encouraging a subset of the population to undertake the treatment can allow for a credible ex-post evaluation to be conducted. Herein lies the main advantage of an RED: there is no need to purposefully deny eligible participants from the program (neither are they compelled to participate) to obtain a useful estimate of the impact. This makes it simultaneously unobjectionable for many policymakers—especially those with voting constituents (at least as long as the result is to their favor)—and for ethical review boards, which are unlikely to approve outright program denial solely for the purpose of facilitating a research environment suitable for impact evaluation.

Genuine RED studies are relatively rare. The Moving To Opportunity experiment [Katz et al. 2001] is one RED evaluation published in an Economics journal.[5] This RED experiment was used to estimate the impact of relocation on family well-being. RED-based evaluations have appeared in the Education [Powers and Swinton 1984], Biostatistics [Hirano et al. 2000], Political Science [Albertson and Lawrence 2009] and Public Health [Martino et al. 2012] literatures. That REDs have not been more widely used is somewhat of a puzzle since, as Holland [1988:453] says, "[e]ncouragement designs can arise in *any* study of human subjects in which the treatments or causes of interest must be voluntarily applied by the subjects to themselves." This paucity of REDs is even more surprising since most discussions (including those in the Economics literature) of alternatives to RCTs include the encouragement design (e.g., Duflo et al. [2008], Gertler et al. [2011], Imai et al. [2013], TenHave et al. [2003], and West et al. [2008]), and credible estimation techniques are now widely developed, even with complications like missing data (e.g., Barnard et al. [2003] and Zhou and Li [2006]).

In the context where there is a potential benefit attributable to the treatment—especially when there is some evidence for this already, however poorly estimated—it would be

[5] A few other papers are Boucher and Mullally [2010], McKenzie and Özler [2014], and Beam [forthcoming].

ethically questionable to deny people this treatment merely for purposes of evaluation. Therefore, a natural evaluation strategy would be the RED, where the control group is not denied the treatment altogether. In an influential publication of the World Bank designed for practitioners [Gertler et al. 2011:153], the authors argue that "[t]he most basic principle in the assignment of program benefits is that the delivery of benefits should never be denied or delayed solely for the purpose of an evaluation." The program—as it is envisioned by its proponents—should dictate how it should be evaluated, and not the other way around (for evaluators to dictate program delivery solely for the benefit of obtaining credible impact estimates).

However, we do not discount that there are instances when treatment denial is unavoidable. As mentioned above, programs may sometimes be oversubscribed, and proponents have to find a way to allocate its provision in a manner that is both transparent and fair. Indeed, in this situation, the randomized assignment of treatment among those who are eligible is perhaps the most transparent and fair way to allocate the scarce resource since everyone who is eligible faces an equal probability of being provided the treatment. But when the program is fully resourced, treatment denial on the basis of the researcher's objective of achieving an experimental setting rests on shaky ethical grounds.

Moreover, existing programs are particularly amenable to an RED analysis because the administrative overhead is already in place. RCTs run by economists in the field are usually implemented for new programs, and these require staff training, the creation of an evaluation team (including, among others, a field team, data managers, and the analysis team), and financing for the program delivery itself. In contrast, in evaluating an existing program via an RED, the researcher can work with a team already experienced in the delivery of the specific program and the team most likely to deliver it in the future. Data that the program staff members already routinely collect can be immediately used by the researcher (subject to the usual data-cleaning process), and any necessary additional information is not likely to be as much as what would be required in an RCT.

All of these imply that an RED evaluation team would not need to micromanage the program-delivery aspect of the evaluation, thereby minimizing the risk of "Hawthorne effects" on the program team.[6] It also minimizes the problem of scaling up—so-called "gold-plating"—

[6] The Hawthorne effect refers to a change in behavior of the subjects under study. The presence of such an effect can limit the generalizability of the results. In the situation mentioned in the text, the "Hawthorne effect" does not arise from the change in the subjects' behavior, but on the behavior of those people who are providing the treatment. Program team members may work better than they usually would if they were not being managed by the research or evaluation team. In principle, this can also arise in an RED, but since the program under evaluation in an RED is usually a pre-existing one, the magnitude of behavioral change is not likely to be large enough to threaten the external validity of the experimental design.

programs that have been shown to be effective in RCTs. Both of these threaten the external validity of results obtained from RCTs. These programs deployed within the framework of an RCT are relatively small in scale, and they are well-managed by the research team. When the program has been shown to be effective, there is generally a call to scale it up, where now usually the state or some other organization (as opposed to the well-trained team implementing it in the trial) has to deliver it. In these cases, the large bureaucratic team's competence in delivering the national program is not typically on a par with the carefully selected and trained trial team, and, as such, the program benefits may no longer necessarily materialize. Bold et al. [2013] document this for an intervention in Western Kenya, where the benefits found in nongovernmental organization-led trials failed to manifest themselves when the Kenyan government took over program delivery.

3 The PASS Program

This section describes the specific program which is the subject of the RED evaluation. We begin with a discussion of peer learning and its implementation as a Supplemental Instruction (SI) or Peer Assisted Study Sessions (PASS) program. We then describe the specific case of PASS at the University of Wollongong.

3.1 Supplemental Instruction or Peer Assisted Study Sessions

A number of nonexperimental studies suggest that peer learning and student leadership programs at university contribute to student learning outcomes, participation, and retention rates [Blanc, DeBuhr, and Martin 1983; Kuh 2003; Pascarella and Terenzini 2005]. There are many variations of such programs and roles. Some are designed solely around pastoral care or transition needs and may take the form of a senior student being assigned to one or several first-year students for a specified period. Others may involve a model of students from a particular background being targeted for assistance, and it may be compulsory for students to participate. Still others offer academic assistance with student leaders as tutors for individuals or groups. These senior tutors may sometimes be involved with grading papers or exams.

One of the more widely known peer learning programs is Supplemental Instruction (SI), which is usually called PASS (Peer Assisted Study Sessions) in Australasia. SI was developed in the US at the University of Missouri–Kansas City (UMKC) in 1973. SI or PASS and its variations are now offered to thousands of students worldwide [Arendale 2002]. By 2009, staff from over 1,500 tertiary institutions from 29 countries had been trained in the implementation of the program [Martin 2009].

PASS is a free—in the sense that students do not pay an upfront or direct cost to attend other than their time cost—and voluntary supplementary academic assistance program that utilizes peer-led group study to assist students enrolled in targeted subjects or courses. The program is specific to each subject, and it consists of informal but regularly scheduled sessions. Each session is independent and is focused on the content- and discipline-specific study strategies in the given subject. PASS is commonly attached to subjects which many students may find challenging, and any student enrolled in that subject is eligible to participate. Marketing and communication of the program stresses that all students are welcome to attend. In an effort to avoid the stigma associated with remedial instruction, PASS is not targeted to specific students or subsets of students. The sessions are facilitated by current students—so-called “PASS Leaders”—who have recently completed (and, in most cases, have excelled in) the subject. The leaders are recruited based on their academic results and interpersonal skills.

The role of the PASS Leader is not to reteach lecture material or to directly answer questions. Using their own experiences and the concerns of participants around challenging topics or questions, they instead facilitate the discussion, utilize the knowledge of participants and resources, such as lecture notes and textbooks, and generally guide the group to arrive at correct answers. Participants are involved in setting the agenda at the beginning of each session, ensuring it meets their learning needs as much as possible. The PASS Leader has no involvement in grading papers or exams, which presumably provides participants with a non-threatening environment to ask questions which they may be hesitant to put to an academic staff in a more senior or formal role [Longfellow et al. 2008].

3.2 PASS at UOW

PASS at UOW is a highly awarded program. Its accolades include an Australian Learning and Teaching Council Program Award and Most Outstanding PASS Program in the World Award from UMKC, both in 2010. It was also the recipient of two commendations (2006 and 2011) from the former Australian Universities Quality Agency and was awarded institutionally in 2007 for Outstanding Contribution to Student Learning. PASS at UOW had its origins in 2002 with Computer Science and Business courses. Since 2007, the program has supported students in all faculties at UOW. Since 2005, PASS at UOW has been accredited by UMKC as the National Center for PASS/SI in the Australasia region, providing training for other PASS programs at some 70 institutions in this part of the world. In this capacity, the National Center at UOW has led the second wave of implementation of the program in this region after earlier attempts by other Australian universities faltered in the 1990s. Thirty-seven of the 39 Australian universities now have staff trained by UOW in implementing PASS.

