SOCIAL SUPPORT AND ACADEMIC SUCCESS: FIELD EXPERIMENTS IN FURTHER EDUCATION IN ENGLAND

By:

Anthona Sophia Bibiana Groot

SCHOOL OF PUBLIC POLICY
UNIVERSITY COLLEGE LONDON (UCL)
SEPTEMBER 2018

A dissertation submitted to University College London in accordance with the requirements for award of the degree of Doctor of Philosophy.

Word Count: 78,003

Author's Declaration

I, Anthona Groot, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

The two field experiments set out in this thesis were conducted as part of the Behavioural Research Centre for Adult Skills and Knowledge (ASK) in collaboration with Professor Todd Rogers (Harvard Kennedy School), and were funded by the UK Government's Department for Innovation, Business and Skills (BIS). Todd Rogers and his team at the Student Social Support Lab were involved in developing the intervention introduced in Chapter 4. I led on trial design and implementation, collected all outcome data, conducted the analyses presented throughout this thesis, and wrote this thesis in its entirety.

SIGNED 04/09/2018

Abstract

Universally, humans have a strong need to feel valued and cared for by their close social relationships (Baumeister & Leary, 1995). The degree to which people can identify such sources of support is strongly correlated with positive emotional and physical health outcomes (Thoits, 2011; Taylor, 2011), as well as academic achievement (Song, Bong, Lee, & Kim, 2015; Wentzel, Russell & Baker, 2016). Yet, few researchers have robustly tested how supportive communication between students and their social networks can be stimulated when it is lacking. This thesis makes a contribution to the academic literature and education policy by developing and testing interventions that motivate, inform, and remind students and their immediate social networks about their learning.

The thesis introduces a fresh approach to the design of social support interventions. Rather than introducing new ties or establishing formal mentoring relationships, students' existing relationships are enlisted to provide support. Students' friends and family are, after individual randomization, sent a series of weekly text messages over the full academic year. These messages contain actionable and relevant information about the student's course, inspired by recent information interventions in education (Kraft & Rogers, 2015; Chande, 2017).

The results indicate that informing study supporters of students' learning improves student attendance and attainment in maths and English qualifications. The follow-up trial shows that communicating with study supporters and students simultaneously is more effective than communicating with study supporters only. Qualitative evidence provides new insights for the design and implementation of supportive information interventions. Additionally, this thesis provides novel qualitative evidence that support from parents and friends helps students overcome challenges through cognitive appraisal processes. This data therefore offers new support for the popular hypothesis in the published literature (e.g. Feeney & Collins, 2015) that improved coping with emotions is a primary mechanism of social support on psychological well-being and academic success.

Impact statement

This thesis may be of interest to academics, policy-makers and practitioners. The findings show that a light-touch intervention can make a real difference to student success in post-16 institutions. It can be readily implemented in schools and colleges: the programme of supportive text messages requires little to no new infrastructure or software. In fact, most schools and colleges have management information systems with built-in text messaging software. At only 4 pence per text message, a small nudge can have big effects.

The replication and extension of these findings have already commenced. First, I gained funding from the Education Endowment Foundation (EEF) to replicate and extend the studies described in this thesis. The project is currently running across 31 further education colleges, and the results will be made public in Spring 2019. Second, the recognition that few education institutions were using their existing communication channels to proactively support students has motivated the decision to build a custom text-messaging platform: Promptable.[1] I work closely with the Ventures team at the Behavioural Insights Team to translate the academic research into a scalable social enterprise. The scheduling of the weekly texts has been automated, so that the cost of implementation can be reduced even further. Promptable will be used by 100 Sixth Form colleges from September 2018 and may be used by organisations locally, nationally and internationally.

The idea that students' close social relationships can be leveraged to boost educational attainment has gained public attention. The dissemination of outputs has occurred both via mainstream media as well as specialist publications. The Sunday Times (Bennett, 2017)[2] reported on the results of the first field experiment introduced in Chapter 4, as well as Forbes Magazine (Morrison, 2017),[3] BBC Radio 4 (2016),[4] and Times Educational Supplement

[1] The website of Promptable, a BI Ventures product, can be accessed at https://promptable.com/
[2] Available at https://www.thetimes.co.uk/article/good-luck-texts-boost-exam-grades-pvrjmn7cj
[3] Available at https://www.forbes.com/sites/nickmorrison/2017/10/18/the-little-nudge-that-makes-a-big-difference-to-student-grades/
[4] The audio clip is available on http://www.bbc.co.uk/programmes/p047xrbj

(Exley & Martin, 2017).[5] The Financial Times (Green, 2017)[6] reported on the replication study funded by the EEF. Finally, both intervention studies introduced in this thesis are published in collaboration with the Department for Education (Hume et al., 2018a, 2018b). The practitioner report actively seeks to disseminate these research findings to non-academic audiences (Hume et al., 2018a, Section 3, p. 20). Finally, policy makers can read about the Study Supporter intervention in The Green Book (HM Treasury, 2018, p. 17),[7] which provides guidance the appraisal and evaluation of spending proposals. The Study Supporter trial was included as a case study to illustrate the value of testing and iteration.

Finally, the iterative approach to intervention development through the mixing of quantitative and qualitative evidence will hopefully benefit the academic disciplines of public policy, social sciences and educational research. On a final note, participating college staff have anecdotally shared stories of positive impacts of the supportive communication interventions: “Our learners have worked really hard this year and I believe that you have made a huge difference to learner attitude” (L. Merceron, personal communication, June 11, 2018).

[5] Available at https://www.tes.com/news/grit-classes-and-supportive-texts-boost-student-attendance-research-reveals

[6] Available at https://www.ft.com/content/3df1d370-a9c7-11e7-ab66-21cc87a2edde

[7] The publication can be accessed at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf

Acknowledgements

I am sincerely grateful to my academic supervisors, Professor Peter John and Jake Anders, for taking on my supervision halfway through the journey. They have shown such generosity and kindness when I was figuring out how to put everything together. I would like to thank them and the staff at UCL School of Public Policy and the Institute of Education for their support in ensuring a smooth transition between departments. David Mallows, thank you for guiding me during the first year of my doctoral studies. I am indebted also to Michael Sanders, my BIT supervisor, who showed me how field experiments are done in the real world, and who introduced me to the wonderful world of cleaning education datasets.

My colleagues at the Behavioural Insights Team, who have made the past few years such a rich experience: Raj Chande, Zhi Soon, Fionnuala O'Reilly, Jess Barnes, Sara Halkiopoulos, Jess Hunt, Susannah Hume, Michael Sanders, Miranda Jackman, and many others. Eliza Kozman, Sara Halkiopoulos and David Nolan for their expert project support. Eliza, thank you also for lending your voice for the video that motivated countless students to sign up. I am also grateful to my fellow BIT PhD students for their friendship and support.

Todd Rogers, who was kind enough to lend me his idea of involving students' social networks in their learning, and who provided significant support and encouragement at critical points in this journey. Thanks also to Emily Bloomenthal at the Student Social Support Lab at Harvard University for the fantastic guidance on writing supportive text messages.

I am particularly grateful to the staff at the further education colleges who helped implement my field experiments. In particular, I would like to thank Michael Smith, Darrell De'Souza, Barbara Worms, Teresa Stroud, Fiona Morey, and Mumtaz Van Der Vord. I am also indebted to the Department for Innovation, Business and Skills and the Department for Education for their generous funding of my PhD scholarship. The additional funding from the Education Endowment Foundation is gratefully acknowledged.

Finally, I've had real Study Supporters supporting me throughout this journey. They checked in with me when I needed it most, showing just how much social support matters - Astrid, Cristina, Francisco, Joan, Jorine, Jose Maria, Marieke and Veerle. Above all, my family - Rob, Ellen, Olivier and Oma Tonny - and the love of my life, João Nuno, this thesis is dedicated to you. Thank you for your unwavering love and support.

Table of contents

List of tables

List of figures

1 INTRODUCTION

1.1 Prologue

Education has a major impact on people's life chances. Attaining adequate levels of literacy and numeracy is associated with a reduced risk of economic disadvantage and poor emotional or physical health (Hanushek, Schwerdt, Wiederhold, & Woessmann, 2015; McIntosh & Vignoles, 2001; Organisation for Economic Co-operation and Development [OECD], 2013). Individuals with low proficiency in literacy are more than twice as likely than their higher skilled peers to be unemployed (OECD, 2013). The wage premium of basic skills is considerable; a one standard deviation in numeracy skills is associated with an 22.5 percent wage increase for UK adults, after controlling for their education levels (Hanushek et al., 2015). The implications of education also move beyond economic and health outcomes. Voter participation and support for free speech are both positively and strongly associated with educational attainment (Dee, 2004).

The latest OECD review of skills levels in England dedicated considerable attention to the low skills levels among those aged 16 – 19 (2016). One-third of those aged 16 – 19 living in England have low basic skills, a proportion three times larger than the three best performing OECD member countries: Finland, Japan and the Netherlands. Additionally, England sees more limited and slower literacy and numeracy progress in later adolescent (20 – 22) years than other OECD countries. These findings provide an interesting contrast to England's relatively high levels of tertiary attainment. In 2016, 48.1 % of adults in the UK are tertiary graduates (European Commission, 2017), yet only 24.1 % and 27.5 % of adults with tertiary education achieved Level 2 or above on literacy and numeracy, respectively (Wheater et al., 2013). Level 2 is equivalent to the level expected from sixteen-year-olds, suggesting that having achieved tertiary education levels is not equivalent to having good literacy and numeracy skills.

To address the issue of poor basic skills (i.e. literacy and numeracy) in England, the government has raised the participation age to 18 (Department for Education, 2012) so that students between 16 and 18 are now required to be in education or training. Additionally, a new policy was introduced in 2015, which requires 16-18 year-olds to continue to work towards maths and/or English qualifications if they failed to obtain an A*-C on the General Certificate of Secondary Education (GCSE) for these subjects at age 16 (Education Funding Agency, 2014). Maths and English study up to the age of 18 has become a condition of government funding, binding students and institutions to the reforms. In 2016/17, the latest available data, 344,621 students (41.5 %) left secondary school without an A*-C (or 9-4) for English and maths, respectively (Department for Education, 2017b, p. 12). Half of students who fail maths and/or English at age 16 move to further education (FE) colleges to resit the qualification alongside taking vocational qualifications (Department for Education, 2016b, p. 10, Figure 5a). In 2016/17, only one out of every four post-16 resit students achieved a passing grade for their maths or English GCSE (23.4% for maths, 23.8% for English; Department for Education, 2018).

The catch-up success rates are considerably higher for schools and sixth form colleges which can partly be explained by the observation that low-attaining students (e.g. below a D grade on GCSEs) are disproportionately likely to move into further education (Impetus- The Private Equity Foundation [PEF], 2017, p. 15, figure 5a and 5b). Only 5 % of catch-up students continue to study at sixth form colleges, and they are more likely to have obtained an almost-pass in English (i.e. D grade, 67.9 %) than their peers at FE colleges (51.9 %) a difference of 16 % points (Department for Education, 2016b, p. 11, Table 5a).[8] This better starting point for sixth form students translates into higher success rates: 32% and 18% of catch-up students achieve a A*-C in English and maths, respectively (Impetus-PEF, 2017, p.16, Figure 6). Catch-up rates are considerably poorer in further education colleges, where 13% and 5% of students who did not achieve A*-C by the end of secondary school achieve a pass for English and maths GCSEs, respectively (Impetus-PEF, 2017, p.16. Figure 6). Further education (FE) colleges are thus in particular need of support of effective strategies to help improve basic literacy and maths skills. Further, a number of structural issues, including significant budget cuts (McNally & Wyness, 2017; Wolf, 2015) and high teacher turnover rates (11.2%; Frontier Economics, 2017), impede the sector's ability to

[8] This observation holds for maths GCSE as well, where 52.5% of sixth form students previously achieved a D grade, contrasted with only 35% of catch-up students at further education colleges having previously achieved an almost-pass (Department for Education, 2016b, p. 11, Table 5b).

provide personalised support to its students. A key policy question is therefore what low-cost, scalable and effective strategies can help FE colleges deliver adequate support to further improve student engagement, retention, and success.

This thesis makes a contribution to this objective by developing and testing interventions that motivate, inform, and remind students and their immediate social networks about their learning. Motivated by the policy issues set out above, this thesis employs the literature on social support to explore ways in which college communication can be enhanced. As such, this thesis sets out an intervention aimed at addressing a specific policy issue, but it also aims to stand alone as a theory piece. Finally, this thesis is inspired by the behavioural science literature that has produced interventions with remarkably positive results, while keeping costs low and implementation straightforward. I focus on 16-18 year-olds as adolescence represents a transitionary period where the foundations for positive adjustment throughout the life course are laid.

1.2 A journey through the literature

In the hope of finding low-cost and easy-to-implement ways to help further education college students achieve their basic skills qualifications, I first turned to the literature on “wise interventions”. This class of interventions promises to increase student engagement by fostering more positive mindsets about the value of learning and schoolwork (Hulleman & Harackiewicz, 2009), their belonging at educational institutions (Walton & Cohen, 2011), and their academic potential (Yeager & Dweck, 2012). These interventions are easy and cheap to deliver since they require only 20 minutes of class time, and find surprisingly large treatment effects on student achievement (Harackiewicz & Priniski, 2018; Hulleman & Barron, 2015). For example, in an attempt to foster perseverance in maths, Bettinger and colleagues (2017) teach high-school students about the malleability of the brain. Students in the control group read an article about the brain, but do not learn about its potential to grow and change. Three weeks later, students in the treatment group who initially believed their abilities were fixed exerted significantly more effort on a maths task than their control group peers, by approximately 30 per cent of a standard deviation (Bettinger, Ludvigsen, Rege, Solli, & Yeager, 2017). Yet, these types of social-psychological interventions are limited by the fact that they rely on students’ ability and willingness to engage in reflective exercises. Especially those who are mandated into learning may struggle to engage meaningfully with such interventions. Many GCSE resit students have negative views of maths and English, do not want to engage with the subjects (Williams, Hadjivassiliou, Marvell, Green, & Newton, 2017), and have little confidence that they will be able to break the 'cycle of failure' (Wallace, 2013). Restoring students' sense of self-efficacy may require more than a writing exercise.

The academic literature on mentoring sparked an interesting new avenue: would it be possible to build a more supportive learning environment through personal attention and support to students at risk of disengagement? Students could be coupled with a mentor who checks in with them regularly, helps out when issues arise, and who builds trust over time. Although the evidence-base for mentoring programmes is mixed at best (DuBois, Portillo, Rhodes, Silverthorn, & Valentine, 2011), some mentoring interventions have shown promise: youth from poor socio-economic background were considerably less likely to drop out from school and enrol in college when they were paired with a proactive mentor (Oreopoulos, Brown, & Lavecchia, 2017; Van der Steeg, van Elk, & Webbink, 2015). Unfortunately, these can be prohibitively expensive for colleges to sustainably implement, especially considering the significant funding cuts to further education over the last decennium (Hupkau, Mcnally, Ruiz-Valenzuela, & Ventura, 2016).

Rather than introducing new ties and establishing formal mentoring relationships, students' existing relationships could be enlisted to provide support. A large albeit mostly correlational literature documents that the resources and interactions provided by people we trust and care about can help people cope with stress (Thoits, 2011). Numerous studies have documented the beneficial effects of social support on emotional and physical health (Thoits, 2011), and educational outcomes (Song, Bong, Lee, & Kim, 2015; Wentzel, Russell & Baker, 2016). Although there is no consensus in the literature as of yet, social support appears to benefit individuals through greater self-esteem, reduced negative affect, and more positive self-conceptions (Burleson & MacGeorge, 2002). Recent studies also find that individuals who regularly interact with supportive others have diminished cortisol reactivity to stressors (Eisenberger, Taylor, Gable, Hilmert, & Lieberman, 2009) suggesting that neurocognitive mechanisms underpin the beneficial effects of social support. It is thus well established that social support is important. But how can we activate social support processes when these are lacking? Not all young people have access to an adult who cares about them and their learning, be it a parent, teacher, neighbour, social worker or relative. Further, even if family and friends want to get involved, they may not feel they have the tools, knowledge, or access to information to do so effectively (Hoover-Dempsey et al., 2005). Thus, this thesis introduces an intervention that aims to inform, remind, and motivate family and friends to get more involved.