In 2014, PASS at UOW delivered 40,000 contact hours to over 4,000 individual students. While the majority of subjects supported are at the first-year level, PASS also supports some second-year and post-graduate subjects, particularly those which may have a high percentage of students transitioning into their first semester at UOW. The program consists of one-hour weekly sessions for 12 of the 13 weeks in a full semester at UOW. Regular participation (five or more for a particular subject) is strongly encouraged.

PASS Leaders at UOW are typically recruited from students who have been regular participants previously. The PASS Leader team normally consists of about 90 to 100 students. All new PASS Leaders receive two days of initial training and one day of team building and professional development before the commencement of the semester. All new Leaders also receive a senior mentor (often a more experienced PASS Leader) to provide observations and feedback and generally support them in their personal and professional development within the role. The UOW model of senior mentors and full-time staff conducting observations and debriefs is designed to ensure that Leaders are undertaking the role for which they have been trained. Ongoing workshops and meetings are held throughout the year to facilitate further skill development around topics such as deeper-level-questioning skills, enhancing positive group dynamics, and facilitating the involvement of all participants.

4 Experimental Design

To estimate the impact of attendance in the PASS program on a number of student outcomes, we implemented a randomized-encouragement-design experiment. The intervention (PASS attendance) was exogenously varied by manipulating the probability of obtaining supplemental instruction via PASS. In this section, we first describe the problem involved in causal inference within the present context and how randomized encouragement solves it. Second, we elaborate on the estimation strategy. Lastly, we discuss the execution of the experiment, including sample selection and incentive randomization.

4.1 Selection Bias in Impact Evaluation

It is not sufficient merely to compare the observed mean outcomes between those who received SI and those who did not and use the difference as an estimate of the impact of supplemental instruction because of the likely presence of self-selection bias. The observed and unobserved characteristics between these two groups could be different, and these differences may influence both the decision to participate and the outcome of interest. For example, innate but unobserved motivation may influence both the student's decision to enroll in an SI program and the final grade obtained in the class. This confounds the estimate of the program impact obtained from a simple comparison of means.

More formally, consider a generic outcome variable y_i for individual i, and let d_i denote a binary variable that equals 1 if individual i received supplemental instruction.[7] Two potential outcomes exist for each individual, $y_i(0)$ without SI and $y_i(1)$ with SI, but we only observe $y_i = y_i(d_i) = y_i(0)(1-d_i) + y_i(1)(d_i)$. At the population level, one can show that

$$\begin{aligned} &\mathrm{E}[y_i|d_i = 1] - \mathrm{E}[y_i|d_i = 0] \\ &\qquad = \{\mathrm{E}[y_i(1)|d_i = 1] - \mathrm{E}[y_i(0)|d_i = 1]\} \\ &\qquad + \{\mathrm{E}[y_i(0)|d_i = 1] - \mathrm{E}[y_i(0)|d_i = 0]\}. \end{aligned} \tag{1}$$

One can see from Equation (1) that the difference in outcomes between the treated and untreated group consists of, first, the impact of the program among those who received SI (the first set of terms in braces, generally called the average treatment effect on the treated (ATET), which is the parameter one is generally interested in estimating) and, second, the selection bias (the second set of terms in braces).

In the context of this study, it is unlikely that there is no selection bias. On one hand, good student outcomes are associated with highly motivated students, but these are precisely the same students who may have a higher probability of engaging SI. They are also the students who may believe, *ex ante*, that they have the most to gain from SI. That is, the inequality $\mathrm{E}[y_i(0)|d_i = 1] > \mathrm{E}[y_i(0)|d_i = 0]$ could hold, so that the selection-bias term is nonzero. Without taking this into account, one would overestimate the impact of SI. On the other hand, students who require SI could be negatively selected as well. Students who are subject to poor educational inputs for which we are unable to control (such as household or parental characteristics) may use SI to compensate. In this case, their potential outcome would be worse than those who chose not acquire SI: $\mathrm{E}[y_i(0)|d_i = 1] < \mathrm{E}[y_i(0)|d_i = 0]$.

This is true even in a regression framework, where we are able to control for differences in observed characteristics. Suppose the estimating equation is as follows:

$$y_{ij} = \tau d_{ij} + \boldsymbol{\beta}'\mathbf{x}_{ij} + a_j + \epsilon_{ij}, \tag{2}$$

where $\mathbf{x}_{ij}$ is a vector of individual-specific exogenous characteristics in subject or class j (including a constant), a_j are subject fixed effects, and ϵ_{ij} is an idiosyncratic stochastic disturb-

[7] The exposition here draws from Angrist and Pischke [2009], Holland [1986], and Imbens and Wooldridge [2009] unless otherwise cited. In the analysis below, the treatment status is actually measured as a treatment intensity, specifically, the number of SI sessions attended in a semester. We present the binary treatment case here merely for ease of exposition; the substantive interpretation is similar in the continuous case.

ance. Ordinary least-squares (OLS) estimates of the parameter τ, which is the impact of attending SI, and vector of parameters $\boldsymbol{\beta}$ will be biased and inconsistent because $\mathrm{E}[\epsilon_{ij}|d_{ij}, \mathbf{x}_{ij}, a_j]$ is not equal to $\mathrm{E}[\epsilon_{ij}]$ due to the influence of unobserved characteristics on both d_{ij} and y_{ij}. Whether τ is over- or underestimated turns on the relative degree of positive and negative selection, although both selection effects are certainly plausible.

If treatment assignment were to be randomized, as is usual in, say, RCTs, the potential outcomes would be independent of treatment status—that is, $\{(y_i(0), y_i(1)) \perp d_i\}$, where $\perp$ denotes statistical independence. Thus, by design, the selection bias will be equal to zero in Equation (1) or, in the regression framework, the conditional mean of the error will be equal to its unconditional mean, leading to unbiased and consistent OLS estimates of the population parameters in Equation (2).

4.2 Estimation Strategy

Due to the experimental design, the treatment parameters can be naturally estimated via OLS and two-stage least-squares (2SLS) instrumental variables. Under certain assumptions, these distinct estimators will recover different parameters. In particular, two parameters are estimated below: first, the "intention-to-treat effect" (ITTE), i.e., the causal impact of being randomly assigned to the incentivized group; second, the average causal response (ACR) since the number of PASS sessions attended can be construed as a measure of treatment intensity [Angrist and Imbens 1995].

Let p_{ij} denote the number of times student i attended the PASS session for subject j, where p_{ij} can take on integer values between 0 and 12, and let z_{ij} denote group assignment, with $z_{ij} = 1$ indicating that student i in subject j was randomly allocated to the group that had a positive probability of winning the near-cash incentive and $z_{ij} = 0$ indicating otherwise.[8] Note that p_{ij}—the treatment-intensity variable—remains a choice variable and is therefore endogenous. It is z_{ij}—the instrumental variable—that is exogenously manipulated by the researchers.[9]

By design, the variation in z_{ij} is exogenous, so $\{(y_{ij}(p_{ij}), p_{ij}(z_{ij})) \perp z_{ij}\}$ necessarily holds—that is, the potential outcomes and intensity of treatment are independent of group assignment. Moreover, it is unlikely that the behavior of students who were not incentivized was affected by their exclusion from the end-of-semester lottery to give away the cash prizes. In

[8] The incentive—discussed in more detail in Section 3.3—is a chance to win one of 50 gift certificates worth AUD 1,000 for the first two semesters and a chance to win one of five gift certificates for the third semester we ran the experiment.

[9] Think of d_{ij} in Section 3.1 as a binary version of p_{ij}, where $d_{ij} = 1$ could mean "high" values of p_{ij} and $d_{ij} = 0$ correspond to "low" values of p_{ij}.

$\{(y_{ij}(p_{ij}), p_{ij}(z_{ij})) \perp z_{ij}\}$, we make it explicit that the treatment intensity, p_{ij}, is a function of the instrument, z_{ij}. OLS estimates of the following regression model, usually called the "reduced form",

$$y_{ij} = \rho z_{ij} + \boldsymbol{\theta}'\mathbf{x}_{ij} + a_j + \varepsilon_{ij}, \tag{3}$$

will produce unbiased and consistent estimates of the impact of being randomly assigned to the incentivized group—represented by the parameter ρ—on the outcome variable, y_{ij}. In other words, the estimate $\hat{\rho}$ from Equation (3) is an estimate of the ITTE.[10]

Note, however, that ρ is clearly not the impact of PASS attendance, p_{ij}, on the outcome, y_{ij}. Nevertheless, the ITTE estimate is relevant for policy in this context. For instance, policymakers may already be committed to the PASS program—or any other intervention, for that matter—and are not necessarily interested in its exclusive effect on student outcomes, perhaps because its positive impact has already been credibly established elsewhere. They may still, in any case, endeavor to find out whether positively incentivizing PASS attendance can significantly contribute to improved outcomes, possibly within the context of a cost–benefit analysis. Policymakers may then compare the costs of running an incentive program in addition to PASS itself against the benefits attributable to PASS attendance. In many other instances, this is the relevant policy question, more so than the question of whether the program itself has an impact.

If we further assume that $p_{ij}(1) \geq p_{ij}(0)$ for all i, then we can use the exogenous variation in z_{ij} to estimate the ACR generated by p_{ij} on y_{ij}. This assumption is the monotonicity condition in Angrist and Imbens [1995] and Imbens and Angrist [1994], and its implication is that being randomized into the incentive group can only impact PASS attendance in a non-negative way.[11] That is, if an individual is likely to attend PASS without the incentive, then she is much more likely to do so when positively incentivized. Although this assumption is fundamentally untestable, it is most probably trivially satisfied in our context since there is no reason for an individual to be discouraged from PASS attendance with the introduction of the incentive. The *ex-ante* utility gain is always going to be higher with the positive probability of winning the monetary reward.

The existence of an instrument, which is the first condition in Imbens and Angrist [1994], is satisfied by the randomization of the encouragement. Furthermore, the instrument,

---

[10] Note that $\mathbf{x}_{ij}$ and the subject fixed effects include controls for the randomization strata [Angrist et al. 2014; Bruhn and McKenzie 2009]. In our case, this is the subject, an indicator for aboriginality, an indicator for being an international student based on broad funding, and an indicator for being female.

[11] Theoretically, it is also possible for the inequality to go in the opposite direction (but that is unlikely in this case). That is not problematic for the estimation, though. The crucial part about the assumption is that the instrument should affect everyone in the same direction.

z_{ij}, can only affect y_{ij} via its impact on p_{ij}. In other words, z_{ij} is excludable from the following outcome equation:

$$y_{ij} = \gamma p_{ij} + \boldsymbol{\delta}'\mathbf{x}_{ij} + a_j + \nu_{ij}. \tag{4}$$

This means that the incentive has no direct impact on student outcomes or that z_{ij} is uncorrelated with the error term, ν_{ij}, in Equation (4). Any observed differential in mean y_{ij} between the incentivized and non-incentivized groups can only be attributed to the variation in PASS attendance induced by the chance of winning the gift voucher. The receipt of an electronic ticket in the lottery is made contingent not on the eventual student outcome but rather on PASS attendance only. Thus, it is not likely that the lottery can increase, say, student motivation (and, hence, student outcomes) unless it was through increased PASS attendance.

The IV estimate of γ in Equation (4) is the ACR attributable to PASS attendance on a specific student outcome measured by y_{ij}. The estimate can be retrieved via the sample analog of Wald's [1940] grouping estimator:

$$\gamma = \frac{\mathrm{E}[y_{ij}|z_{ij} = 1] - \mathrm{E}[y_{ij}|z_{ij} = 0]}{\mathrm{E}[p_{ij}|z_{ij} = 1] - \mathrm{E}[p_{ij}|z_{ij} = 0]}.$$

The ACR parameter captures a weighted average of causal responses to changes in treatment intensity for the subpopulation whose treatment status was induced to change by the instrument [Angrist and Imbens 1995].[12] Wald's formulation makes explicit that it is the ratio of the ITTE (numerator) to the "compliance rate" (denominator) that produces the ACR.

Although the ACR is consistently estimated, there is no *a priori* reason to believe that these estimates are representative for the whole population or even for the subpopulation who attend PASS. This would only be the case if we assume that the impact of PASS attendance is the same for everyone in the population (i.e., the treatment effect is constant) or if compliance is not a function of the individual's treatment effect (their "gain" from PASS). Once we allow for heterogeneous treatment effects or if the compliance rate increases proportionally with the size of the gains, the effect or response parameter recovered by IV is only valid for the population of compliers or those whose treatment intensity or status was affected by the instrument [Angrist 2004; Angrist and Imbens 1995; Heckman and Vytlatcil 2005; Imbens and Angrist 1994].

One may argue that the ACR may not be the most policy-relevant parameter. In this paper, the compliers are those students who, for whatever reason, would not have attended PASS or would not have attended as many PASS instances as they did had they not been offered the chance to win the gift certificate. This peculiar characteristic sets them apart from those people

[12] This paramater is sometimes called the complier average causal effect (CACE).

who would have attended PASS with or without the additional incentive. Unfortunately, it is generally not possible to characterize the differences between compliers and non-compliers since they cannot be identified from the data.

From a policy perspective, however, the compliant group is just as important as the population-average group. The compliers are precisely those students who would increase their participation in PASS if they were appropriately incentivized. Estimating this marginal treatment effect is again useful, for instance, in the context of comparing additional costs and benefits of the incentive against the potential gains in student outcomes. If students select into PASS on gain (i.e., those who have the most to gain from PASS are the ones who take it up), then the ACR (i.e., the parameter that we estimate) is likely to be a lower bound for the average treatment effect on the treated (ATET).

Although one can obtain consistent estimates of the ACR with just the triple of $\{y_{ij}, p_{ij}, z_{ij}\}$, the precision of the estimates can generally be improved with the inclusion of covariates, with little impact on the interpretation of the coefficients [Angrist and Imbens 1995]. This is most easily implemented by estimating the model via 2SLS. The second-stage equation is essentially Equation (4) but with p_{ij} replaced by its predicted value based on parameter estimates from the following first-stage equation:

$$p_{ij} = \pi z_{ij} + \boldsymbol{\phi}'\mathbf{x}_{ij} + a_j + \xi_{ij}. \tag{5}$$

Essentially, π represents the denominator of the grouping estimator, where Equation (5) corrects for the presence of covariates. Estimating Equations (4) and (5) with control variables via 2SLS will produce consistent estimates of the ACR. Robust standard errors, where we allow for an unrestricted correlation structure in i, are used for statistical inference.

4.3 Experiment Execution[13]

The study has two experimental arms: an incentivized group consisting of students who were offered an incentive to participate in PASS and a non-incentivized group. While PASS is the treatment or intervention of interest, members of both groups were allowed to attend PASS.

The study sample was drawn from students enrolled in any of 14 first-year courses at the main campus of the University of Wollongong in Australia over three semesters: Autumn 2014, Spring 2014, and Autumn 2015. Initially, it was planned to conduct the experiment over just two semesters. A decision was made to extend the study to a third semester after it became apparent

[13] Before experiment initiation, ethics approval was obtained from the Social Science Human Research Ethics Committee of the University of Wollongong. The experiment was also registered with the randomized controlled trials registry of the American Economic Association.

that statistical power was compromised by a lower-than-expected first-stage effect of the incentives on PASS participation. The courses cover a broad range of subjects and involves four of the five faculties within the university.[14] Six of the fourteen subjects were from the Faculty of Business. This was motivated by the fact that some of the funding was provided by this academic unit. Students were made aware of the study through a presentation delivered on the first lecture in each nominated subject.[15]

Assignment into the incentivized and non-incentivized groups was based on a computer-generated random number drawn from a uniform distribution, with stratification by subject, aboriginality, broad funding (domestic or international), and sex. Group assignment was communicated to students via short-message service (SMS) and email. To the extent possible, this communication came immediately after the first lecture. Randomization was conducted separately for each semester. Some students were enrolled in two or more subjects under study. These students were randomly allocated into an experimental group for each subject in which they were enrolled. In other words, the randomization units are specific individual–subject combinations.