A final set of interventions provided the necessary inspiration for the design and delivery of the intervention. Recent studies have found that simply providing relevant and timely information to the student or their parents can improve educational outcomes (Bergman & Chan, 2017; Kraft & Rogers, 2015; Kraft & Monti-Nussbaum, 2017; Rogers et al., 2017). These studies show us that relatively simple text messages have the potential to empower both parents and students to get more involved. The texts are designed to address specific behavioural barriers to engagement. First, parents may erroneously believe their child is doing well in class and attend all lessons because their child holds private information about their effort (Bergman, 2015; Bergman & Chan, 2017). Second, parents may not know how and when to get involved with their child's education or may not believe their effort will make a difference (Hoover-Dempsey et al., 2005). Finally, students may simply forget deadlines or may put off important tasks.

Text message reminders can refocus people's attention on the task at hand, and are commonly used in behaviour change interventions (Richburg-Hayes et al., 2014). Text-message reminders of deadlines and tasks have been found to be effective at improving educational outcomes for high school and university students (Castleman & Meyer, 2016; Castleman & Page, 2015). These information interventions have thus far targeted students' parents or guardians. Recognising that students at post-16 institutions may no longer live at home (or indeed, be parents themselves), this thesis took a different approach where students could choose their own 'study supporter' who would then receive informative text messages about the students' learning every week. Finally, to assess whether it is indeed the social aspect that motivates students, the second research question assesses whether direct college-student communication is more or less effective than communication via study supporters.

1.3 Gaps identified in the literature

Social support, or the "individual's perception of general support or specific supportive behaviours (available or acted on) from people in their social network" (Malecki & Demaray, 2003, p. 232), is positively correlated with a range of beneficial outcomes. Adolescents who report they have access to supportive others are more likely to be engaged academically (Pan, Zaff, & Donlan, 2017), have higher expectations of their educational success (Goodenow, & Grady, 1993) and feel lower test anxiety (Song et al., 2015) than their peers who report low support. The academic literature on social support has thus far primarily focused on the description of support processes. Most studies on social support, especially within the domain of educational research, rely on cross-sectional samples. In these studies, students self-report the support they perceive or receive from parents, peers and teachers. It is relevant to know that adolescents primarily rely on their parents for informational support and turn to their peers for emotional support (Malecki & Demaray, 2003), but this correlational data tells us little about ways to build supportive relationships.

Similarly, there is a considerable body of research on the protective benefits of social support on physical and emotional well-being (Greene & Burleson, 2003; Uchino et al., 1996), but few studies have sought to explore how to mobilise greater involvement from people's family and friends. The majority of social support interventions focus on introducing new ties, such as 'peer supporters' who previously suffered from the same condition as the recipient of the intervention (e.g. Ussher, Kirsten, Butow, & Sandoval, 2006). It is less common for interventions to focus on the existing social networks of the individual. Exceptions exist: a family support intervention where obese adolescents' parents learnt about creating a healthier and more autonomy-supportive home environment found that treated youth consumed more fruit and less junk food even a year post-intervention (Straker et al., 2014).

Can supportive communication interventions also result in greater school engagement, in a population of students who have negative prior experiences with education? Several recent interventions to improve parent-school communication have found promising results (e.g. Bergman & Chan, 2017), so there is good reason to think information interventions can boost attainment. Yet, this thesis is one of the first to test the effectiveness of such approaches in a further education college context. The literature review in

Chapter 2 also informed a distinct approach to encouraging friends and family to become more involved, namely, by providing them with positive, forward-looking information about the student's learning. The current body of information interventions delivered through text messages is primarily focused on informing parents of their child's misbehaviour: poor attendance, missed assignments, or low grades (see for example Bergman & Chan, 2017; Rogers & Feller, 2018). This thesis therefore tests whether a softer and more autonomy-preserving approach can prompt students' close relationships to support their learning.

Randomised controlled trials (RCTs) can isolate the causal effect of an intervention (Gerber & Green, 2012), as will be reviewed in depth in Chapter 3. This thesis employs RCTs to answer the primary research questions set out in section 1.5 (p. 19). Although RCTs provide robust estimates of intervention effectiveness, they too often use "a binary understanding of the social reality – an intervention either worked or it didn't" (Hesse-Biber, 2013, p. 50). Often, the social context in which RCTs are implemented is de-prioritised in favour of control over the experimental manipulation (Bonell, Fletcher, Morton, Lorenc & Moore, 2012). This thesis attempts to fill this gap by embedding qualitative inquiry into the randomised controlled trial to facilitate exploration of intervention theory of change. Does greater involvement of a close relationship protect students from college-related stressors through social support, or is it simply the reminder to study for an upcoming exam that propels students into action? This thesis thus employs a mixed methods randomised controlled trial design to uncover causal effects and gain a deeper understanding of the interactions between students and supporters.

1.4 Contributions

This thesis aims to make an empirical contribution to the literature on information interventions by combining the theoretical foundations of need-supportive communication with the low-cost and light-touch design of a text-messaging intervention. The aim of the supportive communication intervention is to provide relevant and motivating information to the student and their social network, and to facilitate friends and family to have more and better conversations with students about their learning. Few intervention studies have focused on strengthening supportive communication between students and their social networks, and none have done so with a cohort of vocational students taking maths and English qualifications. Alongside this empirical contribution, this thesis aims to integrate different areas of research to enhance our understanding of what constitutes supportive communication and whether and how it can be leveraged through interventions. The literature review brings together the adjacent literatures on social support and supportive communication, and then combines it with the growing literature on information interventions.

A second intended contribution of this thesis focuses on the theory of change of the intervention. Two pathways are offered. First, regular informative prompts about the student's learning may empower the immediate social network to become more involved. Having access to a social network and being able to share one's thoughts and worries may also enhance individuals' ability to deal with everyday stressors better. Thus, the intervention's 'active ingredients' may be the stimulation of close social relationships to become more involved in the student's learning. Alternatively, the prompts to student's key relationships may be effective because the student sees the study supporter as an arrangement to help fulfil the academic goals that they may otherwise struggle to see through. The behavioural science literature on commitment devices posits that social pressure to follow through on a commitment can be a highly effective behaviour change strategy (Rogers & Frey, 2015). This thesis utilises qualitative data gathered across the two field experiments to explore students' motives for signing up. The qualitative data brings together different perspectives on the programme and explores whether students primarily hope to receive more support from a third party, the study supporter, or constrain their future selves through the nomination of a social commitment device.

A final set of contributions is expected from the combination of qualitative and quantitative data collection and analysis. The in-depth qualitative interviews with students and tutors will contribute to a better understanding of the behavioural and structural barriers to student engagement at further education. The qualitative component also seeks to enhance the contextualised design of information interventions. The quantitative element, on the other hand, sheds new light on effective and low-cost strategies to enhance student engagement and achievement.

1.5 Research questions and thesis outline

This thesis builds on the literature showing that informative text messages sent to parents or guardians can improve educational outcomes (e.g. Bergman & Chan, 2017; Castleman & Page, 2015; Kraft & Rogers, 2015) and asks whether similar treatment effects can be observed in an intervention where students have full autonomy over the choice of recipient of these communications. It is organised around the following two research questions:

1. **Primary:** Can supportive text messages delivered to students' social networks improve attendance and achievement in maths and English courses at further education colleges?
2. **Secondary:** Are the effects greater if students also receive the text messages?

The first research question is answered by both field experiments set out in this thesis (Chapter 4 and 5, respectively). The first field experiment compares communicating with the nominated study supporter against a business-as-usual control group. The second field experiment addresses research question 2 by isolating the added benefit of providing the same information to students directly. I will now briefly outline the rest of this thesis.

Chapter 2 presents a comprehensive review of the literatures on social support and information interventions and explores how they may be combined into a theoretically-informed intervention. I argue that perceiving support from close others is a strong facilitator of wellbeing and motivation to learn, and then review practical approaches to leveraging existing relationships to become more engaged in students' learning.

Chapter 3 sets out the methodological approach taken in this thesis, namely that of a mixed methods randomised controlled trial (RCT). The chapter reviews the potential challenges to internal and external validity of RCTs and argues that qualitative methodologies can serve to enhance the credibility of field experiments in complex social settings.

Chapter 4 and 5 each present a field experiment. These studies have been published elsewhere in a government report on supporting retention and attainment in maths and English (Hume et al., 2018a; 2018b) and an early working paper (Groot, Sanders, & Rogers, 2017). Chapter 4 presents a two-arm trial, which tests the effectiveness of a supportive communication intervention aimed at students' close relationships. Chapter 5 builds on this foundation by introducing two additional trial arms to isolate the effects of direct information provision to students versus delivery via a third party. Both field experiments are complemented by in-depth qualitative interviews with students, study supporters and tutors to better understand the facilitators and barriers of the interventions.

Chapter 6 explores the potential theory of change of the supportive communication interventions through the integration of qualitative data gathered throughout the two experimental phases. Students' motives at sign-up are interrogated, as well as their subjective experiences of the intervention. The chapter also utilises qualitative data to further explore taxonomies of social support. Finally, Chapter 7 offers concluding remarks.

2 REVIEW OF THE LITERATURE

2.1 Introduction

This chapter appraises the available evidence base on the potential of supportive communication interventions to improve academic engagement and attainment. It lays the foundations for the development of a supportive communication intervention questions through a critical review of the literature. This chapter has three main objectives: (1) to locate the thesis in the academic literature on social support and social support interventions; (2) to review the similarities and differences between parent-teacher communication interventions and the interventions introduced in this thesis; and (3) to summarise the potential pathways through which a supportive communication intervention produces beneficial academic outcomes. In each section, the aim is to identify what is known, assess how robust the underlying research is, and draw implications for the research design set out in Chapter 3.

In this thesis, I attempt to form a picture of how students' social context can be leveraged to enhance learning and motivation, focusing specifically on close ties within students' existing social networks. The primary research question throughout this thesis is whether supportive college communications with friends and family can improve attendance and achievement in maths and English courses at further education colleges. The secondary research question asks whether direct college-student communications are more, equally, or less effective than communications delivered via close relationships. Essentially, this question asks if social support is the active ingredient of the intervention. The majority of this chapter therefore focuses on the role social support and supportive communication play in promoting well-being and engagement with learning.

Section 2 defines social support and examines the available literature on the importance of having access to supportive others. It explores the psychological and physiological benefits of social support. Greater school engagement and achievement is the focus of this section, although it should be noted that the literature on the relationship between social support and health offers relevant perspectives as well. This section focuses particularly on different sources of support: parents, teachers and peers.

Section 3 appraises the available evidence base on mechanisms through which social support benefits individuals. It is posited that having access to positive social and community ties reduces individuals' stress responses through improved coping (Cohen & Wills, 1985; Uchino, Cacioppo, & Kiecolt-Glaser, 1996). It proposes that social support interventions would benefit from a clear focus on pathways of social support. It is also noted that the picture is more complex than often acknowledged: social support is beneficial to recipients when it meets their specific needs but may fail to benefit them when it is not provided in a skilful manner.

Section 4 reviews design elements of social support interventions. Design features of social support interventions are reviewed to support intervention development work. It proposes that intervening on individuals' existing social ties within a dyadic setting is a promising avenue to improving educational outcomes for students in post-16 institutions.

Section 5 then reviews the growing evidence-base on information interventions. Similarities and differences between existing interventions are reviewed. Most information interventions focus on remedying low parental monitoring of their child's behaviour through personalised information. A notable minority of studies also seek to encourage positive parent-child conversations about learning. This section also reflects on potential barriers individuals may face enlisting the support they need, and barriers close relationships may experience to being actively involved in students' learning.

Section 6 concludes by highlighting how the thesis may contribute to the literature and address existing gaps. There is little research on the relationship between social support and educational outcomes, especially at post-16 institutions. Furthermore, few intervention studies have actively sought to encourage frequent and positive communication between colleges, students, and their family and friends.

2.2 The relationship between social support and student adjustment

Relationships between students and their peers, parents, teachers and wider social networks are key factors determining whether students complete their courses and stay motivated throughout the year (Goodenow & Grady, 1993; Rosenfeld, Richman, & Bowen, 2000; Soenens & Vansteenkiste, 2005; Wentzel, 1998). This literature may help us understand under what conditions communication interventions optimally leverage close social relationships.

I first discuss the definitions and measurement of social support. The following section then introduces the role of proximal social contexts and salient relationships, such as peers, teachers and parents, and the ways in which they may nurture or thwart student psychological adjustment and educational outcomes.

2.2.1 What is social support?

The concept of social support has been studied across various disciplines, which has resulted in numerous partially overlapping definitions (Williams, Barclay, & Schmied, 2004). The definition used in this thesis is "the individual's perception of general support or specific supportive behaviours (available or acted on) from people in their social network" (Malecki & Demaray, 2003, p. 232). This definition was chosen due to its inclusion of both types of social support: perceived and received social support. Perceived support focuses on the anticipated support available from others, such as parents, classmates or close friends (Lakey & Cohen, 2000), and has been a focal construct in the literature on academic achievement and social support. The evidence for the link between perceived social support and academic achievement is relatively robust and consistent (Ahmed, Minnaert, van der Werf, & Kuyper, 2010). For example, Rosenfeld et al. (2000) asked a large representative sample of middle and high school students to report the social support they perceived to be available from their peers, parents and teachers, and found that those who perceived support to be available achieved better grades. Similarly, Wentzel (1998) found that adolescents' perception of support correlated strongly with various indicators of school motivation, including interest in school and their grade point average (GPA).

Perceived social support is by definition subjective. Some scholars have therefore opted to chart the actual enactment of social support or the number of social ties to assess the relationship between social support, well-being and academic outcomes. Received support is typically measured by asking respondents to list the frequency of contact with others in their social network, or count number of close friends and relatives. Perceived and received support are not as closely correlated as one might expect, at around r = 0.30 (Kaul & Lakey, 2003). One explanation might be that perceived support reflects generic positive evaluation of relationships (Kaul & Lakey, 2003) whereas received report is a measure of the size of the social network. Paradoxically, received support is sometimes found to be associated with negative outcomes (Kaul & Lakey, 2003). It has been posited that receiving social support signals to the recipient that others perceive the problem too great for the recipient to cope with (Shrout, Herman, & Bolger, 2006), creating feelings of weakness, guilt or indebtedness.

Beyond perceived and received social support, researchers are often interested in the types of social support provided. Taxonomies of social support may help identify what types of support should be encouraged by social support interventions. Chapter 6 focuses specifically on student-supporter interactions and the types of social support enacted (Section 6.4, p. 221). House's (1981) influential and pioneering framework outlines four broad categories of social support: emotional, instrumental, informational and appraisal support. Emotional support involves the provision of care, encouragement, acceptance and nurturance (Cutrona & Russell, 1990). Instrumental support, on the other hand, involves active helping or the provision of material resources, such as help with transportation or college fees. Informational support involves the communication of knowledge, guidance and feedback, and occurs when "one individual helps another to understand a stressful event better and to ascertain what resources and coping strategies may be needed to deal with it" (Taylor, 2011, p. 192). Lastly, appraisal support is defined as receiving positive appreciation for one's efforts (Gottlieb & Bergen, 2010). Recent studies commonly exclude the final category, appraisal support, since it is difficult to distinguish from emotional support in practice. In line with Taylor (2011), this thesis focuses on informational, instrumental and emotional classifications of social support.

Finally, this thesis draws on the literature on supportive communication, which focuses on the interactive processes of enacting social support. Social support is "fundamentally communicative in character" (Burleson & MacGeorge, 2002, p. 384); it necessarily involves a recipient and provider, and some form of non-verbal or verbal communication. The outcome of such interactions can be either positive or negative as a function of characteristics of the recipients, supporter, and the interactional context (Bodie & Burleson, 2008). The next section examines how characteristics of support providers influence student adjustment and student success.