Members of the incentivized group were encouraged to attend the PASS program with a chance to win a near-cash incentive. The incentive was substantially larger in the 2014 semesters. Across nine subjects in 2014, the prizes consisted of 50 gift certificates worth AUD 1,000 valid at a number of retail outlets, for a total prize value of AUD 50,000. In Autumn 2015, only five such gift certificates were offered across five subjects for a total value of AUD 5,000. Students in the incentivized group who attended PASS at least five times over five separate weeks were eligible for the prize draw; those who attended at least eight times over eight separate weeks were allocated two entries in the prize draw, thereby doubling their chances of winning. Although it is possible to attend more than one PASS session for the same week, these "repeat" attendances were counted only once. Several reminder emails were sent out to all students in the incentivized group throughout each semester. The prize draws were conducted during the exam period of each semester. Winners were notified via SMS and email.

[14] The five subjects in Autumn 2014 are Accounting Fundamentals in Society (ACCY111), Statistics for Business (COMM121), Engineering Materials (ENGG153), General Mathematics 1A (MATH151), and Introduction to Anatomy and Physiology I (SHS 111). The four subjects in Spring 2014 are Statistics for Business (COMM121), Macroeconomic Essentials for Business (ECON101), Theory Design and Statistics in Psychology (PSYC123) and Molecules, Cells, and Organisms (BIOL103). In Autumn 2015, the included subjects were Accounting Fundamentals in Society (ACCY111), Statistics for Business (COMM121), Foundations of Engineering Mathematics (MATH141), General Mathematics 1A (MATH151), and Introduction to Anatomy and Physiology I (MEDI111/EDPS101).

[15] In Spring 2014, a second presentation was also delivered in the Week 2 lectures of each subject.

A participant information sheet with detailed information on the research purpose, investigators, method and demands on participants, risk disclosure, funding sources, and ethics approval was made available for download.[16] A special email account was set up at UOW to communicate directly with students. Throughout the duration of the study, the individual students' identities were never known to the principal investigators except when the prizes were given out in a public ceremony.

Consent to participate was presumed, but students were given an opportunity to opt out of the study at any point over its duration by sending an email to the designated email address. No student chose to opt out of the study. The full study population consists of 6,954 student–subject observations. Of these, 969 (14 percent) do not have a final grade assigned for the relevant subject, leaving 5,985 observations in the estimation sample (1,993 in Autumn 2014, 1,729 in Spring 2014, and 2,263 in Autumn 2015) consisting of 4,397 students. Attrition is explained by students having the option to withdraw from a given subject without academic penalty up to approximately two-thirds through each semester's teaching period. Attrition is similar for the incentivized group (13.4 percent) and the non-incentivized group (14.5 percent). The difference in the probability of attrition between these groups is not statistically significant, whether baseline control variables are excluded ($p = 0.226$) or included ($p = 0.794$).

Table 1 shows descriptive statistics from the estimation sample, which includes all observations with non-missing final grades. Table 2 shows the results of balance tests on baseline characteristics between the incentivized and non-incentivized groups. The difference between groups is not statistically significant on any of these characteristics.

5 Results

In this section, we present the estimation results. We begin with the first-stage regressions that show the strength of the incentive to induce the desired behavioral changes. This is followed by the OLS results that do not take into account the selection bias. Finally, we discuss the results from the instrumental-variable approach, where we use the randomized design of the study for consistent estimation of the treatment effect.

5.1 First-Stage Results

Table 3 shows first-stage regression results which summarize the effectiveness of the incentive at inducing PASS participation. The upper panel shows results for the full estimation sample, with and without the inclusion of control variables. Given the results of the balance tests, it is

[16] The participant information sheet is included in the appendix of this paper and can be downloaded from http://www.uow.edu.au/~siminski/PASS_research.pdf.

not surprising that the inclusion of controls does not greatly affect the estimates, which show that the incentive clearly increased PASS participation. With the inclusion of controls, the estimated first-stage effect is 0.474, almost half of one PASS session, which is equal to 19 percent of the control group mean. The first-stage F-statistic is 24.85, well above conventional thresholds of instrument relevance. With a sample of 3,013 in the incentivized group, this translates to 1,428 additional sessions of PASS attended by students in that group. The total value of incentives provided was AUD 55,000. Each additional hour of PASS participation therefore cost AUD 38.51.

The lower panels of Table 3 show first-stage results for various sub-groups. The two columns on the right show the difference in first-stage estimates between sub-groups and the standard errors of those differences. The point estimates suggest that the incentive had a larger effect on attendance for females (0.57) than for males (0.37), but the difference is not statistically significant. The first-stage effect does not differ greatly by age, by ATAR[17] score, or between domestic and international students. The largest apparent difference is by socioeconomic status. The estimated first-stage effect for students from low-SES areas is 0.891 sessions, more than twice as large as for the remainder of the sample (0.422). While the difference is not statistically significant ($p = 0.134$), it is reasonable that students from low-SES areas may respond more to such incentives.

Perhaps the most interesting finding in the first-stage results is the comparison between Autumn 2014 and Autumn 2015, for which the estimates are similar (0.459 versus 0.439). This is despite considerably larger (approximately five times larger) incentives being offered in 2014. For Spring 2014, the corresponding effect was slightly larger (0.555). In an attempt to isolate the role of the size of incentives, however, the discussion is focused on the Autumn semesters due to other differences between Autumn and Spring.[18] The implied cost (in terms of the value of the prizes only) of inducing one hour of PASS participation is approximately AUD 50 in Autumn 2014 and AUD 10 in Autumn 2015.

[17] The Australian Tertiary Admission Rank (ATAR) is used by most universities in Australia for admission purposes.

[18] Since Autumn is the first semester of the academic year and the included subjects are all first-year subjects, the study population drawn from the Autumn semesters is predominantly comprised of students in their first semester of university study. The composition of subjects included in our experiment is also similar in the two Autumn semesters (three subjects were included in both of the Autumn semesters). Further, our strategy for informing students about our experiment was more intensive in Spring. In Spring 2014, we visited each class in the first and second weeks, while in Autumn of each year, we visited only in the first week. In Spring 2014, we also provided students in the incentivized group an approximate probability of winning a prize ("1 in 10"). All of these factors may have affected the size of the first-stage effects.

This is an encouraging finding in the sense that the size of the incentive (and hence the total cost of RED experiments) seems to not be a major factor in the design of effective RED experiments. Similar future studies could be designed in ways that increase statistical power through methods that do not necessarily involve greater cost. Including a larger sample size (perhaps through multi-institutional collaboration) would seem to be a better option than increasing the incentive pool. More generally, the design of effective incentives in this context warrants further research.

Table 4 further examines the nature of the first-stage inducement effect. For these results, the dependent variable is binary (indicating PASS participation in bands): zero sessions, 1–4 sessions, 5–7 sessions, and 8+ sessions. Given the structure of the incentive, we hypothesized that students in the incentivized group may be more likely to bunch around the attendance thresholds: five and/or eight PASS sessions. The results suggest that students in the incentivized group were less likely to attend PASS zero times (–0.042) and more likely to attend 8+ times (0.046). The probability of attending 1–4 times or 5–7 times was unaffected. A possible explanation is that students who were incentivized to try a PASS session continued attending for the majority of the semester.

5.2 OLS Results

Table 5 shows OLS results. The upper panel shows results for the full sample with and without controls. In both specifications, one hour of PASS attendance is associated with a 0.07 increase in standardized grades associated with small standard errors. Since the coefficients on PASS attendance do not change with the inclusion of very detailed control variables, there is no evidence of positive (or negative) selection on observables into participation into PASS. Nevertheless, self-selection into PASS may be associated with unobserved determinants of grades.

The lower panel of Table 5 shows OLS results for various subgroups. Overall, the results do not vary greatly between subgroups. However, there is clear evidence that the association is larger for older students versus younger students, and for low-ATAR versus high-ATAR students. There is also suggestive evidence that the association is larger for low-SES students.

5.3 2SLS Results

Table 6 shows 2SLS results for the full sample. With control variables, the point estimate suggests that one hour of PASS improves grades by 0.065 standard deviations, which is very similar to the OLS estimate. This estimate is not statistically significant ($p = 0.167$), but it is subject to a large standard error. Assuming a constant effect for each hour of PASS, this suggests that an average quantity of PASS sessions (6.25 hours among those who attend PASS) increases marks by 0.41 standard deviations, equivalent to 7.9 marks on the non-standardized 100-point scale. But the corresponding confidence interval ranges from –0.17 to 0.98 standard deviations, or –3.3 to 19.1 marks.

In our pre-analysis plan, we documented our intention to study heterogeneity of effects between various sub-groups (by age, sex, domestic/international status, high school grades and socioeconomic status), but the relatively weak first-stage effects of the incentive have been prohibitive. As shown in Table 3, the first-stage F-statistics do not exceed 10 for any pair of sub-groups in the pre-analysis plan, suggesting that comparisons between these sub-groups are not appropriate.[19] If we ignore the weak first-stage relationships, none of the differences between sub-groups in 2SLS effects are statistically significant at the 5-percent level. That said, the point estimates sometimes do vary considerably between subgroups, so we have chosen to not show those to avoid misinterpretation.

We do, however, show one comparison between sub-groups: between students who are enrolled in their first-ever semester at the University of Wollongong and other students. The first-stage F-statistic exceeds 10 for both of these subgroups. The 2SLS results suggest that effect of PASS may be considerably larger for the new students. The difference in the estimates for new students versus others (0.201 standard deviation) is borderline statistically significant ($p = 0.051$).

We stress that the comparison between new students and others was not included in our pre-analysis plan, and so it should be treated as an exploratory result rather than a confirmatory result. In other words, we did not explicitly hypothesize a difference between these sub-groups *ex ante*. Nevertheless, this seems a plausible result, given the issues around transitioning into a university environment, including the more independent and self-directed study skills and time management required in tertiary study. Many students also leave home to attend university, thereby losing the structured and supportive environment which some parents can provide. Students at university are also faced with making new social connections and friends within much larger cohorts than they have previously encountered. Programs such as PASS aim to address many of these issues. The systematic review conducted by Dawson et al. [2014] commented on earlier studies, including Ogden et al. [2003], Court and Molesworth [2008], van der Meer and Scott [2009], and Bronstein [2008], which suggested that PASS assisted in the areas of students developing effective study skills and exam preparation techniques, understanding course expectations, feeling more supported, enhancing social relationships and personal well-being, and reducing anxiety.

[19] For example, the first-stage F-statistic is 8.8 (i.e., less than 10) for males, preventing meaningful comparisons by gender. Similarly, the F-statistics are less than 10 for international students, for older students (aged 20+), low-ATAR students, regional students, low-SES students, and regional students.

In contrast to the discrepancy between new students and others in the 2SLS results, the OLS analysis suggests that the effects of PASS are very similar for new students and other students. A possible explanation for the discrepancy between OLS and 2SLS is the nature of selection into PASS. The divergent results would be explained by negative selection on unobserved characteristics into PASS among new students, combined with positive selection among other students. Under this interpretation, students in their first semester of study are more likely to select into PASS if they have unobserved characteristics associated with lower university grade performance (e.g., if they are particularly likely to struggle with the transition from high school to the university environment). Conversely, there may be positive selection into PASS among students who are not in their first semester. This would be the case if, for example, those students who have the most available time and commitment to invest heavily into study are most likely to choose PASS as part of their overall study regime.

6 Conclusion

There is a widely recognized need among stakeholders to ensure that policymaking is evidence-based, relying on rigorous scientific methods to generate new knowledge about the impacts of certain policy instruments. With respect to the hierarchy of evidence, knowledge generated from randomized controlled trials is widely considered to be the most credible. Yet there exists is a substantial divide between the ability of RCTs to generate credible evidence and the socio-political realities associated with successfully conducting an RCT.

One reason for the paucity of experimental evidence is that deploying an RCT to analyze many public policy interventions may not overcome ethical hurdles. In particular, the outright denial of access to specific services is difficult to justify, especially when there is putative evidence on its effectiveness or when the political constraints do not allow for creating an authentic control group. Randomly assigning eligibility and compelling treatment in this case simply for the purposes of evaluation are unlikely to pass ethical review (or, indeed, voter review at the ballot box).

Our proposed alternative is to conduct an RED experiment, where the evaluator can generate an exogenous variation in the probability of program participation by incentivizing a randomly selected subset of the eligible population to participate. This design allows us to estimate the program impact using a conventional instrumental-variable approach, which we estimate by two-stage least squares to recover the average causal response. This kind of experimental design does not suffer from many of the challenges associated with RCTs, particularly the denial of access to the voluntary and universal program. As such, REDs should form part of the evaluation toolkit available to researchers, along with RCTs, quasi-experimental approaches, and other observational studies.

However, an RED creates its own challenges that researchers should attempt to overcome. Two are associated with the fact that the estimation approach in an RED is fundamentally an instrumental-variable approach. Therefore, it must satisfy the requirements for an IV estimator to be consistent. Two are worth highlighting here. First, the instrument must not have a direct causal relationship with the outcome variable (i.e., the exclusion restriction). Second, the instrument must be strongly associated with the endogenous variable (i.e., the first-stage relationship must be substantial). Otherwise, the bias that arises out of the IV approach may even be worse than the bias associated with the OLS estimator.

We use an RED to examine supplemental instruction or peer-assisted study sessions, which are integrated into tertiary education in many higher-educational institutions around the world. Its proponents claim that it contributes to the academic success of students, as well as to other relevant student outcomes, such as satisfaction with university life. To our knowledge, there is scant experimental studies on the effectiveness of PASS or SI, no quasi-experimental approach has been used to evaluate any PASS program, and there is no natural experiment available which lends itself to a similar analysis. Our study attempts to fill this gap by deploying a randomized-encouragement-design experiment to estimate the impact of PASS on academic achievement as measured by the final marks. We find that there is suggestive evidence that PASS contributes to higher grades, and that this contribution can potentially be large, but that this is rather imprecisely estimated in our experiment.

At this stage, it is important to note that the PASS program could potentially have impacts over and above the final marks of students. In particular, we recognize that PASS may contribute positively to student life. A number of first-semester students may not know anyone from the university. Attending PASS sessions can expand a student's social network, thereby possibly accelerating the acclimatization process to university life. Other skills which can be useful in the present and then later in life, such as working with a group of peers or learning to be assertive, may be acquired over PASS sessions. These are beyond the scope of the present study, but we are cognizant of the fact that these are relevant to the holistic development of the student within the context of higher education.

In our application, the first-stage relationship was quite weak, and this led to a very low statistical power to discriminate between a potentially large effect or none at all. Since the instrument is weak, inference is problematic because the standard error associated with the 2SLS estimate of the effect increases as the participation rate decreases. Our application of the RED can serve as a lesson to future researchers. Although the endogeneity of the treatment or program indicator is a serious econometric concern, the incentive design is perhaps just as important. Estimation and inference become a precarious exercise for all the reasons identified in the weak-instruments literature (e.g., Bound, Jaeger, and Baker [2005]) when the incentive simply does not have sufficient power to induce eligible program participants to subsequently enroll in the program.

Another lesson learned in our specific application is that the size of the incentive for the kind of behavior we wanted to promote did not seem to matter. For roughly similar numbers of eligible students across semesters, the decision to devote either AUD 50,000 or AUD 5,000 to promote PASS attendance did not generate any notable difference in the likelihood of attending the PASS program. This additionally demonstrates that there is much more to be learned in the incentive space, especially the design of effective incentives for a specific population and a target behavior.

Finally, the RED also inherits the causal interpretations from the IV framework. This means that the parameter that the 2SLS estimator recovers is a kind of local average treatment effect or, in the case of a continuous treatment, the average causal response. The subpopulation over which the treatment effect is averaged is also contingent on the kind of instrument (i.e., incentive) that the researcher decides to use. The LATE/ACR may or may not be the ultimate parameter that policy evaluators are interested in since this would be context-specific. However, it is important to emphasize that there are a number of treatment effects that can be estimated within the IV sphere, and not all of them may be informative for the purposes of the evaluation or for program stakeholders.

References

Allan, Bradley M. and Roland G. Fryer, Jr. [2011]. “The Power and Pitfalls of Education Incentives”, The Hamilton Project Discussion Paper No. 2011-07.