2.2.2 Sources of social support: parents versus peers

The relationship between social support and academic engagement is complex and few studies have been able to identify causal directions due to study designs. The majority of studies on social support and engagement rely on cross-sectional designs and self-reported data. The study of naturally occurring social support lends itself well to observational research. The studies reviewed here therefore provide descriptive evidence of the importance of a supportive environment. Such research is relevant to this thesis because it highlights the importance of interaction with close others. Since this thesis proposes an intervention which intervenes on students' existing social ties, it is important to understand how different sources of support are associated with educational outcomes. This section first addresses parental or family support, and then reviews support from peers.

Adolescence is often seen as a transition period where "the network of significant others is restructured" (Helsen, Vollebergh & Meeus, 2000, p. 320) as peer support becomes crucial while parental support diminishes in importance. After all, for most young people, peers are their primary source of social interaction (Lynch, Lerner & Leventhal, 2013), so the peer group may be a potentially powerful setting for encouraging positive learning-related behaviours. Students also perceive parents and teachers as less supportive as they move from primary- to secondary school and into college education, while perception of peer support peaks (Furman & Buhrmester, 1992).

The current evidence base on the relationship between social support and wellbeing contests the overriding importance of peer support in adolescence. A number of studies find that parental social support is closely related to adolescent well-being, and more so than peer support (Helsen et al., 2000; Malecki & Demaray, 2003, 2006; Stice, Ragan, & Randall, 2004; Wentzel, 1998). Perceived parental support is associated with lower incidence of internalising problems in adolescence, such as anxious and depressive symptoms (Stice et al., 2004; Piko, Luszczynska, & Fitzpatrick, 2013). Autonomy-supportive parenting, which is characterised by consistent and sensitive parenting behaviour, is associated with higher autonomous motivation for school, greater perceived well-being and competence, and increased engagement and effort (Soenens & Vansteenkiste, 2005; Vasquez et al., 2016). Finally, supportive parenting practices such as providing emotional warmth, clear expectations, and autonomy support are predictive of high academic engagement (Bempechat & Shernoff, 2012).

Peer support effects are typically discussed in relation to mental wellbeing, such as depression or dealing with an illness (e.g. Stice et al., 2004). Few studies directly examine the associations between peer support and academic outcomes. This appears to be an issue of jargon, since there is a large literature on peer connectedness, acceptance and relatedness (for example, see Ruzek et al., 2016; Wentzel & Caldwell, 1997). This more generalised sense of identification with peers and students is positively related to academic motivation, but only when one's friends value academics as well (Nelson & DeBacker, 2008). Similarly, feeling supported by peers is positively associated with interest in class (Wentzel, Battle, Russell, & Looney, 2010) and prosocial goal pursuit (Wentzel et al., 2016), again, only when these peers were not resistant to school themselves. Intuitively, it is not surprising that students who feel connected to their classmates and enjoy mutual relationships also tend to be motivated to do well in school. Additionally, the effects of peer support on student adjustment appear to depend at least in part on the level of perceived parental support (Helsen et al., 2000).

Longitudinal studies shed light on how both perceived support and academic engagement increase and decline over time and interact with one another. Wang and Eccles (2012) examined the effects of peer, parent and teacher social support on adolescent school engagement. In a longitudinal study with twelve to seventeen year- olds, school engagement was measured not only by student grades, but also by compliance with school rules, absence of disruptive behaviours, participation in extracurricular activities and social identification with the school (Wang & Eccles, 2012). All self-reports of school engagement declined over the 6-year period of the study. Notwithstanding the general decline in self-reported engagement as students aged, different sources of social support had different effects on the adolescents' school engagement. Supportive teachers were particularly beneficial to the students' valuing of learning and identification with school. Whether peer social support benefited the student was largely dependent on the type of peers they interacted with (i.e. peers exhibiting prosocial versus antisocial behaviour). Lastly, parental social support was positively associated with all measures of school engagement, and a stronger predictor of school engagement than peer support (Wang & Eccles, 2012).

In sum, the available evidence base suggests that each source of social support can be beneficial to young people's well-being and academic engagement, provided the parent or peer has a positive attitude towards education themselves. The next section delves deeper into the potential pathways via which social support protects the individual from stressors and helps them to feel motivated in school.

2.3 Mechanisms, pathways and contextual factors

The beneficial and protective effects of social support on various life domains including physical and mental well-being have been extensively documented. For example, university students who reported that they had access to people to interact and socialise with reported fewer physical health symptoms and rated their general health more positively than those who perceived a lack of social intimacy (Hale, Hannum, & Espelage, 2005). In a meta-analysis of 81 studies on the effects of social support on health, Uchino et al. (1996) conclude that support is reliably related to better immune responses to acute stress, lower rates of morbidity and mortality, lower coronary heart disease, and lower blood pressure, after controlling for health-related factors.

2.3.1 Theoretical models of social support pathways

The precise mechanism through which social support influences well-being is still debated. The discussion has thus far focused on two theoretical models, that each attempt to explain how and when social support produces positive outcomes. Neither of the below theoretical models appear to fully explain the relationship between social support and positive adjustment, and both types of support effects have been found in the empirical literature (Feeney & Collins, 2015). Yet, these theoretical orientations are of interest because the choice between these models ultimately influences the design of a social support intervention.

2.3.1.1 The stress buffering theory of social support

The stress-buffering perspective posits that social support buffers the individual from the negative effects of life stressors (Cohen & Wills, 1985), and that social support is thus beneficial only in the face of stressful life events. Proponents of this perspective argue that social support bolsters the individual's perceived ability to cope with stressors (Cobb, 1976). There is empirical evidence that social support moderates stressful life events. For example, a recent longitudinal study with first- and second-generation immigrant adolescents in the United States found that internalising mental health symptoms increased over time in young people who experienced acculturative stress (Sirin et al., 2013). However, greater perceived social support altered this relationship between stress and poorer mental health, (Sirin et al., 2013). The relationship between depressive symptoms and negative stressors was weaker for the adolescents who reported high social support in comparison to those who perceived less support to be available.

2.3.1.2 The main-effects model of social support

The main-effects model of social support asserts that the provision of social support benefits the recipient regardless of whether they are experiencing stress or adversity (Cohen, 1988). Proponents of this theoretical perspective argue that social support improves individuals' overall psychological state, including sense of security and belonging (House, Umberson, & Landis, 1988). Support may promote well-being by providing individuals with regular positive experiences in ordinary social interactions (Lakey & Orehek, 2011).

In a study where couples reported their day-to-day mood changes, researchers found that companionship and positive talk predict greater well-being than discussions of stress and how to cope with it (Hicks & Diamond, 2008). The researchers posit that having a conversational partner to discuss both ordinary and stress-related topics with is beneficial to the recipient's well-being (Hicks & Diamond, 2008). Social support interventions could be particularly effective if they encourage recipients and providers of support to develop supportive relationships over time. Additionally, if perceived social support improves through positive interaction, social support interventions should encourage individuals to have meaningful conversations not only about stress and coping, but also about their interests and positive events. The interventions reported in this thesis were designed with these considerations in mind.

The main-effects model of social support has guided the intervention development phase of this thesis. The starting point for the intervention was the belief that social support is beneficial to all adolescents; not only those at risk of failing their courses or experiencing significant stress. Researchers who are particularly interested in the buffering effects of social support would need to identify individuals who are at-risk, and specifically recruit them into the study. The current study does not enforce such restrictive inclusion criteria.

2.3.2 Why is social support beneficial?

Social support is hypothesised to benefit recipients through increased positive mood and self-esteem (Collins & Feeney, 2004; Feeney, 2004; Feeney & Collins, 2015), improved self-efficacy (Coffman & Gilligan, 2002) and more effective coping with stressful events (Taylor, 2011). Finally, social support may benefit the individual by strengthening the relationship between provider and recipient (Burleson, 2003). Such increases in positive mood, coping and self-esteem are found across a range of beneficial outcomes, including physical and emotional well-being (Burleson & MacGeorge, 2002). It should be noted that most studies on the underlying mechanism between social support and well-being are focused on marital or intimate relationships (e.g. Collins & Feeney, 2004) which may not generalise well to support from friends or relatives. The study of social support mechanisms has gained prominence within the field of health psychology, but educational research has thus far provided limited empirical evidence on the pathways between social support and academic achievement.

One of the few empirical studies on this topic assessed the relationships between adolescents' self-reported maths enjoyment and anxiety, perceived support from parents, peers and teachers, and maths grades (Ahmed et al., 2010). The researchers propose that increased motivational beliefs (e.g. competence beliefs, interest in school) and positive affective experiences (e.g. enjoyment) as well as reduced negative affective experiences (e.g. anxiety) mediate the relationship between perceived support and academic achievement. Adolescents' perceived social support from parents, peers and teachers each correlated significantly with maths achievement (Ahmed et al., 2010). Multiple mediation analyses showed that support from parents significantly predicted competence, interest, enjoyment and lowered anxiety. In turn, each of these mediators significantly predicted maths achievement, (Ahmed et al., 2010). Fifty-five per cent of the effect of perceived parental support on maths achievement was indirectly explained by the motivational and affective variables. The pattern was similar but somewhat weaker for perceived peer support and teacher support. In summary, adolescents who perceived their parents, peers and teachers to be supportive felt less anxious, more confident, and enjoyed maths more, which in turn predicted better maths grades (Ahmed et al., 2010). This empirical study is consistent with earlier theoretical models on the links between social support (from parents) and academic achievement, such as the expectancy-value model (Eccles, 2007).

A greater understanding of mechanisms underlying the relationship between social support and academic outcomes may help develop effective social support intervention programmes. The qualitative components of this thesis will further explore students' reports of perceived social support, competence, interest, anxiety and enjoyment of maths and English.

2.3.3 When is social support beneficial?

Early studies on social support conceptualised social support as universally beneficial to the recipient (Taylor, 2011). More recently, it has become clear that the picture is more complex: social support can sometimes have negative consequences for recipients (Maisel & Gable, 2009; Rafaeli & Gleason, 2009). A number of contextual and relational factors determine whether the offered support is perceived as beneficial, including responsiveness of the provider, relationship closeness, spontaneity of supportive behaviour, gender of support provider and communicative skill (Bolger, Zuckerman, & Kessler, 2000; Cutrona, Cohen, & Igram, 1990; Cutrona, 1991; Gleason, Shrout, & Bolger, 2008; Maisel & Gable, 2009; Rafaeli & Gleason, 2009; Shrout et al., 2006)

In a series of experiments, Cutrona and colleagues (1990) studied the effect of contextual factors on the perceived helpfulness of supportive behaviours. The authors asked participants to read short stories about a student who just received the news that his/her mother was seriously injured in a car accident (the gender of the support recipient and provider was systematically varied). Participants then rated the perceived helpfulness of the offered support and general supportiveness of the relationship. In each of the scenarios, one person requests support and a second person offers support. The gender of the recipient, relationship closeness, type of support offered, and spontaneity of support offer were systematically varied, resulting in 16 versions of the story. Spontaneous support was judged more helpful than requested support, as was support provided within the context of a close relationship rather than casual friendship. Especially relevant to this thesis is the finding that optimal matching between desired support (e.g. emotional support) and received support (either emotional or informational support) resulted in higher perceived supportiveness (Cutrona, Cohen, & Igram, 1990). It therefore appears critical that supporters offer help spontaneously, but only when it is desired by the recipient. In this thesis, weekly text messages are intended to be a starting point for a supportive conversation, but the communications do not dictate the type of support offered.

The importance of matching between desired and provided support may explain the seeming paradox that well-intentioned support sometimes leads to feelings of inadequacy and inequity in the recipient (Maisel & Gable, 2009; Rafaeli & Gleason, 2009). Support attempts may be unskilled or misguided, for example through insufficient attention or sensitivity. When a student receives a poor grade for her maths exam she may be hoping for a warm embrace instead of advice on how to study better next time. As reviewed above, the mismatch between requested support (e.g. caring) and actual support transaction (e.g. advice giving) can lead to reduced trust and dissatisfaction with the relationship (Bolger et al., 2000). Support can be well-intentioned but poorly executed: if an individual receives support they did not ask for, feels misunderstood or undervalued, the experience may be a negative one.

Interestingly, invisible support, where "the provider reported enacting, but the recipient did not report receiving support" (Maisel & Gable, 2009, p. 928) is associated with more adaptive responses to stressors than support noticed by the recipient. Invisible support may take the shape of a subtle conversational approach. Several explanations for the finding that the clear provision of support is not always beneficial have been offered. First, the visible provision of support may signal to the recipient that they are incapable of coping independently, reducing their sense of autonomy and self-efficacy. Second, visible support may ironically highlight the stressful situation to the recipient. Third, receiving support may result in a feeling of indebtedness. Finally, visible support can be appraised as over-involvement (Rafaeli & Gleason, 2009).

Both observational and experimental studies find evidence for the relevance of support visibility. Observational studies have used daily diary responses (Bolger et al., 2000) and videotaped interactions (Maisel & Gable, 2009) and find that the perception of having received support is not unambiguously associated with better adjustment (Bolger et al., 2000). Bolger and Amarel (2007) randomly assigned participants to receiving support or no support before an anticipated public speech. In this procedure, a confederate peer either asked the experimenter for advice about presentation skills (i.e. invisible support), directly provided advice or reassurance to the participant (i.e. visible support) or did not offer support. Participants who received visible support prior to delivering the public speech reported feeling more anxious and upset than participants who received no support at all. Invisible support, on the other hand, lowered participants' emotional reactivity to the stressful speech (Bolger & Amarel, 2007).

The studies reviewed in this section underline that providing effective support is a skill: it certainly requires nuance to support others without giving them the feeling they are indebted or inadequate. Support providers can be guided to provide skilful support. For example, highlighting the importance of the recipient' need for autonomy may help providers of support account for such desires. Rather than dictating the type of support provided, the provider could ask the recipient for direction. Similarly, support providers can be reminded to adopt a constructive and problem-solving approach and to set aside their own anxieties and needs during such supportive interactions (Collins & Feeney, 2000). Finally, creating opportunities for reciprocated support where the initial recipient provides support in return may help counteract feelings of indebtedness. This can be encouraged by creating a habit of discussions about both partners' daily activities and concerns.

In summary, the literature on supportive communication suggests that the provision of social support is beneficial only when (1) it fits with the actual needs that arise, and (2) it is sensitive and responsive rather than controlling or intrusive. Social support interventions should thus be designed with the above considerations in mind: a rigid prescription of how to provide support is unlikely to aid skilful support. Further, interventions that aim to harness social support may benefit from educating support providers on the importance of matching support to needs.

This section introduced social support as a multidimensional concept and set out proposed mechanisms through which support aids student adjustment. The final segment discussed various contextual factors that may determine whether support is perceived as helpful. The next section reviews the practical application of the literature to the design of social support intervention and offers suggestions for optimal matching of support.

2.4 Mobilising support: a review of social support interventions

The term ‘social support intervention’ is used to refer to interventions that seek to foster interactions and supportive communication between individuals and their social networks. These studies are concerned with real-life interactions rather than lab-based studies reviewed above (see for example Bolger & Amarel, 2007; Cutrona et al., 1990). Social support interventions can be roughly divided into two types: those that introduce new ties, and those that intervene within the existing social network. I illustrate these in turn.

2.4.1 Group-based versus dyadic social support interventions

The first distinction within social support interventions is whether the interaction takes place in a group format, or in a dyad. A popular example of group-based support interventions is Alcoholics Anonymous (AA), where individuals hoping to recover from drugs or alcohol dependency join a small group of others who are dealing with the same issue but may be at different stages of the same journey. Having access to a group of similar others, who can provide both emotional and informational support, is hypothesised to aid addiction recovery (Kaskutas, 2009).

Group-based support interventions have shown to be effective across various settings. Breast cancer patients who were encouraged to offer each other support, discuss feelings, experiences and new ways of coping in twelve weekly sessions showed reductions in depression and anxiety symptoms at 6- and 12-month follow-up (Spiegel et al., 1999). Peer support groups are popular as they provide a safe space where participants can both receive and provide support (Hogan, Linden & Najarian, 2002). Reciprocal support was previously introduced as one of the relevant contextual factors that determine the effects of social support (Taylor, 2011). Finally, group-based support interventions also provide the peer supporters with an opportunity to build lasting social networks; which is predictive of well-being in itself (Cohen et al., 2000).

Examples of dyadic interventions are less plentiful. A typical dyadic intervention trains a family member or friends of the individual to provide more emotional support or improve their communication (Hogan et al., 2002). The recipients of support may struggle with a specific problem, such as a cancer diagnosis, eating disorders, substance abuse, obesity or chronic stress (Hogan et al., 2002). A pilot study with diabetes patients tested a peer-support system to test whether assigning patients a buddy could lead to improvements in health-related decisions (Rotheram-Borus et al., 2012). Volunteer peer mentors, who had diabetes themselves but had lost weight and increased exercise (i.e. positive role models), offered support to the women assigned to the buddy treatment by sending daily text messages (Rotheram-Borus et al., 2012). Unfortunately, the sample size is very small (N = 22), and the results were mixed, which is hardly a surprise given methodological limitations.

2.4.2 Introducing new ties, versus intervening on natural ties

A second distinguishing factor between social support interventions is whether they introduce new ties to the person in need of support or seek to leverage existing relationships. Examples of the first category come to mind easily: mentors, coaches, counsellors or home visitors have been mobilised to improve the support people receive or perceive to be available. In these cases, the researcher introduces a new social tie to the individual and assesses whether the recipient of support improves on a pre-defined dimension (Gottlieb, 2000). For example, Colella and King-Shier (2017) investigated whether a peer support intervention could improve recovery of men who had recently undergone surgery to improve symptoms of cardiovascular disease. They recruited 'peer volunteers' who had successfully undergone the surgery themselves. The volunteers were trained to provide telephone support to the patients. They learnt about active listening, the value of sharing experiences, and building support. Eligible patients were then randomised to receive either weekly telephone calls over a period of 6 weeks, or usual care. The intervention did not significantly improve patients' depression scores or perception of social support. However, treated individuals used fewer health services, such as emergency rooms (Colella & King-Shier, 2017). The sample size of the treatment group was relatively small (N = 61), so future studies should seek to replicate this study at a larger scale. Nevertheless, this study is a good example of an intervention where new ties are introduced with the aim to enhance social support.

An alternative strand of social support interventions relies on intervening within the individual's natural network. The studies reported in this thesis are an example of this approach. Intervening not only on the individual of interest but also their immediate social environments may help create environmental conditions that support the uptake and maintenance of behaviour change. For example, a family support intervention where obese adolescents' parents learnt about creating a healthier and more autonomy-supportive home environment found that treated students consumed more fruit and less junk food than non-treated peers even a year post-intervention (Straker et al., 2014). Most social support interventions reviewed in this section concern health-related behaviour change. A number of education interventions can also be categorised as social support interventions, but intriguingly they rarely use social support theory (see Froiland, 2011; Miller, Davison, Yohanis, Sloan, & Gildea, 201). Parental engagement interventions aiming to improve parent-child communication are a clear example, as they target the family structure.

A modest but growing number of studies attempt to boost supportive communication between students and their parents with simple, personalised prompts (Bergman, 2012; Bergman & Chan, 2017; Chande, 2016; Kraft & Dougherty, 2013; Kraft & Monti-Nussbaum, 2017; Kraft & Rogers, 2015; Miller et al., 2017; Robinson & Lee, 2017; Rogers & Feller, 2018; York, Loeb, & Doss, 2018). The studies reviewed below have in common that they convey specific information to parents on how to change their own and their child's behaviour. These interventions attempt to alter the home environment, enabling parents to better support their children through provision of actionable advice. I argue that this type of intervention is easily extended from parents to the adolescent's wider social network. Whether the young person relies on her parents for support and guidance, or whether another adult assumes this responsibility may not be of crucial importance. Instead, it may be key that the intervention prompts those whom the young person identifies as a close, trusted relationship.

2.4.3 Questioning the support process in interventions

Few social support intervention studies examine the messages communicated by support providers, or how these are processed by recipients. The previous section detailed the importance of matching between desired and enacted support, but it is challenging to assess to what degree such matching occurs during implementation. For example, Carlson and colleagues (2002) implemented a non-experimental smoking cessation intervention and invited participants to bring along a support person to the sessions (only 26% of the sample did so). Access to the social support component was offered to all participants; this was not randomised. The supporters met to discuss a variety of topics, including techniques for smoking cessation, expected withdrawal symptoms and supportive versus critical behaviour. These sessions were designed to educate close relationships of the individual enrolled in the programme. The rates of successful smoking cessation at three months post-quitting were significantly higher for individuals who brought a support person along in comparison to those who did not (Carlson et al., 2002).

Since cessation rates are the only outcomes reported, it is unclear why the inclusion of the social element had any effect on cessation rates, or indeed, whether individuals who brought along a support person were different to those who did not from the start. Did the information sessions help the support people to communicate more tolerantly, rather than critically? Did the support people feel empowered to take additional measures, such as removing all cigarettes from the home? What type of people self-select buddy support and to what degree is this related to successful smoking cessation? Studies that are able to shed light on the support processes taking place within intervention settings are in short supply. Later chapters in this thesis aim to fill this gap through a qualitative exploration of supportive communication processes.

This section has introduced a variety of configurations of social support interventions. They can focus on existing ties or introduce new ones; and foster social support within a dyadic relationship or a larger group. This thesis focuses on existing social ties and dyadic relationships. I argue that a focus on existing relationships can help build long-lasting behaviour change through positive behaviour modelling and modifying home environments.

The next section delves deeper into the barriers close relationships face to becoming more involved in the student' education. I then turn to studies that shed light on the ways family members and other close relationships can be encouraged to become more involved in building a supportive learning environment.

2.5 Improving educational outcomes by informing friends and family

2.5.1 Why would close relationships need a nudge to become more involved?

Consistent parenting is not an easy task: multiple tasks may be competing for the parent's attention, and the returns of spending time with one's children are only visible far into the future (Mayer, Kalil, Oreopoulos, & Gallegos, 2015). Parental involvement with the child's education declines as their child moves from primary to secondary education and beyond (Hoover-Dempsey et al., 2005). Parents may feel they do not have sufficient knowledge of the more advanced topics, or believe that they do not have the resources to help (Hoover-Dempsey et al., 2005). This set of findings can be generalised to the student's wider social support network. Grandparents, aunts, uncles, friends, or brothers or sisters may want to be involved in the student's education, but not know how to.

What parenting behaviours are particularly beneficial to the development and wellbeing of children? The social support literature suggests that it is especially important for parents to foster positive learning environments (Wentzel et al., 2016). Parents who take an interest in their child's education, participate in parents' evenings, communicate with their child about homework or school activities raise children who do better in school (Desforges & Abouchaar, 2003; Fan & Chen, 2001; Jeynes, 2007; Pomerantz, Moorman, & Litwack, 2007). Such positive parenting behaviours can be stimulated. Jeynes (2005) conducted a meta-analysis of 52 studies and found that parental involvement programs lead to 0.36 of a standard deviation increase in grades and other measures of academic achievement. Beyond getting involved with school events and checking homework, parental beliefs and communication also appear to be of importance. For example, mothers' positive attitude towards and communication about maths and science positively predicts their adolescent children's' interest in these courses, as well as their actual subject choice (Hyde et al., 2017).

The next section reviews a number of text-messaging interventions which show that greater involvement can be stimulated through relatively simple and low-cost text-message alerts.

2.5.2 Supportive information interventions

Two studies provided the initial motivation for this thesis. I first introduce a study conducted by Kraft and Rogers (2015) which focuses on providing parents with frequent and personalised information about their child's learning. The second study (York, Loeb & Doss, 2018)[9] was conducted with a much younger cohort of students but helped inform the thematic categories of the text messages in my work.

A pioneering field experiment with 435 high school students and their parents sought to improve regular parent-teacher and parent-child communication about school (Kraft & Rogers, 2015). Teachers wrote one-sentence individualised messages to parents about their child's behaviour and performance in a high school credit-recovery programme, over the course of four weeks. Parents were randomly assigned to one of three conditions: (1) positive feedback: what their child was doing well, (2) improvement feedback: what their child needed to improve on, or (3) control: no communications. Both parents and teachers were blind to condition. Taking both types of messages together, the weekly message resulted in a 6.5 percentage point increase in the probability the student passed the class; or a 41 per cent reduction in failing to earn the credit. This study falls within a category of information interventions aiming to increase parental monitoring.

York, Loeb and Doss (2018) designed an intervention to help parents support their preschool aged child's literacy development. Parents randomly assigned to the treatment group received three texts per week about specific activities they could engage in. Control parents received texts about a topic irrelevant to literacy development. The texts were worded positively and were designed to inform and motivate. In this 8-month texting program, treated parents more frequently told stories, recited nursery rhymes, worked

[9] An earlier version of this study was published in 2014, as a working paper with the National Bureau of Economic Research (NBER).

on puzzles with their children, and pointed out words that rhyme than those who received only placebo texts. This increase in parental involvement translated into learning gains for treated children of 0.11 standard deviations. Teachers also reported that parents in the treatment group asked more questions about their children's learning than control group parents. The intervention was highly cost-effective, as it cost on average only one dollar per family to deliver. Additionally, the program placed few demands on the parents' time as the text messages were designed to be easy to implement (York, Loeb, & Doss, 2018).

In the years since these two studies were conducted, the evidence base on the effectiveness of information interventions has continued to grow: timely information delivered to parents can help improve academic outcomes. I will now review a number of subsequent intervention studies and focus specifically on variations in design and delivery. The interventions summarised in Table 2.1 have in common that they aim to (1) send information about the child's learning, (2) highlight the importance of schooling, and (3) encourage parents to communicate with and offer guidance to their child. All studies reported in the table below were evaluated using a randomised controlled trial (RCT) design.

The common denominator of the information interventions that inspired this thesis is not the mode of delivery, but their focus on providing actionable and simple information to parents. Two types of interventions can be distinguished. The first type of information interventions delivers factual and tailored information about student behaviour to parents. They primarily aim to prompt parents to take a more active interest in behaviours such as truancy, missed homework or low grades. For example, Bergman and Chan (2017) and Rogers and Feller (2018) seek to address incorrect parental beliefs about their child's absences or missed assignments. They hypothesise that parents overestimate the prevalence of absenteeism in general (i.e. among other classmates) or may not be aware their child missed assignments. The researchers seek to correct these incorrect beliefs by providing parents with personalised and timely information about the actual occurrence of such events. These prompts, whether delivered via mail (Rogers & Feller, 2018) or text message (Bergman & Chan, 2017), are designed to increase parental monitoring and with it the frequency with which their children disclose their academic progress and effort.

Table 2.1: Overview of similar interventions

Paper and intervention	N	Delivery	Results
Avvisati et al. (2014). Inviting low SES parents to parent-school meetings on how to become more involved in educations.	970	Face-to-face parent-school meetings	Positive and significant: students' truancy and disciplinary sanctions reduced by 0.10 – 0.20 SD, and grades improved by 0.12 SD, but no effect on standardised tests.
Bergman & Chan (2017). Parents received automated text alerts about missed assignments, missed attendances, and test scores.	1,137	Automated text messages	Positive and significant; 18% increase in class attendance and 39% reduction in course failure. Effects on standardized tests are non-significant, scores on class tests improve by 0.13 SD.
Kraft & Dougherty (2013). Teachers called parents daily to discuss the week's teaching, 4-week summer academy.	145	Teacher-parent phone calls	Positive and significant; increased odds of homework completion by 40%, on-task behavior by 25%, and class participation by 15%.
Miller et al. (2017). Texting parents of secondary school pupils about upcoming deadlines and class topics.	15,697	Text messages	Positive and significant for maths, at 1 month additional progress. Effects on science and English subject not statistically significant.
Rogers & Feller (2018). Parents received letters about their children's absences, up to three times throughout the year.	28,080	Letters	Positive and significant, most effective treatment reduced total absences by 6% and chronic absenteeism by 10% relative to control.
Castleman & Page (2017). College-intending high school seniors and their parents receive texts about college and financial aid tasks over the summer.	4,754	Text messages	Positive but not significant. Timely enrollment increases by 3.2 % points when both students and parents receive texts, and 2.9% points when students only are texted ($p < 0.10$).
York, Loeb & Doss (2018). Parents of pre-school children received weekly texts, encouraging them to engage in low-level literacy exercises.	935	Text messages	Positive and significant, weekly texts increased home literacy activities by 0.21 – 0.34 SD, and gains in child's early literacy scores of 0.11 SD.

A second set of information interventions listed in Table 2.1 seeks to help parents overcome barriers to good parenting, by providing them with easy-to-achieve steps, encouragement, and reinforcement. York, Loeb and Doss (2018) provided simple ideas for early literacy exercises, encouraging parents to build positive home literacy habits. Avvisati and colleagues (2014) invited parents to attend school meetings in order to boost their knowledge of and confidence with the school environment. The sessions offered advice to parents about how to help their children at home, highlighted the importance of homework, and underlined the importance of parents as role models. All in all, they argue that "what matters most is that children feel their parents are interested in their school experience, and feel encouraged to talk often about it" (Avvisati, Gurgand, Guyon, & Maurin, 2014, p. 61).

Interventions that aim to foster shared knowledge and understanding between the young person and her immediate social context are the focus of this thesis. Rather than informing friends or family about poor behaviour or results, I focus on encouraging them to have engaging conversations about course topics and planning for upcoming exams. Finally, this thesis is unique in its focus on empowering wider social networks to become more involved. All of the reviewed information interventions focus on the parent-child relationship only, which is logical for school-aged children but less so for post-16 students. The next section asks whether the provision of personalised and detailed information is a pre-requisite for effective information interventions.

2.5.3 Is personalised information about student behaviour necessary?

Informing parents of their child's behaviour in school can help lower monitoring costs. The informative texts may also increase the salience of the benefits of parental monitoring (Cunha, Lichand, Madeira, & Bettinger, 2017). It is also possible that these information interventions simply increase the cognitive accessibility of a behaviour (e.g. asking the student what they learnt in class). Lastly, parents may have biased beliefs about the child's behaviour and performance in school (Bergman, 2017; Rogers & Feller, 2018).

Few studies have attempted to disentangle the relative merits of the above three potential explanations. A recent study attempted to disentangle whether texts lowers monitoring costs or increases the salience of monitoring benefits, or both (Cunha et al., 2017). Specifically, they test whether providing fine-grained information drives the effectiveness of information interventions (arguably lowering monitoring costs) or whether it is possible to provide parents with salient information about the benefits of attendance and assignment completion while omitting customised information. The researchers randomly assigned parents of 19,300 secondary school pupils in São Paulo, Brasil, to receive either weekly information messages, awareness messages, or no communication over the course of 18 weeks. The information messages conveyed information about the child's attendance, lateness and assignment completion, mirroring both Bergman and Chan's (2017) and Rogers and Feller's (2018) approach. The awareness messages did not contain personalised information, and only aimed to raise parents' awareness about punctuality, assignment completion and class attendance. An awareness message read, for example, that "for a good school performance, it is important that [student name] doesn't miss school for no reason" (Cunha et al., 2017, p. 10).

The awareness messages led to comparable and sometimes larger improvements in student attendance and GPA in comparison to the messages that contained detailed and personalised information about student behaviour. The simple awareness messages improved outcomes by 89 - 129% of the effects of information messages (Cunha et al., 2017). The authors conclude that information interventions may not be effective because they correct parents' misbeliefs, but because they focus their attention on variables we know are important for educational outcomes: attendance, punctuality and assignment completion (Cunha et al., 2017). This study shows that general text messages about students' learning and class attendance may be just as effective as personalised and administrative data-focused messages. When the infrastructure for integration between college data and text-messaging platforms is lacking, general text messages about students' learning is the only option. The two experiments discussed in Chapter 4 and 5 were implemented under such constraints.

2.5.4 Parental monitoring and students' need for autonomy

Some information interventions show large effects (e.g. Bergman & Chan, 2017), other small or insignificant ones (e.g. Castleman & Page, 2017).