Angrist, Joshua D. [2004]. “Treatment Heterogeneity in Theory and in Practice”, *The Economic Journal* 11(494):1167–1201.

Angrist, Joshua, Eric Bettinger, and Michael Kremer [2006]. “Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia”, *The American Economic Review* 96(3):847–862.

Angrist, Joshua D. and Guido W. Imbens [1995]. “Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity”, *Journal of the American Statistical Association* 90(430):431–442.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin [1996]. “Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association* 91(434):444–455.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos [2009]. “Incentives and Services for College Achievement: Evidence from a Randomized Trial”, *American Economic Journal: Applied Economics* 1(1):136–163.

Angrist, Joshua and Victor Lavy [2009]. “The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial”, *The American Economic Review* 99(4):1384–1414.

Angrist, Joshua, Philip Oreopoulos, and Tyler Williams [2014]. “When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards”, *Journal of Human Resources* 49(3):572–610.

Angrist, Joshua D. and Jörn-Steffen Pischke [2009]. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press.

Arendale, D. [2002]. History of Supplemental Instruction (SI): Mainstreaming of developmental education. In D.B. Lundell and J. Higbee (Eds.), *Histories of developmental education* (pp. 15–27). Minneapolis: Center for Research on Developmental Education and Urban Literacy, General College, University of Minnesota.

Baicker, Katherine and Amy Finkelstein [2011]. “The Effects of Medicaid Coverage – Learning from the Oregon Experiment”, *The New England Journal of Medicine* 365:683–685.

Barnard, John, Jiangtao Du, Jennifer L. Hill, and Donald B. Rubin [1998]. “A Broader Template for Analyzing Broken Randomized Experiments”, *Sociological Methods Research* 27(2):285–317.

Barnard, John, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. Rubin [2003]. “Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City”, *Journal of the American Statistical Association* 98(462):299–311.

Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock [2012]. “Paying for Performance: The Education Impacts of a Community College Scholarship Program for Low-income Adults”, Federal Reserve Bank of Chicago Working Paper No. 2009-13.

Beam, Emily A. [forthcoming]. “Do job fairs matter? Experimental evidence on the impact of job-fair attendance”. *Journal of Development Economics*.

Blanc, R., L. DeBuhr, and D.C. Martin [1983]. “Breaking the Attrition Cycle: The Effects of Supplemental Instruction on Undergraduate Performance and Attrition”, *Journal of Higher Education* 54(1):80–89.

Bloom, Howard S. [1984]. “Accounting for No-Shows in Experimental Evaluation Designs”, *Evaluation Review* 8(2):225–246.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng’ang’a, and Justin Sandefur [2013]. “Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education”, Center for Global Development Working Paper No. 321. URL: http://goo.gl/02hzmA.

Boucher, S. and C. Mullally [2010]. “Case study: Evaluating the impact of index insurance on cotton farmers in Peru” in P. Winters, L. Salazar, and A. Maffioli (eds.), *Designing Impact Evaluations for Agricultural Projects*. IDB Technical Note IDB-TN-198. Washington, D.C.: Inter-American Development Bank.

Bound, John, David A. Jaeger, and Regina M. Baker [1995]. “Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak”, *Journal of the American Statistical Association* 90(430):443–450.

Bronstein, S.B. [2008]. “Supplemental instruction: Supporting persistence in barrier courses.” *Learning Assistance Review* 13:31–45.

Bruhn, Miriam and David McKenzie [2009]. “In Pursuit of Balance: Randomization in Practice in Development Field Experiments”, *American Economic Journal: Applied Economics* 1(4):200–232.

Cha, Paulette and Reshma Patel [2010]. “Rewarding Progress, Reducing Debt: Early Results from the Performance-Based Scholarship Demonstration in Ohio”, MDRC, October 2010. Available at SSRN: http://sssrn.com/abstract=1786941.

Court, S. and M. Molesworth [2008]. “Course-specific learning in peer assisted learning schemes: A case study of creative media production courses”. *Research in Post-Compulsory Education* 13:123–134.

Dawson, P., van der Meer, J., Skalicky, J., & Cowley, K. [2014]. “On the Effectiveness of Supplemental Instruction: A Systematic Review of Supplemental Instruction and Peer-Assisted Study Sessions Literature Between 2001 and 2010”. *Review of Educational Research* 84(4):609–639.

Deci, Edward L., Richard Koestner, and Richard M. Ryan [1999]. “A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation”, Psychological Bulletin 125(6):627–668.

Duflo, Esther, Rachel Glennerster, and Michael Kremer [2008]. “Using Randomization in Development Economics Research: A Toolkit”. In T. Paul Schultz and John Strauss, editors, *Handbook of Development Economics*, Vol. 4. Amsterdam and New York: North Holland.

Fryer, Roland G., Jr. [2011]. “Financial Incentives and Student Achievement: Evidence from Randomized Trials”, *The Quarterly Journal of Economics* 126(4):1755–1798.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M.J. Vermeersch [2011]. *Impact Evaluation in Practice*. Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Gibson, John, David McKenzie, and Steven Stillman [2011]. “The Impacts of International Migration on Remaining Household Members: Omnibus Results from a Migration Lottery Program”, *The Review of Economics and Statistics* 93(4):1297–1318.

Heckman, James J. [1979]. “Sample Selection Bias as a Specification Error”, *Econometrica* 47(1):153–161.

Heckman, James J. and Edward Vytlacil [2005]. “Structural Equations, Treatment Effects, and Econometric Policy Evaluation”, *Econometrica* 73(3):669–738.

Hirano, Keisuke, Guido W. Imbens, Donald B. Rubin, and Xiao-Hua Zhou [2000]. “Assessing the effect of an influenza vaccine in an encouragement design”, *Biostatistics* 1(1):69–88.

Holland, Paul W. [1986]. “Statistics and Causal Inference”, *Journal of the American Statistical Association* 81(396):945–960.

Holland, Paul W. [1988]. “Causal Inference, Path Analysis, and Recursive Structural Equations Models”, *Sociological Methodology* 18(1):449–484.

Huang, Tairan K., Matthew P.J. Pepper, Corinne L. Cortese, and Sally Rogan [2013]. “Faculty and academic staff perceptions, experiences and expectations of the PASS Program: A case study”, *Journal of Peer Learning* 6(1). Available at http://ro.uow.edu.au/ajpl/vol6/iss1/10.

Imai, Kosuke, Dustin Tingly, and Teppei Yamamoto [2013]. “Experimental designs for identifying causal mechanisms”, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(1):5–51.

Imbens, Guido W. and Joshua D. Angrist [1994]. “Identification and Estimation of Local Average Treatment Effects”, *Econometrica* 62(2):467–475.

Imbens, Guido W. and Jeffrey M. Wooldridge [2009]. “Recent Developments in the Econometrics of Program Evaluation”, *Journal of Economic Literature* 47(1):5–86.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman [2001]. “Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment”, *The Quarterly Journal of Economics* 116(2):607–654.

Kremer, Michael, Edward Miguel, and Rebecca Thornton [2009]. “Incentives to Learn”, *The Review of Economics and Statistics* 91(3):437–456.

Kuh, G. [2003] “What We’re Learning About Student Engagement from NSSE”, *Change* 35(2):24–32.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw [2010]. “The Effect of Financial Rewards on Students’ Achievement: Evidence from a Randomized Experiment”, *Journal of the European Economic Association* 8(6):1243–1265.

Lewis, Don, Martin O’Brien, Sally Rogan, and Brett Shorten [2005]. “Do Students Benefit From Supplemental Education: Evidence From a First-Year Statistics Subject in Economics and Business”, University of Wollongong Economics Working Paper Series No. 05-21.

Longfellow, E., S. May, L. Burke, and D. Marks-Maran [2008]. “‘They had a way of helping that actually helped’: A case study of a peer-assisted learning scheme”. *Teaching in Higher Education* 13:93–105.

Martin, D. [2009]. “Foreword”, *Australasian Journal of Peer Learning* 1(1):3–5.

McKenzie, David and Berk Özler [2014]. “Quantifying some of the impacts of economics blogs”, *Economic Development and Cultural Change* 62(3):567–597.