Section 2.3.3, which addressed the question of when social support is beneficial, highlighted the importance of context (see p.30). As reviewed in detail, the abovementioned information interventions aim to increase parental monitoring behaviour (see for example Kraft & Monti-Nussbaum, *2017*). Parental monitoring may have different effects for different groups of individuals or family settings (Jacobson & Crockett, 2000). The relationship between parental monitoring and student adjustment varies by gender, age, and family composition (Jacobson & Crockett, 2000). Some interventions may have been implemented in environments that benefited from greater parental monitoring. Another factor that may determine the effectiveness of such parent-school information interventions is whether the young person feels that their need for autonomy is respected. Educational interventions often fail because they have not sufficiently accounted for the adolescent's enhanced desire to feel respected (Yeager, Dahl, & Dweck, 2018). Controlling parenting styles, such as constant monitoring of their child's behaviour, being highly demanding, or not asking for the child's input in discussions, are associated with diminished self-efficacy and less effective family communication (Givertz & Segrin, 2014). These findings highlight the need for positively-worded communications that encourage family and friends to provide support and comfort.

2.6 Why would students sign up voluntarily?

An important distinction between the information interventions reviewed above and the studies described in the next chapters of this thesis, is that students in our sample voluntarily opted in to take part in a year-long programme of text messages. None of the information interventions listed in Table 2.1 relied on students to consent to having information about their learning shared with their parents (see p. 40). Chapter 3, which discusses the methodological choices made in this thesis, reviews how this design feature affects the inferences that can be made from the studies. In this section, the following question takes a central place: what might motivate adolescents to voluntarily sign up to an educational intervention that potentially increases monitoring?

Perhaps the answer can be found in the emerging literature on commitment devices. Students constantly have to make choices between immediate and delayed gratification (e.g. leisure versus assignment completion). We often discount the future costs of a behaviour, and instead prioritise its immediate benefits (Brocas, Carrillo, & Dewatripont, 2004). A student may desire to complete her qualification by the end of the year, but may also desire to spend her afternoon out with her friends instead of spending it in the college library. Indeed, many people intend to improve their behaviour in the future, but when the future arrives they may fail to follow through (Rogers, Milkman, & Volpp, 2014). After all, the future has now become the present, in which leisure activities may be more immediately desirable than long-term goals such as achieving a qualification.

In order to self-regulate behaviour, people may choose to sign up to a commitment device. Commitment devices help individuals to “pre-commit their ‘future selves’ to follow-through” (Rogers, Milkman, & Volpp, 2014, p. 1). An example commitment device is to schedule gym workouts with a friend to increase the embarassment one would feel if they skipped (Rogers, Milkman, & Volpp, 2014). The programme of text messages could serve a function akin to a commitment device: students may want to sign up to increase their sense of accountability. In this scenario, students deliberately limit their future behaviour by imposing monitoring upon themselves, in the form of a supporter checking in with them about their learning. Alternatively, students could sign up because they value the prospect of receiving support from their friends and family, or hope the weekly texts breaks down communication barriers. These alternative explanations for the attractiveness of an information intervention are further explored in Chapter 6 (Section 6.2, p. 215).

2.7 Conclusion

The primary research question addressed in this thesis is whether encouraging friends and family to take an active interest in the student’s learning helps improve attendance and attainment. As a secondary question, this thesis asks whether the informative text messages are more effective when addressed directly to the student or delivered via a close third party. In asking the secondary question, this thesis aims to gain a better understanding of the mechanism of the information intervention: is it the informational content of the texts that benefits students, or does the activation of students’ social network produce benefits over and above the information contained in the texts?

Two literatures underpin this thesis: the social support literature and the emerging literature of behavioural science and information interventions. The first forms the theoretical foundation for the development of the intervention and informed the qualitative component of the randomised controlled trials. The literature on social support illuminates the importance of interpersonal relations for students' educational outcomes and well-being. The literature points to reduced negative effect and better coping with stressors as mechanisms between the provision of support and improved academic success. Increasing the quality and quantity of supportive conversations between students and the influential others who care about their learning is hypothesised to trigger a recursive process of improved self-confidence, motivation and effort, in turn leading to better educational outcomes.

Second, the recent wave of information interventions has informed the design of this thesis. These interventions attempt to boost supportive communication between students and their parents with simple, personalised prompts. The messages aim to overcome behavioural barriers to engagement, such as inattention and inertia. Parents or influential others may not know how to get involved, may not have the tools or resources to do so, or may put it off. Similarly, the student may not know how to best ask for help. These barriers to supportive communication are addressed in this thesis.

Providing individuals with positive encouragement positively impacts academic achievement, as well as motivation and emotion (Soenens & Vansteenkiste, 2005). However, this literature has been rarely applied to intervention studies in the field, instead focusing on lab-based experiments or observational studies. At the same time, the majority of parent-child communication interventions focus solely on providing detailed but often negative information about the child's behaviour (e.g. low grades or missing homework). This thesis combines the literatures on harnessing social support with parent-child information interventions and asks the question whether friends and family can be encouraged to provide skilful support to students via low-cost and light-touch prompts.

3 METHODOLOGY REVIEW

3.1 Introduction

Recent experimental studies suggest that informing students' family about their learning can promote classroom attendance and attainment (Bergman & Chan, 2017; Cunha et al., 2017; Kraft & Rogers, 2015; Miller et al., 2017; Rogers & Feller, 2018). These studies were natural field experiments (see Harrison & List, 2004, p. 1014): parents received communications from the school their child attended, and the communications were often (at least in part) authored by the teachers themselves. The second important attribute of these information interventions is that they used randomisation. They randomly assigned a subset of the total participant pool to receive the school communications, and those assigned to a 'control group' received the business-as-usual. The latter group may still receive phone calls or letters from the school, but they do not receive the intervention materials. Randomisation ensures that every subject has the same probability of being treated as every other subject in the pool (Gerber & Green, 2012). This type of study design is commonly referred to as a 'field experiment' or 'randomised controlled trial' (RCT).

RCTs, if well-designed and properly implemented, are capable of isolating the causal effect of an intervention on outcomes of interest (Gerber & Green, 2012). This methodology is especially well-suited to establish whether an intervention is effective at achieving a specific and measurable outcome (Hanley, Chambers, & Haslam, 2016). The Education Endowment Foundation (EEF), a charity set up by the Department for Education (DfE) to identify, fund and evaluate promising educational innovations, states that "wherever possible this will mean using a randomised control trial – the gold standard of educational research" in their first annual report (Education Endowment Foundation, 2012, p. 17).

The challenge with many RCTs is that the social context in which they are implemented is de-prioritised in favour of control over the experimental manipulation. RCTs are popular especially because they do not require substantive theoretical assumptions about covariates or confounders (Deaton & Cartwright, 2017). Although this is certainly a strength of RCTs, the transferability of findings can be threatened by insufficient focus on "the complexity of social causation" (Bonell et al., 2012, p. 2299). Combining a field experiment with qualitative instruments and designs has the potential to triangulate and contextualise its design, implementation and findings (Paluck, 2010). Furthermore, some outcome measures and mechanisms cannot be collected quantitatively, such as subtle descriptions of social support (see Chapter 6). Additionally, qualitative inquiry allows the researcher to uncover unexpected processes or experiences. Exploratory quantitative analysis is in some sense confirmatory because it requires ex-ante decisions about what measures to include.

In isolation, both quantitative and qualitative approaches have their weaknesses, such as the lack of focus on nuances within communities studied for the former and poor generalisability for the latter (Bamberger, 2012). Mixed methods as a research method explicitly draws on an integration of the two, and therefore seeks out their commonalities rather than differences (Johnson, & Onwuegbuzie, 2004). Using each method to answer a related question is a key strength of mixed methods designs (Palinkas et al., 2011). The chosen research design of this thesis is a field experimental approach, combined with an embedded qualitative component. The qualitative data are used to triangulate and contextualise the implementation and findings.

Finally, a note on terminology. Randomised experiment, randomised controlled trial (RCT), natural field experiment, random assignment studies, or randomised evaluation all describe the same research design, where whether one receives a treatment is determined randomly. Harrison and List (2004) helpfully devised a taxonomy of field experiments, and refer to experiments "where the environment is one where the subjects naturally undertake these tasks" (p. 1014) as natural field experiments. Throughout this thesis, I will use the vocabularies of 'RCT' and (natural) 'field experiment' interchangeably.

This chapter will firstly review the underlying theory and design features of the natural field experiment, and critically discuss potential threats to internal and external validity. The latter part of this chapter examines the potential of a mixed methods approach to field experimentation. As part of this section, the challenges inherent to mixing methods and qualitative inquiry are addressed. Finally, the approach taken in this thesis is laid out.

3.2 Causal inference and randomised controlled trials

An experiment, when designed and implemented well, is a relatively straightforward way to uncover the causal effect of a programme, intervention, or policy (Athey & Imbens, 2017; Shadish, Cook, & Campbell, 2002). In the most basic sense, when one set of units is assigned to treatment, the new approach, and the other set of units to control, this is an experiment (Tymms, 2012). When assignment is random, it is known as a true experiment, and when it is not it is known as a quasi-experiment (Tymms, 2012). As introduced above, random allocation sets the RCT apart from other evaluation approaches, such as quasi-experimental or observational approaches. An RCT is able to address questions that other methods cannot, as two examples below illustrate. I first take a narrative approach to the importance of randomisation before turning to a more formal statement of causal inference in Section 3.2.1 (p. 49).

Pakter and Chen (2013) implemented and evaluated a text messaging intervention where parents received regular communication from teachers. The design of the intervention is comparable to the intervention described in this thesis, and the researchers ask similar research questions. They aim to understand "what impact the use of text messages between teachers and parents have [...] on the student's learning" (Pakter & Chen, 2013, p. 358). The researchers contacted the parents of students who volunteered for the study; 35 parents agreed to receive text messages every two days. The parents who were not contactable or who did not opt in served as a comparison group. Outcome variables collected included grades in course and the percentage of school days missed. Pakter and Chen (2013) found that neither student attendance nor their attainment was affected by the texts. They conclude that "the results seemed to indicate that whatever influences the text messages may have had, this influence was dwarfed by other events in the students' lived" (Pakter & Chen, 2013, p. 362). In reality, the research design is not able to answer the research questions in a robust and unbiased manner. Parents self-selected into groups; those who did not wish to receive text communications from parents may be different from parents who did sign up both in observable and unobservable ways. The issue of selection bias is clear-cut in the above example. In the absence of random allocation, any difference in means between the texting and non-texting group may be due to pre-existing differences (Duflo, Glennerster, & Kremer, 2008). Randomisation allows the researcher to create credible counterfactuals. The importance of randomisation and the issue of selection bias are further discussed in section 3.4.3 (p. 60). Further, the sample size of the Pakter and Chen (2013) study ($N_{treated}$ = 29) is very small. Even if the researchers had used randomisation, the probability that their finding reflects a true effect is small due to potential randomisation failure and the under-powered nature of the study (Button et al., 2013).

Researchers are unable to make causal inferences when they rely on correlational data. The following study examined the relationship between perceived social support from parents, teachers and peers and academic outcomes. Wentzel and colleagues (2016) surveyed 398 secondary school students and their teachers in the US, about perceived emotional support, parental expectations, effort in school, and prosocial behaviour. Using multilevel modelling techniques, the researchers explored whether emotional support from parents, teachers and peers are predictors of students' academic outcomes. Perceived emotional support from parents was predictive of better academic grades, and peer social expectations was predictive of responsible behaviour (Wentzel et al., 2016). Although these findings can shed light on the importance of emotional support for academic outcomes, they cannot directly evaluate the direction of effects. For example, good grades could elicit increased emotional support from parents, or emotional support could enable students to focus on their learning. We can only conclude that there is an association between the two observed variables. Indeed, "correlation does not imply causation" (Pearl, 2009, p. 99). Observational studies are unable to substantiate causal claims. To understand why randomised experiments provide particularly reliable assessments of cause and effect, I now turn towards the potential outcomes framework as first introduced by Rubin (1974) and Holland (1986).

3.2.1 Principles of causal inference

Researchers who are interested in measuring the causal effects of a new social support programme would theoretically need to observe the participant both under the treatment and control condition, at the same point in time. They would want to understand whether being assigned to having a study supporter, who then receives weekly texts, leads to better attendance and achievement than not having a study supporter. However, one can never simultaneously observe a participant's outcomes when they both receive and do not receive a treatment. If a participant is assigned to having a supporter at t_1, and their outcomes are assessed at the end of the academic year at t_2, the researcher cannot go back in time to t_1 to expose the same participant to the control condition. This is a problem of missing counterfactuals (Rubin, 2005).

A counterfactual can be described as "what would have happened to the participant under the treatment condition and what would have happened to the same participant under control condition under identical circumstances" (West, Biesanz, & Pitts, 2000, p. 41) and vice versa. We can therefore never directly observe the causal effects, as participants are always assigned to one level of the treatment only (Athey & Imbens, 2017). In other words, "each potential outcome is observable, but we can never observe all of them" (Rubin, 2005, p. 323), which is known as the "fundamental problem of causal inference" (Holland, 1986, p. 947). One can only compare different units with different levels of the treatment against one another. It is therefore essential to find a valid counterfactual; the two groups should, on average, have similar characteristics. A randomised controlled trial is a relatively straightforward way to construct valid counterfactuals, and is the chosen method in this PhD thesis. To justify this research design, I introduce Rubin's Causal Model, which provides an answer to the missing counterfactual problem.

Rubin's Causal Model (RCM) labelled as such by Holland (1986), provides us with a compelling model for causal inference. In a randomised experiment, each subject is "*potentially exposable"* to the action of a cause (Holland, 1986). As set out in the previous section, subjects are randomly assigned to two or more conditions, where (1) implies the active treatment and (0) implies the control condition. The causal effects are then measured on the outcome variable, Y for a particular subject (*i*) and interval of time. The word *potential* is crucial to this model. It denotes that we can observe the subject's educational outcomes if they had a Study Supporter, $Y_i(1)$, or the subject's educational outcomes if their Study Supporter would not receive texts, $Y_i(0)$. The treatment assignment determines which potential educational outcome we observe. For any unit, the potential outcomes would be (Angrist & Pischke, 2008, p. 11):

[1] Potential outcome = $\begin{cases} Y_{1i} \text{ } if \text{ } D_i = 1 \\ Y_{0i} \text{ } if \text{ } D_i = 0 \end{cases}$

$Y_i(0)$ is the potential outcome for subject *i* had they been assigned to control (0) irrespective of whether they were actually assigned to control, while $Y_i(1)$ is its potential outcome if they were treated (1). *D* indexes the binary treatment variable {0, 1}. The difference between $Y_i(1)$ and $Y_i(0)$ is the causal effect of the treatment on subject *i*. However, we are facing an empirical challenge. If subject *i* was assigned to the treatment group ($D_i = 1$), we would never be able to actually observe $Y_i(0)$. The actual observed outcome can be written in terms of potential outcomes as:

[2] $$Y_i = \begin{cases} Y_{1i} \text{ } if \text{ } D_i = 1 \\ Y_{0i} \text{ } if \text{ } D_i = 0 \end{cases} = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

Because D_i is either 0 or 1, one of the two terms will always be zero, thus we observe the potential outcome resulting from treatment $Y_i(1)$ if treatment is administered, ($D_i = 1$) and we observe $Y_i(0)$ if no treatment was administered.

We are generally interested in estimating the average causal effect[10] among a population of interest (such as FE college students, where i = 1, ..., N), which is defined as the difference between the two potential outcomes:

[3] $$\frac{1}{N} \sum_{i=1}^{N} Y_i(1) - Y_i(0)$$

As set out above however, due to missing counterfactuals we are never able to directly observe the unit-level, nor population-level (average) causal effect. We expose some units to the treatment (1) and others to the control (0). By comparing those who were allocated to treatment versus control, we hope to learn about the effects of the active treatment on outcome Y. We would like to take the expected average of both groups, $E[Y_i(1)|D_i=1]$ and $E[Y_i(0)|D_i=0]$ and calculate the difference between means to assess the difference in educational outcomes between treated and non-treated subjects. However, we have to take account of the alternative potential outcomes, or the expected outcome for a subject in the treatment group had they not been treated and vice versa. These potential outcomes cannot be observed, but can be imagined (Gerber & Green, 2012). In the absence of randomisation, this converges to:

[10] Which falls within the category of "summary causal effects" (Rubin, 2005, p. 323), and is also often referred to as the average treatment effect, or ATE (Gerber & Green, 2012).

$$[4] \qquad E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] = E[Y_i(1) - Y_i(0)\ |D_i = 1]$$
$$+ E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

E denotes the expectation operator, or the average outcome of a variable. The top right equation calculates the average treatment effect on the treated (ATT), which is the treatment effect we would like to isolate. The bottom right part of the equation represents the selection bias.

$E[Y_i(0)|D_i{=}1] - E[Y_i(0)|D_i{=}0]$ captures the difference in potential untreated outcomes between the treatment and control subjects. To illustrate selection bias, I return to the example of the social support intervention. Treated students whose supporter receives text messages may have had different educational outcomes on average, even if they had not been treated. This could be true when students who are motivated to do well in school were especially keen to sign up for the programme. In this case, $E[Y_i(0)|D_i{=}1]$ would be larger than $E[Y_i(0)|D_i{=}0]$. Alternatively, schools could encourage especially the students that struggle most in school to take part, in which case $E[Y_i(0)|D_i{=}1]$ would likely be smaller than $E[Y_i(0)|D_i{=}0]$. In other words, there may be systematic differences between students who sign up to the programme, and those who do not. $E[Y_i(0)|D_i{=}1]$ is not observed which makes the causal effect of the treatment D_i difficult to uncover.

Random assignment addresses this inferential problem as it eliminates selection bias (Angrist & Pischke, 2008). Random assignment of D_i means that potential outcomes $Y_i(0)$ and $Y_i(1)$ are completely independent of allocation to $D_i = 0$ versus $D_i = 1$. Because of this independence, subjects assigned to treatment versus control only differ through their exposure to D. This principle is illustrated below:

$$[5] \qquad E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$
$$= E[Y_i(1)|D_i = 1] - E[Y(0)_i|D_i = 1]$$

This independence allows us to swap $E[Y_i(0)|D_i{=}1]$ for $E[Y_i(0)|D_i{=}0]$ in the second line, simplifying the equation to:

$$[6] \qquad E[Y_{1i} - Y_{0i}|D_1]$$

Which corresponds to the average causal effect for those units who were treated, or the ATT. Since D_i is now completely independent of Y_0 and Y_1, we can further simplify the equation to:

[7] $$E[Y_{1i} - Y_{0i}],$$

or the average treatment effect, ATE (Gerber & Green, 2012). If the treatments are randomly assigned, the selection bias $E[Y_i(0)|D=1]$ - $E[Y_i(0)|D=0]$ is zero. Under randomisation the difference in Y between $Y_i(1)$ and $Y_i(0)$ is an *unbiased* estimate of the causal effect. Subject to large enough sample sizes to overcome small sample bias (Button et al., 2013), random assignment ensures conditions differ only with respect to the treatment assignment; variation in characteristics of units are evenly distributed between conditions. Additionally, being assigned to (1) or (0) is an observed characteristic for the units in any trial. Randomisation allows us to be confident that assignment does not depend on unobserved characteristics of the units (Athey & Imbens, 2017). A number of assumptions need to be satisfied under the potential outcomes framework, which I will turn to now.

3.2.2 Assumptions of the potential outcomes framework

Obtaining causal effects from comparisons between subjects necessitates the assumption of SUTVA, or 'stable unit treatment value assumption' (Rubin, 2005). SUTVA holds that (1) there are no hidden variations in treatment, and (2) the potential outcomes of unit Y_1 depend solely on its treatment received, or that there is no interference between units (Rubin, 2005). This assumption of non-interference stipulates that potential outcomes for subject *i* reflect only their own treatment or control status and do not depend on the treatment or control status of any other observations (Gerber & Green, 2012; Rubin, 2005). SUTVA is essential to the inference of causal effects, and an important consideration when designing field trials. For example, researchers may decide to randomise at the cluster level (e.g. classroom) if they are worried that the assignment of individuals within a given classroom influences the outcomes for others. If a student who receives informational text messages shares these with their peer assigned to control, there is interference between units. In this case, treatment is transmitted from treated to untreated subjects. It is common for educational RCTs to randomise at the cluster-level in order to protect against spillover effects (Torgerson & Torgerson, 2008). As mentioned above, SUTVA also requires that there are no different levels of a given treatment. This can be an issue in field experimentation and will be further addressed in the section on internal validity, 3.3.

The second assumption is that of excludability, where potential outcomes only respond to random assignment of the treatment (Gerber & Green, 2012, p. 39). For example, if researchers used different methods to assess outcomes in the treatment and control groups, potential outcomes could respond to differences in measurement, rather than whether or not they were exposed to the treatment. The same procedures and questionnaires should always be administered to both treatment and control groups, and ideally at the same time and under similar conditions. In this thesis, individuals are individually randomised to treatment and control within classrooms. This helps protect against variation in measurement, as different colleges may report class attendance in different ways. It would be unlikely that a teacher measures attendance in different ways, as they adhere to a register key set by the college. Randomisation at the college level could potentially be problematic, especially as I rely on college-reported administrative data to make causal inferences.

Another example where the assumption of excludability is violated, is when parents know that their child is assigned to the control condition and provide them with additional time and support to ensure they do not feel left out. For this reason, it is important that experiments are blinded: where possible, schools, parents, pupils and ideally researchers should not be aware of treatment assignment. This ensures uniform handling of treatment and control groups (Gerber & Green, 2012). In this thesis, teachers and administrators at participating colleges are not informed of students' assignment to conditions. Whereas it remains important to chart whether colleges launched any additional interventions during the trial period, it is unlikely that random assignment set additional college initiatives in motion since they were blinded to allocations. Nevertheless, teachers could have asked their students if their nominated study supporter received any text messages, and any additional effort on their part to support non-treated students would jeopardise the field experiment. Fortunately, these assumptions can be checked and solutions to potential violations can be built into experimental designs, as reviewed in the next section.

3.3 Confirming internal validity

The internal validity of RCTs can be high, provided that they are implemented well (Peters, Langbein, & Roberts, 2016). Various issues in the design and implementation of RCTs can impair internal validity, threatening the researcher's ability to conclude that the measured effect is indeed caused by the intervention (Angrist & Pischke, 2008; Duflo, Glennerster, & Kremer, 2008). Thus far, only the simplest of experimental designs has been considered, where units are individually assigned to conditions. This design is well-suited when the risk of spillover is low and the research question concerns individual-level behaviour such as giving to charity, signing up to become a donor, or increasing physical activity levels. If the researcher were to believe that the behaviours of interest result from a complex interplay between the individual and their environment or that treatments are taken up by others, alternative experimental designs must be used. This thesis employs individual randomisation but benefits from a closer look at spillover, as introduced in the next section.

3.3.1 Accounting for spillover effects

The potential issue of spillover is especially pertinent in education research, where students and teachers interact with one another regularly and over extended periods of time. For example, when researchers want to assess whether programmes for smoking prevention are effective at preventing smoking uptake in secondary schools (see Campbell et al., 2008), it would be ill-advised to assign some pupils within classrooms to receive the new smoking education, and other to receive the usual education. Within classrooms and schools, pupils communicate with one another, so that pupils assigned to the control group may learn about the intervention content. Smoking is also a relatively public behaviour that takes place in the school environment, so a change in behaviour will likely be noticed. Communication between subjects blurs the line between intervention and control groups, complicating the identification of causal effects considerably, and violating SUTVA (Sinclair, McConnell, & Green, 2012). Cluster-level randomisation may address this issue, as it allows researchers to avoid local interactions between units assigned to different conditions (Athey & Imbens, 2016). In the above example, the researchers assigned entire clusters, such as classrooms or whole schools, to treatment groups (Campbell et al., 2008).

Spillover was deemed a moderate risk in the two studies described in this thesis due to their explicit focus on increasing communication. It was technically possible for students to nominate one another (but strongly discouraged), and therefore some students may have been allocated to the control group (as recipient) while also receiving updates about their peers (as supporter). If these students attended the same class, they would receive informative updates about upcoming exams and course content. In this case, spillover effects threaten the unbiasedness of our causal estimation. Treated students may exhibit more positive in-class behaviours, and therefore influence their peers. Evaluating a similar text-messaging intervention where parents receive information about their child's grades, behaviours and attendance, Berlinski and colleagues (2016) constructed classes with a high (75%) or low (25%) share of treated students via random assignment. They found that the effect of being assigned to treatment is larger for students in high-treatment classrooms in comparison to low-treatment classrooms (Berlinski et al., 2016). Cunha et al. (2017) also find positive spillover effects of their texting intervention, as within-classroom control students improve in terms of attendance and GPA almost as much as their treated peers. As other studies consistently find positive spillover effects on untreated students (e.g. Bergman, 2016; Cunha et al., 2017; Xu, 2017), the likely effect of any spillover in the present study is the attenuation of treatment effects. This potential issue is further examined in chapters 4 and 5 (p. 92 and p. 184, respectively), where students who nominated each other in pairs are identified and spillover effects are further considered.

3.3.2 Partial compliance

It is helpful to make explicit the analysis choices employed in this thesis, as they have implications for the inferences that can be drawn from the treatment estimates. In this thesis, treatment consists of two parts. First, students nominate a key individual in their social network, who then receives weekly communications about the student's learning. The text messages are sent out by the experimenter, and their (un)successful delivery is observed. The second element of treatment is not observed by the experimenter, but is arguably the true active ingredient: the quantity and quality of learning-related conversations between the nominated study supporter and the student. It is hypothesised that the informative text messages prompt the supporter to enquire about the student's learning and offer more frequent reminders or support.

It may be tempting to restrict analysis to subjects in the treatment group who received the weekly communications. Some supporters were never actually treated, for example because the student provided an incorrect mobile phone number, or because the supporter opted out of receiving the texts. However, comparing average outcomes of the treated subset of the treatment group with subjects in the control group is erroneous. This analysis is no longer focused on groups created by the initial randomisation which will lead to biased inference (Duflo et al., 2008). Those who are actually treated are a non-random subset of the treatment group (Gerber & Green, 2012). For example, less motivated students might have intentionally nominated non-existing individuals as supporters. Similarly, supporters who opted out of the programme may not be as close to the student as those who do not opt out. For those in the control group, it is not known whether their supporter would opt out, or whether these supporters turn out to be uncontactable. In short, restricting analysis to 'treatment-on-the treated' (TOT) may exaggerate the effect of our social support intervention. The endogeneity of students' choice of whom to nominate influences educational success as well as whether their nominated supporter received treatment.

It is therefore established practice to analyse field RCTs using the 'intention-to-treat' (ITT) estimate. All participants assigned to treatment and control, whether they complied with the treatment or not, are included in the analysis. Interpretation of ITT is also more straight-forward than TOT, as the ITT effect is calculated as the average effect of assignment to treatment (Athey & Imbens, 2017). TOT only produces an effect for the 'compliers', or those who are treated only when assigned to the treatment group (Gerber & Green, 2012). Using ITT, researchers are able to test whether a new intervention works, on average, for the total population rather than a specific subgroup of people who accepted participation. It is therefore an unbiased estimate of the average treatment effect (Duflo et al., 2008).

With the exception of CACE estimates presented in section 5.6.2 (p.182), treatment groups are constructed on the basis of random assignment, not on actual text message receipt. The delivery and dosage of the intervention are secondary outcomes of interest, but they are not used as exclusion criteria. Additionally, it is challenging to ensure all subjects in a treatment group are treated, especially in real-world settings. If the social support intervention were to scale up, it would be unreasonable to expect all students to nominate engaged study supporters with valid phone numbers. The estimates produced by an ITT analysis carry more external validity than if we were to restrict analysis to compliers only. Both ITT and complier only (i.e. CACE) analyses are performed in the second empirical study, Chapter 5.

3.3.3 Attrition

Attrition occurs when a subject's outcomes cannot be observed at the point of data collection (Gerber & Green, 2012, p. 211). Missing outcome data is often an issue in education, as students may drop out, transfer to a different school, or relocate. Attrition is a particularly severe issue when data is not missing at random, but systematically related to the trial (Torgerson & Torgerson, 2008). For example, an intervention could cause students in the treatment arm to drop out of the participating school. If relatively few students drop out of a control arm, but a higher proportion of students drop out of the treatment arm, it could be erroneously concluded that an intervention has a beneficial effect when in fact those who were harmed by it dropped out and are no longer observed. The most favourable case therefore, is when levels of attrition are similar across both trial arms (Torgerson & Torgerson, 2008).

In this thesis, attrition is fortunately relatively low, as will be further evidenced in chapters 4 and 5 (see sections 4.4.2 (p. 103) and 5.3.2 (p. 162), respectively). The recruitment phase lasted several weeks, so that the intervention was only launched 1.5 months into the academic year. Whereas this risk the chance that those who 'need it most' would not benefit from the intervention, the risk of attrition is lower. Anecdotally, it is suggested that students are most likely to drop out of college within the first 42 days of the academic year; the so-called 'qualifying period' (Education and Skills Funding Agency, 2018), which occurred before randomisation.

3.4 Validity and reliability of field experiments

3.4.1 The narrow focus of RCT designs

A popular critique of RCTs is that they lack external validity (Asmussen, 2011), or the ability to determine "whether the study that is established in the study will be true elsewhere" (Cartwright, 2010, p.60). Like most methods, experiments are often highly localised, conducted in a restricted range of settings, with a convenience sample of subjects (Shadish, Cook & Campbell, 2002). Additionally, experiments use a pre-defined set of measures, which by definition measure some narrowly-defined and pre-specified constructs of interest (Bamberger, Tarsilla, & Hesse-Biber, 2016). Rather than assessing 'what works', the RCT assesses whether it 'worked there' in a particular setting, time, and population (Hanley et al., 2016). Finally, experimental designs often limit data collection points to pre- and post-intervention. Although recent pragmatic trials (e.g. Bonell et al., 2012) are spearheading more integrated data collection, the vast majority of RCTs do not include observation of "formal and informal implementation processes" (Bamberger et al., 2016).

Additionally, RCTs are typically limited to evaluating short-term to medium-term outcomes and are restricted to regions in a way that observational studies are not. Observational studies can follow whole countries over several decades. For example, Ravallion (2012) analyses data on a hundred developing countries to study how initial poverty is related to subsequent growth over a period of 30 years. The data was gathered using household survey data, and time periods between surveys (within countries) were up to 27 years. It would be prohibitively expensive and lengthy to implement an RCT at such a scale. I now turn to several relevant concepts that may impair the validity and reliability of field experiments, and apply these to the studies described in this thesis. The final part of this section describes how the generalisability of experiments may be bolstered.

3.4.2 The Hawthorne Effect

The external validity of RCTs may be impaired by the participants' awareness that they are taking part in an experiment. Although this is explicit in lab-based experiments, this may be true also in more natural settings. Participants may be informed of the study objectives in order to gain consent. It can therefore be expected that they may not behave in the way they would if they were not being observed (Peters et al., 2016). This idea is referred to as the 'Hawthorne Effect' and occurs when study outcomes are influenced by participants' awareness of the changes produced by the intervention. In other words, by virtue of participation, trial subjects may do better (or worse) than those in business-as-usual conditions (McCarney et al., 2007; Merrett, 2006).

In this thesis, students are informed about the aims and data collection procedure of the field experiment before they make the choice to sign up. They therefore know that the purpose of the intervention is to help improve attendance and attainment in their maths or English subject, and might change their behaviour as a result. Once students signed up they continued to attend class as usual and the study was not referred to throughout the remainder of the year. This potential hazard to external validity is further discussed in chapter 4, where qualitative data is used to explore whether students were aware of experimental aims and expectations (see Section 4.8.7, p. 127).