Ogden, P., D. Thompson, A. Russell, and C. Simons [2003]. “Supplemental instruction: Short- and long-term impact”. *Journal of Development Education* 26(3):2–8.

Parkinson, M. [2009]. “The effect of peer assisted learning support (PALS) on performance in mathematics and chemistry”. *Innovations in Education and Teaching International*, 46, 381–392.

Pascarella, E.T. and P.T. Terenzini [2005]. How College Affects Students: A Third Decade of Research. Volume 2. *Jossey – Bass, An Imprint of Wiley* 848.

Powers, Donald E. and Spencer S. Swinton [1984]. “Effects of Self-Study for Coachable Test Item Types”, *Journal of Educational Psychology* 76(2):266–278.

Schultz, T. Paul [2004]. “School subsidies for the poor: evaluating the Mexican Progresa poverty program”, *Journal of Development Economics* 74(1):199–250.

SI Staff Manual from UMKC [2005]. *Supplemental Instruction Supervisor Manual*. University of Missouri–Kansas City.

Siminski, Peter and Simon Ville [2012]. “I Was Only Nineteen, 45 Years Ago: What Can We Learn from Australia’s Conscription Lotteries?”, *Economic Record* 88(282):351–371.

Slavin, Robert E. [2010]. “Can financial incentives enhance educational outcomes? Evidence from international experiments”, *Educational Research Review* 5(1):68–80.

Small, Dylan S. and Paul R. Rosenbaum [2008]. “War and Wages: The Strength of Instrumental Variables and Their Sensitivity to Unobserved Biases”, *Journal of the American Statistical Association* 103(483):924–933.

Strotz, Robert H. [1955]. “Myopia and Inconsistency in Dynamic Utility Maximization”, *The Review of Economic Studies* 23(3):165–180.

TenHave, Thomas R., James Coyne, Mark Salzer, and Ira Katz [2003]. "Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial", *General Hospital Psychiatry* 25(2):115–123.

van den Meer, J., S. Scott, and T. Neha [2010]. "Retention of first-year Māori students at university". *MAI Review* (2). Available at http://www.review.mai.ac.nz/index.php/MR/article/view/309.

Wald, Abraham [1940]. "The Fitting of Straight Lines if Both Variables Are Subject to Error", *Annals of Mathematical Statistics* 11(3):284–300.

West, Stephen G., Naihua Duan, Willo Pequegnat, Paul Gaist, Don C. Des Jarlais, David Holtgrave, José Szapocznik, Martin Fishbein, Bruce Rapkin, Michael Clatts, and Patricia Dolan Mullen [2008]. "Alternatives to the Randomized Controlled Trial", *American Journal of Public Health* 98(8):1359–1366.

Zhou, Xiao-Hua and Sierra M. Li [2006]. "ITT analysis of randomized encouragement design studies with missing data", *Statistics in Medicine* 25(16):2737–2761.

Tables

Table 1 Descriptive Statistics

Variable	Mean	Std. Dev.	Non-missing Observations
Final Grade, Standardized	0.00	1.00	5,985
Incentivized Group	0.50	0.50	5,985
PASS sessions attended	2.67	3.98	5,985
PASS sessions attended = 0	0.57	0.49	5,985
Male indicator	0.55	0.50	5,985
Age	20.31	3.93	5,985
International-student indicator	0.13	0.34	5,985
Equity-student indicator	0.28	0.45	5,985
Aboriginality indicator	0.01	0.09	5,985
Regional indicator	0.20	0.40	5,985
SES Area == High	0.20	0.40	5,985
SES Area == Low	0.12	0.32	5,985
SES Area == Medium	0.45	0.50	5,985
SES Area == missing	0.24	0.42	5,985
Student's first semester at UOW indicator	0.51	0.50	5,985
Original ATAR	76.08	11.92	4,157
Indicator for missing ATAR	0.31	0.46	5,985
High school math grade, standardized	0.00	1.00	3,537
High school math level: Basic	0.29	0.45	5,985
High school math level: Standard	0.18	0.38	5,985
High school math level: Extension 1	0.11	0.31	5,985
High school math level: Extension 2	0.02	0.14	5,985
High school math grade: missing	0.41	0.49	5,985
Semester==Aut_2014	0.33	0.47	5,985
Semester==Aut_2015	0.38	0.48	5,985
Semester==Spr_2014	0.29	0.45	5,985
Subject_code==ACCY111	0.22	0.41	5,985
Subject_code==BIOL103	0.09	0.28	5,985
Subject_code==COMM121	0.22	0.41	5,985
Subject_code==ECON101	0.08	0.28	5,985
Subject_code==EDPS101	0.01	0.10	5,985
Subject_code==ENGG153	0.06	0.24	5,985
Subject_code==MATH141	0.06	0.24	5,985
Subject_code==MATH151	0.05	0.23	5,985
Subject_code==MEDI111	0.08	0.27	5,985
Subject_code==PSYC123	0.04	0.21	5,985
Subject_code==SHS 111	0.07	0.26	5,985

Table 2 Balance Tests for Estimation Sample

Baseline Characteristics	Incentiv-ized Group Mean	Control Group Mean	Difference	P-value for difference
Male indicator	0.551	0.542	0.009	0.490
Age	20.351	20.275	0.076	0.448
International-student indicator	0.132	0.133	-0.001	0.926
Equity-student indicator	0.284	0.283	0.001	0.923
Aboriginality indicator	0.009	0.009	0.001	0.817
Regional indicator	0.198	0.207	-0.009	0.399
SES==H	0.193	0.204	-0.011	0.279
SES==L	0.123	0.113	0.010	0.236
SES==M	0.449	0.447	0.002	0.883
SES==.	0.235	0.236	-0.001	0.935
Student's first semester at UOW indicator	0.509	0.511	-0.002	0.879
Original ATAR	76.052	76.111	-0.059	0.874
Indicator for missing ATAR	0.312	0.299	0.013	0.291
High school math grade, standardized	0.001	-0.001	0.003	0.933
High school math level: Basic	0.290	0.284	0.006	0.584
High school math level: Standard	0.174	0.179	-0.004	0.653
High school math level: Extension 1	0.105	0.108	-0.002	0.755
High school math level: Extension 2	0.023	0.019	0.003	0.358
High school math grade: missing	0.408	0.410	-0.003	0.817
Semester==Aut_2014	0.334	0.332	0.002	0.841
Semester==Aut_2015	0.377	0.379	-0.002	0.905
Semester==Spr_2014	0.288	0.289	-0.001	0.935
Subject_code==ACCY111	0.218	0.218	0.000	0.974
Subject_code==BIOL103	0.089	0.088	0.000	0.951
Subject_code==COMM121	0.219	0.220	-0.001	0.925
Subject_code==ECON101	0.083	0.082	0.001	0.866
Subject_code==EDPS101	0.012	0.010	0.002	0.492
Subject_code==ENGG153	0.065	0.062	0.003	0.617
Subject_code==MATH141	0.063	0.062	0.001	0.855
Subject_code==MATH151	0.052	0.056	-0.003	0.561
Subject_code==MEDI111	0.082	0.080	0.002	0.787
Subject_code==PSYC123	0.042	0.046	-0.005	0.385
Subject_code==SHS 111	0.074	0.075	-0.001	0.880
N	3013	2972		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 First-Stage Results – Dependent Variable is Number of PASS sessions attended

	Full Sample (without controls)		**Full Sample (with controls)**			
Incentive	0.464***	(0.102)	0.474***	(0.095)		
N / first-stage F	5985	20.630	5985	24.853		
	Sub-Group Analyses Included in Pre-Analysis Plan (with controls)					
	Male		Female		Difference	
Incentive	0.374***	(0.126)	0.572***	(0.145)	-0.198	(0.192)
N / first-stage F	3269	8.830	2716	15.489		
	Domestic		International		Difference	
Incentive	0.461***	(0.102)	0.593**	(0.263)	-0.131	(0.282)
N / first-stage F	5192	20.495	793	5.084		
	Young (< 20)		Older (20+)		Difference	
Incentive	0.497***	(0.127)	0.444***	(0.143)	0.054	(0.191)
N / first-stage F	3517	15.377	2468	9.585		
	Low ATAR		High ATAR		Difference	
Incentive	0.484***	(0.154)	0.584***	(0.162)	-0.100	(0.223)
N / first-stage F	2018	9.859	2139	13.040		
	Non-Regional		Regional		Difference	
Incentive	0.522***	(0.105)	0.328	(0.223)	0.194	(0.246)
N / first-stage F	4771	24.802	1214	2.172		
	Not Low SES		Low SES		Difference	
Incentive	0.422***	(0.100)	0.891***	(0.296)	-0.469	(0.313)
N / first-stage F	5276	17.717	709	9.062		