3.4.3 Issues of self-selection

Most RCTs rely on collaboration with practitioners in the field such as school teachers, social workers or local authority personnel, and require individuals to opt in or out of participation. This two-stage process of selection carries potential threats to external validity. First, it may be that the colleges who approached the experimenters and requested to take part in the trial are qualitatively different from colleges who did not show an interest in collaboration. Perhaps the selected colleges were more interested in increasing English and maths attainment, or struggled more with student attendance or attainment than other (non-participating) colleges. Second, students within these colleges self-selected into the trial, and participation was voluntary. Self-selection into a treatment group is dealt with effectively by the randomisation procedure, but the potential threat of self-selection into the experiment receives much less attention (Allcott, 2015).

It is often thought that self-selection results in positive selection bias, where the Average Treatment Effects (ATEs) are larger for trial participants than they would be for those who chose not to participate (List & Rasul, 2010). Belot and James (2014) propose an alternative perspective, which posits that positive and negative selection bias could take place at the same time. Depending on whether there are alternative or competing interventions available and participants' expectations of the effectiveness of the intervention, opting in to the experiment could be too costly as there is a good chance one would end up in the control group (Belot & James, 2014). It is difficult to assess the direction and size of the selection bias as this would require an understanding of individual expectations and available alternatives.

Fortunately, the magnitude of selection bias can be studied empirically by comparing the observable characteristics of those who opted in to the experiment with the broader population of interest. In Sanders and Groot (*forthcoming*), we describe a field experiment (set out in Chapter 4) where colleges provided outcome data on an opt-out basis, whereas trial participation was administered through an opt-in process. We therefore have access to outcome data for students who did and did not self-select into the treatment. Due to the nature of our college recruitment strategy, self-selection at the level of colleges cannot be assessed. We can however observe student-level self-selection.

Only a small number of observable characteristics were correlated to selection. Students on Functional Skills courses are significantly less likely to consent (8.2 % points), as well as those students who failed to report their gender (11.2 % points). Age, subject and gender were uncorrelated to selecting into the trial. It is therefore plausible that self-selection occurred on non-observable characteristics, which is less straightforward to correct for than if selection was driven by observables. We found considerable evidence of selection bias. Students who consented to take part but assigned to the control condition (N = 743) had 3.9% higher attendance than those who did not consent, which was statistically significant at the $p < .05$ level (N = 6089, Sanders & Groot, *forthcoming*, see Appendix 44, p. 289). Treated participants had 4.8 % points higher attendance than their peers in the control condition. A naïve regression where self-selection was not taken into account, over-estimated the true effect by 77% (i.e. 8.5 % points rather than 4.8 % points).

Perhaps those who feel more confident or motivated to do well in their GCSEs are also more likely to want to take part in extra-curricular interventions such as the Study Supporter programme. These findings show that self-selection is a potential threat to our ability to generalise the findings. It is possible that the study populations in Chapter 4 and 5 are qualitatively different from the policy population the interventions may eventually be scaled up to. However, we did not find evidence of participants self-selecting into the trial based on prior beliefs of the effectiveness of the treatment. The vast majority of participants cited the financial lottery as the primary reason for signing up.

3.4.4 Transparency and replicability

Replication is an essential strategy for knowledge generation (Barnow, Burt & Greenberg, 2016) and allows researchers to explore whether experimental results have external validity. The importance of replications has become a popular topic in public discourse and policy-making recently, but is by no means a modern concern. In 1969, Campbell remarked that “too many social scientists expect single experiments to settle issues once and for all” (p. 28). Recent efforts to replicate seminal experiments in the field of psychology found that the reproducibility of the original effects was low for several of the studies (Klein et al., 2014).

It has become clear that poor measurement in noisy research settings does not necessarily mean the observed effect would be even larger in ‘clean’ settings (Loken & Gelman, 2017). The authors caution against “assuming that which does not kill statistical significance makes it stronger” (Loken & Gelman, 2017, p. 584). Experiments with small sample sizes (e.g. N = 50) have low statistical power (unless the effect size they hope to uncover is very large), which can lead to overestimation of effect size, rather than underestimation (Button et al., 2013). Unreliable research, both in terms of measurement error and sample size, leads to overstated conclusions, exacerbating the replication crisis.

The replication crisis has resulted in several important changes to the way experiments are designed, implemented and reported. First, it is now common practice in the behavioural sciences to pre-register an analysis plan. It details how the researchers will collect, clean and analyse data before randomisation occurs. After collecting the final outcome data, the analyses listed in the pre-analysis plan are primary analyses and any additional tests are labelled as exploratory analyses. Pre-analysis plans limit researchers’ freedom in choosing model specifications that have the most satisfactory outcome, cutting the data in multiple ways, or cherry-picking hypotheses (Olken, 2015). The trials reported in this thesis were pre-specified both in a trial protocol (Experiment 1 and 2) and online trial databases (Open Science Framework (Chapter 4) and the American Evaluation Association (AEA) repository (Chapter 5). The pre-analysis plans focus on a narrow set of primary outcomes. Finally, the narrow focus of pre-specified analysis plans is complemented by incorporating qualitative inquiry into the study design, which is more exploratory by definition.

Experimental replications are also important tools in the field’s quest to reduce false positive results. Pre-specified analysis plans can remedy concerns about “p-hacking”, but only replications can provide us with a sense of the robustness of the findings (Coffman & Niederle, 2014). Replications can further elucidate the conditions that lead to larger or smaller effects.

When a single trial contains too few members of subgroups of interest, multiple trials can help illuminate whether the impact varies with variations in the environment or participants (Barnow et al., 2016). Unfortunately, (independent) replication studies are rare. It has, however, become more widespread to plan progressive stages of evaluation. A promising intervention can be tested in a controlled pilot setting first, and then scaled up and tested across various settings of interest (Campbell et al., 2000).

The Education Endowment Foundation (EEF) advocates this cumulative approach, where interventions are first evaluated using small-scale development studies, scaled up to efficacy trials, and if they stand up to scrutiny, to large-scale effectiveness evaluations. Through such iterative evaluation of the intervention, it is shaped from 'proof-of-concept" to a scalable policy or programme (Banerjee et al., 2017). This approach has been advocated by several scholars, recognising that sequential testing also allows researchers to deal with possible heterogeneous treatment effects (Barnow et al., 2016). The studies reported in this thesis contain elements of the multiple trials approach. The first field experiment was conducted in the 2015/16 academic year, its design was iterated in the following academic year, and is scaled up to reach 4000 students in 2017/18 and evaluated by an external evaluator.[11] When conducting multiple follow-up studies, uncertainties about the essential features or active ingredients are reduced since every experiment sheds some additional light on the boundary features and prior conceptualisations.

3.5 Overcoming the limitations of randomised controlled trials

Thus far, I have reviewed the theoretical underpinnings of the RCT and discussed approaches to ensuring high internal and external validity. RCTs enable researchers to obtain reliable and statistically unbiased estimates of the causal impacts of a program, policy or intervention, and are consequently regarded as the 'gold standard of educational research' (Hanley et al., 2016). However, as with any other method, its application can be weak. When the

[11] For more information about the scaled-up version of the intervention, which is not part of this thesis, see the Education Endowment Foundation website: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/texting-students-and-study-supporters/

RCT is not adequately designed, implemented rigorously, or analysed appropriately, the validity of the inferences is at risk. The above sections on internal and external validity illustrate the various threats to RCTs. In essence, issues such as attrition, spillover, and noncompliance demonstrate the importance of careful observation of what happens when the execution of an RCT deviates from its design. Such violations of the design can be anticipated and planned for in the pre-analysis plan and can additionally be studied during implementation (Anders et al., 2017; Jamal et al., 2015). The latter approach is further discussed below.

3.5.1 Identifying what works, for whom, and under what circumstances

Critiques of RCTs extend beyond discussions of imperfect design and implementation and how to remedy issues with statistical tools. First, several scholars have criticised the 'black box' approach of the RCT (Bonell, Fletcher, Morton, Lorenc, & Moore, 2012; Cartwright, 2009; Deaton & Cartwright, 2017; Hawe, Shiell, & Riley, 2004; Jamal et al., 2015). Simply testing whether the means of treatment and control groups are sufficiently different does not explain 'why' things work (White, 2013; White & Philips, 2012). Deaton and Cartwright (2017) argue that it is essential to understand the cultural and social environments in which trials are set, to identify why the treatment is or is not effective. Without this understanding, we cannot know what populations the findings can be generalised to (Deaton & Cartwright, 2017).

Contextual factors are often downplayed in RCT designs in favour of control over its implementation, to the degree that we are no longer evaluating a real-life situation (Hawe et al., 2004). Approaches to ensure the internal validity of the RCT, such as double-blind randomisation, may make it difficult to assess how a programme would be received in the absence of such strict rules. Realist researchers have also challenged the prevailing reliance on standardisation, proposing that intervention delivery should be allowed to take on different forms according to context (Marchal et al., 2013). When the 'active ingredients' or mechanisms of a given intervention are clearly theorised, they argue, the key program components should be delivered at the optimal dose for each intervention site (Hawe et al., 2004). To some degree, the interventions described in this thesis are tailored to the context, as every participating college received a unique text messaging schedule. The schedule is based on conversations with English and maths staff, and the college curriculum. However, the intervention dosage was kept uniform across participants. Future studies could allow even greater customisation of the intervention, for example by allowing participants to set their preferred time for delivery of the text messages, or by allowing tutors to write text messages specific to classes, rather than the entire cohort of maths or English students. It should be noted that the essential features of the intervention should be kept fixed, so that only peripheral elements vary across contexts (Abry, Hulleman, & Rimm-Kaufman; 2015). This approach requires clear theorisation of intervention components, which I turn to next.

Critics of field experiments may argue that interventions are too often designed without a clear theory in mind. It may be tempting to conceptualise interventions as "collections of resources, equipment and personnel" (Pawson et al., 2004, p.4), and think in terms of 'when the intervention X is switched on, cause Y follows'. Paying no attention to how the underlying mechanism M connects X and Y may not be harmful to one's ability to test whether the treatment is effective, but it does lead to ambiguity in attempts to explain the results (Morris, Edovald, Lloyd, & Kiss, 2016; Pawson et al., 2004). Field experiments can be strengthened by including a theory phase before implementation, where theoretical mechanisms of change are used to design the intervention (Michie, Johnston, Francis, Hardeman, & Eccles, 2008). During the implementation phase, the feasibility and fidelity of the intervention should be assessed using implementation and process evaluation (IPE; De Silva et al., 2014). Finally, further refinement of the theory will help build a strong evidence base.

3.5.2 Realist approaches to evaluation

In recent years, the realist approach has attempted to build a bridge between the RCT methodology and the exploration of adaptation to context (Pawson, 2004; Pawson et al., 2004). Its core principle is that the theoretical underpinnings of an intervention should be made explicit, and refined through iterative and systematic evidence collection (Pawson et al., 2004). Realist evaluations explicitly capture contextual factors that affect the implementation and outcomes of the RCT, and often draw on qualitative data to gather such insights (Bonell et al., 2012; Moore et al., 2015; Oakley, Strange, Bonell, Allen, & Stephenson, 2006). By doing so, they attempt to integrate the more detailed focus on contextual factors into randomised controlled trial designs.

The more zealous realists argue that the experimental method is not well-suited to answer questions such as “what questions are worth asking?” or “what value should be attached to the experimental findings?” (Shadish, Cook & Campbell, 2002, p. 9). More recent ‘realist RCTs’ bridge this apparent chasm by deploying qualitative approaches alongside - or ahead of – RCTs to ensure these important questions are addressed. In summary, the critique of the RCT as a context-free research endeavour can be overcome by moving from the simple RCT to mixed-methods intervention evaluations (Anders et al., 2017; Drabble & O’Cathain, 2015).

Although these critiques by realist researchers have helped stimulate the important discussion about limitations of RCTs, they do at times exaggerate the limitations of RCTs. Some realist researchers argue that realist RCTs as presented by Bonell et al. (2012) or Jamal et al. (2015) fail to sufficiently take into account the “dynamic interplay among the intervention, actors, context, mechanisms and outcomes” (Van Belle et al., 2016, p.1). Some critics go as far as to say that RCTs cannot be usefully applied to complex interventions (Marchal et al., 2013). Sanderson (2000) states that “approaches founded upon [...] linearity in the relationship between variables, and of proportionality of change in response to causal influences – such are not appropriate in seeking to understand social systems that exhibit complexity” (p. 442). This critique is somewhat unhelpful. If researchers incrementally study the interaction between intervention elements and pay attention to the context, this is arguably more beneficial than not doing so at all. In this thesis, I take the view that a mixed use of quantitative and qualitative methods will help explore programme theory and context, while retaining the ability to make causal inferences due to random allocation of treatment.

In summary, there appears to be much value in the combination of qualitative evaluation approaches and the RCT methodology. The combination of both approaches requires methodological flexibility and integration. The various ways in which quantitative and qualitative data can be synthesised is discussed next.

3.6 Mixing quantitative and qualitative methods

The previous section illustrates the importance of contextual analysis and theory-driven interventions but did not introduce evaluation designs that integrate quantitative and qualitative methods. In this section, the value of qualitative inquiry is explored, and practical approaches to mixing both methods are introduced. I argue that qualitative data is key to understanding how an intervention was implemented in actuality and how it was experienced by its recipients.

Qualitative research within or alongside randomised controlled trials (RCTs) can take many shapes. The majority of embedded qualitative studies in intervention trials focus on just one aspect of the trial, such as improving recruitment practices (e.g. Donovan et al., 2002), gaining an in-depth understanding of participant experiences (Whittemore, Rankin, Callahan, Leder, & Carroll, 2000), or charting non-compliance (Campbell et al., 2000). Yet, by focussing on only one implementation dimension researchers can miss factors relevant to implementation and theory. Three primary qualitative aims support the quantitative components of this thesis, each focussing on a different facet of the intervention: theory, design, and implementation.

It should be noted that the integration of qualitative evidence into field experiments is typically one-directional: it is used to facilitate the interpretation of trial results (Drabble & O'Cathain, 2015). The field experiments described in this thesis follow this principle, and therefore this chapter concerns designs that prioritise quantitative data collection (see Creswell & Plano-Clark, 2011; Creswell, 2009, 2013 for detailed summaries of designs where qualitative data is the dominant mode of data collection).

3.6.1 A mixed methods movement

The introduction of 'mixed methods' (MM) in the 1980's arose from a growing frustration with the dominant discourse on the dichotomy between quantitative and qualitative research (Creswell & Plano-Clark, 2011). The 'paradigm war' between the two approaches created an either-or-dilemma (Tashakkori, Teddlie, & Johnson, 2015). Although there are exceptions, most qualitative paradigms such as constructivism or interpretivism posit that knowledge is socially constructed, and is therefore subjective (Creswell, 2013). The underlying worldview of quantitative approaches is better summarised by the belief that there is truth and that objective knowledge about the world can be gained through careful measurement (Creswell, 2013). The philosophical foundation of mixed methodology, unsurprisingly, is pragmatic. The pragmatic worldview is not committed to either paradigm or approach to analysis: pragmatists apply all approaches in order to answer their research questions (Johnson & Onwuegbuzie, 2004). Throughout this PhD thesis, I ascribe to the pragmatic paradigm.

Two aspects are important in any mixed methods research project: the timing and weighting of quantitative and qualitative data. Qualitative inquiry is embedded within the RCT in this thesis. Both types of data collection can occur alongside one another or can be deployed sequentially. This thesis applies a concurrent approach, where qualitative data is collected during the trial. Its strategy therefore most closely corresponds to the concurrent embedded design (Creswell, 2009). In this approach, both methods address different questions. The quantitative data is used to make causal inferences, whereas the qualitative inquiry is used to enrich the description of context, participant experiences, and implementation fidelity (Palinkas et al., 2011).

This approach takes advantage of the strengths of both approaches. The mixed methods field experiment uses administrative data on class attendance and attainment to isolate treatment effects, and qualitative data supports a more nuanced understanding of the processes that underlie the intervention. The qualitative element also allows us to study outcomes that are not easily quantified. In summary, mixed methods are used to compensate for the weakness of using just one set of methods. The triangulation of qualitative and quantitative data enhances the credibility of field experiments, and allows the RCT to become more transparent and accountable (Hesse-Biber, 2012, 2013).