	Non-Equity		Equity		Difference	
Incentive	0.464***	(0.109)	0.535***	(0.190)	-0.071	(0.219)
N / first-stage F	4288	18.060	1697	7.968		

Additional Sub-Group Analyses (with controls)

	Large Incentive (Autumn 2014)		Small Incentive (Autumn 2015)		Difference	
Incentive	0.459**	(0.183)	0.439***	(0.151)	0.020	(0.237)
N / first-stage F	1993	6.275	2263	8.479		
	First Semester at University		Not First Semester at University		Difference	
Incentive	0.503***	(0.146)	0.438***	(0.122)	0.065	(0.190)
N / first-stage F	3053	11.940	2932	13.015		

Full controls are included in each regression. Robust standard in parentheses, clustered on student.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 First-Stage Results by Treatment Intensity – Dependent Variable is PASS Attendance Band (Binary)

	(1) attendance = 0	(2) attendance = 1 - 4	(3) attendance = 5 - 7	(4) attendance = 8+
Incentive	-0.042^{***}	-0.009	0.005	0.046^{***}
	(0.012)	(0.010)	(0.007)	(0.010)
N	5985	5985	5985	5985

Full controls are included in each regression. Robust standard in parentheses, clustered on student.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 OLS Results - Dependent Variable is Standardized Final Grade

	Full Sample (without controls)		**Full Sample (with controls)**			
ATTENDANCE	0.072***	(0.003)	0.069***	(0.003)		
N	5985		5985			
Sub-Group Analyses Included in Pre-Analysis Plan (with controls)						
	Male		Female		Difference	
ATTENDANCE	0.073***	(0.004)	0.064***	(0.004)	0.009	(0.006)
N	3269		2716			
	Domestic		International		Difference	
ATTENDANCE	0.068***	(0.003)	0.074***	(0.009)	-0.006	(0.009)
N	5192		793			
	Young (< 20)		Older (20+)		Difference	
ATTENDANCE	0.059***	(0.003)	0.088***	(0.005)	-0.030***	(0.006)
N	3517		2468			
	Low ATAR		High ATAR		Difference	
ATTENDANCE	0.072***	(0.005)	0.050***	(0.004)	0.022***	(0.006)
N	2018		2139			
	Non-Regional		Regional		Difference	
ATTENDANCE	0.071***	(0.003)	0.062***	(0.006)	0.009	(0.007)
N	4771		1214			
	Not Low SES		Low SES		Difference	
ATTENDANCE	0.066***	(0.003)	0.081***	(0.008)	-0.015*	(0.008)
N	5276		709			

	Non-Equity		Equity		Difference	
ATTENDANCE	0.068***	(0.004)	0.070***	(0.005)	-0.002	(0.006)
N	4288		1697			

Additional Sub-Group Analyses

	Large Incentive (Autumn 2014)		Small Incentive (Autumn 2015)		Difference	
ATTENDANCE	0.075***	(0.004)	0.070***	(0.004)	0.006	(0.006)
N	1993		2263			

	First Semester at University		Not First Semester at University		Difference	
ATTENDANCE	0.068***	(0.003)	0.069***	(0.004)	-0.000	(0.006)
N	3053		2932			

Robust Standard errors in parentheses, clustered on student.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6 2SLS Results - Dependent Variable is Standardized Final Grade

	Full Sample (without controls)		**Full Sample (with controls)**			
ATTENDANCE	0.049	(0.053)	0.065	(0.047)		
Observations	5985		5985			
			Additional Sub-Group Analysis			
	First Semester at University		Not First Semester at University		Difference	
ATTENDANCE	0.153**	(0.063)	-0.048	(0.081)	0.201*	(0.103)
Observations	3053		2932			

Appendix

Participant info sheet

Lecture slides

PARTICIPANT INFORMATION SHEET FOR STUDENTS

TITLE: *The Causal Effects of the Peer Assisted Study Sessions (PASS) on Educational Outcomes*

PURPOSE OF THE RESEARCH

This is an invitation to participate in a study conducted by researchers at the University of Wollongong. This project aims to estimate the effects of the PASS (Peer Assisted Study Sessions) program on educational outcomes of students, using a quasi-experimental impact evaluation technique known as randomised encouragement design. A random set of students selected from COMM121, ACCY111, ENGG153, SHS111 and MATH151 (in Autumn Session 2014) and certain Spring Session subjects (to be announced) will be offered a substantial incentive to participate in the PASS program – which should raise their PASS participation relative to the control group. Students in the treatment group will gain eligibility to win one of 50 Coles Group & Myer Gift Cards worth $1,000 each. A draw will be held during the exam periods in Autumn and Spring semesters, 2014.

INVESTIGATORS

Assoc Prof Peter Siminski (Faculty of Business) web page

Assoc Prof Leona Tam (Faculty of Business) web page

Ms Sally Rogan (Director of Peer Learning) web page

Dr Alfredo Paloyo (Economics) web page

The study is funded by the Australian Government's Higher Education Participation and Partnerships Program (HEPPP). Associate Professors Peter Siminski and Leona Tam and Dr Paloyo have responsibility for the study design and have no current or previous relationship with the PASS program, other than by virtue of being employed by UOW. As Director of Peer Learning, Ms Sally Rogan (Director of Peer Learning) has overall responsibility for the UOW PASS program.

METHOD AND DEMANDS ON PARTICIPANTS

There will be no primary data collection. We can foresee no participant burden. The research team will analyse de-identified administrative student data on PASS participation, background characteristics, grades and retention. The intuition of the methodology is to compare the outcomes of all students who were offered the incentive to all who were not (and scaling up the difference by the proportion of such students who were induced into PASS by the incentive).

POSSIBLE RISKS, INCONVENIENCES AND DISCOMFORTS

We can foresee no risks for you. Your involvement in the study is voluntary and you may withdraw your participation from the study at any time. The decision not to participate, or to withdraw from the study, will not affect any current or future relationship with the University of Wollongong. To opt out or withdraw from the study, or for further information, e-mail: chscr-1@uow.edu.au

FUNDING AND BENEFITS OF THE RESEARCH

This study has been funded by an Educational Inclusion Grant, through the UOW Faculty of Business. This research will contribute to the evidence base on the effectiveness of the PASS program for improving student outcomes, as well as the effectiveness of incentives to encourage student participation in PASS sessions. Findings from the study will be published in a report to the UOW and likely be published in economics or educational journals. Confidentiality is assured and you will not be identified in any part of the research.

ETHICS REVIEW AND COMPLAINTS

This study has been reviewed by the Social Sciences Human Research Ethics Committee of the University of Wollongong (HE13/360). If you have any concerns or complaints regarding the way this research has been conducted you can contact the UOW Ethics Officer on (02) 4221 3386 or email rso-ethics@uow.edu.au

Thank you for your interest in this study.

BIG Prizes for attending PASS

FIFTY $1,000 Gift Cards

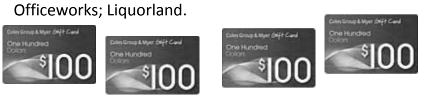
Across 9 subjects this year

Half of You are Eligible

Eligibility notified by SMS and/or E-mail today or tomorrow

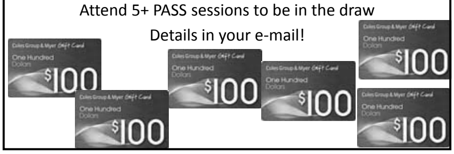
Coles Group & Myer Card

Valid at Coles; Coles Express; Target; Myer; Kmart; Officeworks; Liquorland.



Attend 5+ PASS sessions to be in the draw

Details in your e-mail!



Why these extra prizes?

For Research Purposes

See http://www.uow.edu.au/~siminski/PASS_research.pdf

Any Questions?