3.6.2 Challenges inherent to mixing methods

It should be noted that mixing methods can be challenging. Qualitative fieldwork can be both time- and resource intensive, and is therefore constrained to a smaller sample size. As most researchers are primarily educated in either quantitative or qualitative methods, the mixed methods elements are at times combined in an ad hoc manner (Bamberger et al., 2016). Lewin and colleagues sampled one hundred RCTs to assess the use of qualitative inquiry alongside trials within healthcare (Lewin, Glenton, & Oxman, 2009). From the 100 trials, 30 cases included qualitative research. The methodological quality of these studies varied; several did not describe their qualitative sampling approach or failed to clarify their approach to analyse data (e.g. thematic or content analysis). Mixed methods field experiments can suffer from researcher bias when the trial manager is also the qualitative researcher, or selection bias when sampling strategies are not clearly theorised (Drabble & O'Cathain, 2015).

3.6.3 The contribution of interviews in field experiments

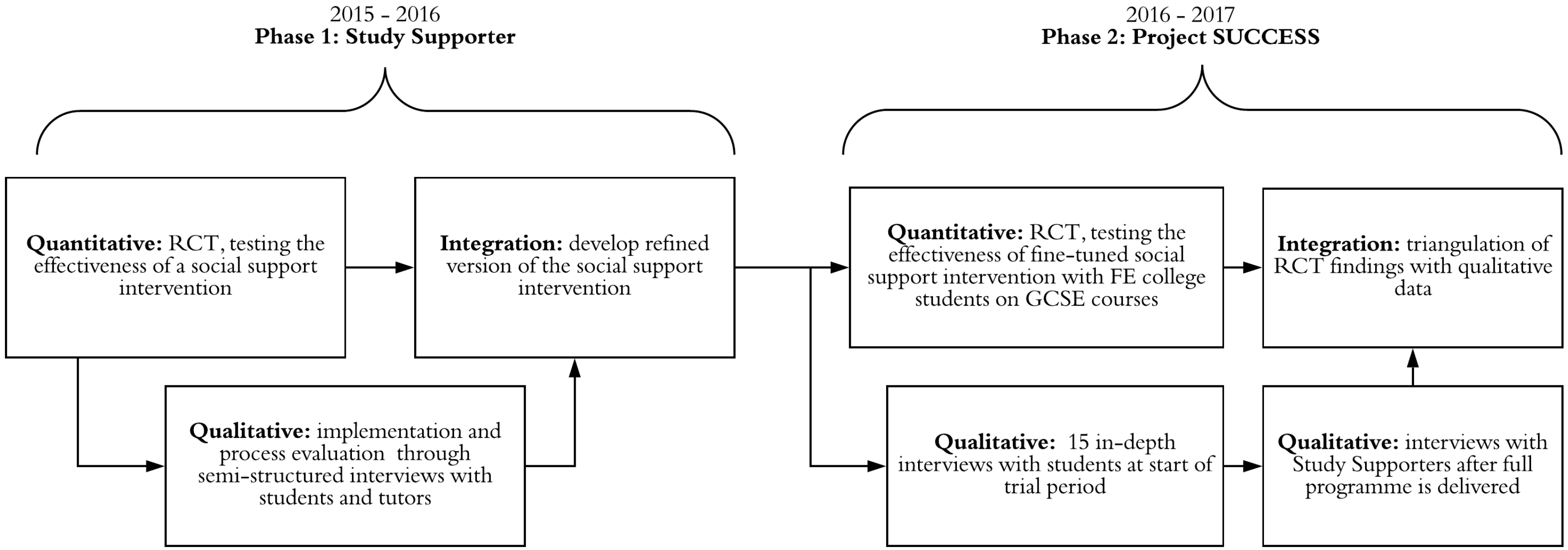
Qualitative research methodologies can be broadly categorised into two categories: they either rely on naturally occurring data, or involve generating data (Lewis & Ritchie, 2003). Ethnography or participant observation places the researcher within the context of interest, and natural interaction between the researcher and participants is of interest. In-depth interviews or focus groups, on the other hand, "give participants a direct and explicit opportunity to convey their own meanings and interpretations" (Lewis & Ritchie, 2003, p. 57). Qualitative research embedded within quantitative evaluations typically takes the form of interviews or focus groups. Generating data allows for probing and clarification, is much less resource intensive, and is more standardised in nature. This thesis employs semi-structured interviews to explore how students seek and receive social support, and how they interacted with their study supporters and the text message content.

If it were not feasible to conduct in-depth interviews, focus groups with groups of students or study supporters would be a reasonable alternative. Focus groups are shaped by group interactions and are especially valuable when the phenomenon of interest is of a social, rather than personal nature (Lewis & Ritchie, 2003). For example, bullying inside the classroom is a topic that is shaped by multiple players. A focus group may be able to explore the social norms surrounding bullying behaviour, or how group composition influences the discussion. Although the social nature of the focus group is its primary strength, it can also induce social desirability bias, conformity or 'groupthink' (Boateng, 2012). The primary reason for selecting the interview as data collection method, therefore, is that the detailed personal focus allows deep exploration of participants' personal contexts and reflections. Second, the semi-structured approach simplifies comparison across cases, but also allows for further clarification and follow-up.

3.6.4 Embedding qualitative inquiry

Qualitative inquiry is wrapped around the two field experiments in order to provide a richer account of students' experience of learning at an FE college and how they and their nominated supporter interact with the text messaging intervention. Particularly since the intervention is mediated by human behaviour - in the sense that nominated supporters need to engage with the text messages for the intervention to ultimately affect student outcomes - qualitative methods help facilitate the interpretation of trial results. Thematic analysis is used to capture depth of understanding of participants' lived experiences of education, supportive relationships, and the intervention. In order to gain a rich and triangulated description of the themes of interest, the different viewpoints of students, their supporters, and tutors are considered. A graphical representation of the flow between quantitative and qualitative components of this thesis is displayed in Figure 1.

Figure 1: A graphical representation of quantitative and qualitative components



3.6.5 Qualitative research questions

Few interventions have been implemented in the Further Education college context (Anderson et al., 2001; Dalby & Noyes, 2015; Swan, 2006), and none using an experimental approach. It was thus deemed important to chart the implementation context in colleges and assess feasibility and acceptability of programme elements. The interview schedule was developed to gather detailed information on the most important factors for successful implementation. Finally, the qualitative inquiry deployed alongside the first field experiments aimed to develop and refine the intervention for the second trial.

The qualitative component of the second trial (Chapter 5) explores barriers and facilitators of the intervention. Qualitative data is also collected to provide a rich description of the participants and further education college environment, and students' lived experiences of the GCSE resit policy.

The final aim was to gain a deeper understanding of the potential mechanisms of the two interventions set out in this thesis. Chapter 6 is entirely focused on the qualitative data gathered during both trials. It explores how the intervention is enacted by participants, and what meanings students attach to interactions with their study supporter(s).

To answer these questions, students, tutors and study supporters were purposefully sampled. The vast majority of qualitative studies rely on purposive sampling, including many mixed-methods implementation studies (Palinkas et al., 2015). Students assigned to treatment groups in which study supporters received text messages were prioritised since the interaction between students and study supporter was of primary interest. This choice explains why the qualitative component does not seek to explain average or heterogeneous treatment effects: few students assigned to control and 'student only' arms were interviewed because these were not deemed the "information-rich cases" (Palinkas et al., 2015, p. 533) within the context of student-supporter communication as the focal phenomenon of interest. Qualitative research is time-consuming, and interviewing individuals who are not especially knowledgeable of the phenomenon of interest may therefore not be the best use of resources.

3.6.6 The qualitative analysis sequence

A thematic analysis approach is used to analyse the interview data, as this approach allows for an in-depth exploration of respondents' views, motivations and experiences through a systematic coding process and identification of themes and patterns (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2008).

Existing research informed an initial coding framework, which was then applied to transcripts (Potter & Levine-Donnerstein, 1999). This approach is not overly restrictive, as text passages that cannot be coded with the existing coding scheme are added to the scheme (Hsieh & Shannon, 2005). They may either represent a new subcategory of an existing code, or refine *a priori* codes (Forman & Damschroder, 2007). That way, coding is performed both deductively and inductively, allowing for a theory-relevant yet also data-led approach to analysis.

The inductive element of the analysis process is important – few studies have directly examined the way students seek and receive social support in maths and English within the mandatory resit context. After coding of all transcripts is completed, the resulting coding framework is rearranged into categories or 'themes'. This data reduction process ensures that only the most relevant text passages are applied to address the research questions. Since the qualitative research questions addressed in Chapter 4, 5 and 6 all require subtly different approaches to the collection and analysis of qualitative data, the specific analysis sequences are set out in more detail in each chapter.

Finally, qualitative research is different from quantitative research in its primary purpose. The goal of qualitative inquiry is not to make generalisations from a sample to the general population using statistical inference, but rather to understand a phenomenon of interest in-depth (Forman & Damschroder, 2007). The qualitative component relies on narrative rather than frequencies or probabilities. Trustworthiness of the qualitative findings is established through detailed record-keeping, researcher reflectivity, the construction of a clear coding framework, and detailed note-keeping as themes emerge. Approaches to ensure the validity and reliability of qualitative research are addressed in the next section.

3.6.7 Attaining rigour in mixed methods research

Qualitative data analysis is an inherently subjective process since the researcher is the primary instrument of the research (Forman & Damschroder, 2007). It is therefore important to critically reflect on the measures taken to ensure the interpretation of qualitative data is credible. This section introduces constructs used in qualitative research to demonstrate that the phenomenon of interest is portrayed truthfully. It requires "confronting your own assumptions and recognizing the extent to which your thoughts, actions and decisions shape how you research and what you see" (Mason, 2002, p. 5). The validation strategies for the qualitative component of the mixed methods randomised controlled trial are informed by the guidelines provided by Giddings and Grant (2009) and Plano Clark and colleagues (2013). Well-known guidelines for the trustworthiness of qualitative research are put aside (e.g. credibility, transferability, dependability and confirmability; Lincoln & Guba, 1985) because they focus on purely interpretative qualitative research. The below validation strategies fit within an embedded qualitative study approach and the pragmatic paradigm.

Clear articulation of research questions is the first step. The qualitative research questions should support the quantitative research questions clearly, for example by enabling exploration of the subtleties in the intervention process that are difficult to detect statistically, enhance the communicability of trial findings, or further description of within- and between-subject variation (Sandelowski, 1996). The overarching qualitative research questions were introduced in section 3.6.5 (p. 72). In sum, the qualitative data support the integrity of the thesis through (1) further refinement of the intervention, (2) exploration of barriers and facilitators of intervention effectiveness, and (3) generation of hypotheses for further investigation of intervention mechanisms.

Second, triangulation is incorporated into the thesis through the collection of in-depth interview data across both experimental studies. Interviews were conducted with various stakeholders: student participants, tutors, and study supporters. Triangulation of the qualitative and quantitative data is further supported through complementing interview data with survey data collected during the consent procedure (Chapter 6, section 6.3.2, p. 219).

Third, auditability is ensured through clear description of the data collection and analysis process. An audit trail through detailed record-keeping of the interview transcripts and analytic memos aids confirmability of trustworthiness (Hsieh & Shannon, 2005).

The relevance of the study is established through detailed description. Other qualitative researchers should be able to assess the degree of transferability (generalisability, in quantitative terms) between the context of the two studies in this thesis and theirs. Therefore, the qualitative analysis sections of Chapters 4 and 5 both start with a detailed description of the further education college environment. Interview participants were also sampled using a stratified sampling approach to ensure representation from a range of settings.

A number of validation strategies were not feasible within the context of this thesis. Since the research was carried out by a single researcher, frequent 'expert critique' where another researcher who is intimately familiar with the study provides feedback on coding decisions was unfeasible. Inter-coder agreement requires the agreement of two or more independent coders, which was unviable for the reasons set out above. Therefore, I reviewed the initial coding scheme one month later to assess whether discrepancies arose between T1 and T2, and continually checked codes and themes against the raw data. The initial coding scheme was revised slightly upon the second review (T2), to reflect addition of more granular codes around typologies of social support. All interview transcripts were revisited and recounted episodes of social support were re-coded to reflect the greater detail of codes. The final coding scheme is presented in Chapter 6, Table 6.3 (p. 222).

The quantitative and qualitative analysis were not carried out by independent researchers, as is often the case in mixed methods randomised controlled trials (e.g. Plano Clark et al., 2013). Researchers who deliver both the intervention and subsequently conduct the interviews may be more likely to find evidence that confirms their expectations. To minimise the potential for bias, the qualitative data collection was carried out prior to collection of quantitative outcome measures. Whether the intervention was successful at improving student outcomes was not yet known at the time of qualitative data collection to minimise the potential for leading questions or bias in the analysis process.

3.7 Conclusion

This chapter stresses the importance of randomisation for causal inference. The treatment effect of an intervention can be isolated using a randomised controlled trial as the methodology eliminates the potential for selection bias. The chapter also introduced the core concept of potential outcomes as a helpful approach to thinking about causal inference (West & Thoemmes, 2010) This chapter addressed potential threats to the validity of causal inferences. It was shown that several threats to internal validity are particularly important in this thesis, including self-selection, non-interference (SUTVA), treatment nonadherence and differential attrition. The literature shows that methodological vulnerabilities are abundant and multifaceted, especially when the experiment is conducted in a real-world situation within complex social and cultural systems. Where possible, potential violations of assumptions and threats to internal validity are identified before trial design and then built into the research design. Specific issues, for example on treatment non-compliance, are addressed in the empirical sections of Chapters 4 and 5.

Additionally, in order to capture the richness of the intervention context and potential mechanisms, this chapter advocates for the integration of quantitative and qualitative methods. Combining field experiments with qualitative inquiry contributes to the validity and generalisability of causal explanations and enables theory development and refinement. An intervention is more likely to be effective when it targets real barriers to behaviour change, which requires an in-depth and theory-based understanding of causal determinants. Mixed methods approaches also shed further light on the theory across settings, contexts, and populations (Michie et al., 2008).

This thesis utilises a mixed methods field experimental design. Qualitative methods are weaved into Chapters 4 and 5 to enhance some of the limitations inherent to randomised controlled trials, and both sets of qualitative data are combined in Chapter 6, in order to:

1. Refine the intervention between the Study Supporter trial (chapter 4) and the Project Success trial (chapter 5), including improved strategies for participant recruitment and text message content development;

2. Explore recipients' responses to the intervention, in order to understand what qualities of interpersonal relationships enable the intervention to reach its full potential;
3. Examine the appropriateness of competing underlying theories identified in the literature review (Chapter 2), to investigate whether social support theories or the commitment device literature fit best with the empirical findings.

The first objective is practical in nature, as the opt-in design of the field experiment necessitates a clear recruitment strategy. The second objective sheds further light on the lived experience of participants during the trial. The third and final objective enables theory-refinement.

This thesis addresses two research questions, accompanied by specific hypotheses, as set out in Table 3.1.

Table 3.1: Research questions and hypotheses

Research question	*Hypotheses*	*Addressed in:*
RQ1: Can supportive text messages improve students' attendance and attainment?	1. Supportive and personalised text messages sent to students' study supporters will generate positive average treatment effects on attendance and course achievement. 2. The programme will work more effectively for students who(se supporter) receive the full schedule of texts.	1. Chapter 4 and 5 2. Chapter 5
RQ2: Are the effects greater if students also receive the text messages?	3. Text messages sent to supportive others in students' social networks will generate larger average treatment effects than texts sent to the student alone. 4. A more intensive communication approach, involving both the student and their nominated supporter, will generate larger average treatment effects than sending texts to only one party.	5. Chapter 5 6. Chapter 5

4 FIELD EXPERIMENT 1 - STUDY SUPPORTER[12]

4.1 Introduction

This chapter provides empirical support for the hypothesis that encouragement from friends and family can positively impact students' class attendance and exam performance. This chapter describes an intervention that is designed to empower students' key relationships to become more involved in their learning. The programme of supportive messages is delivered via text message, a popular tool in the delivery of behavioural interventions due to its low cost, adaptability and the ubiquity of mobile phones (Miller et al., 2017).

In many respects, it is therefore similar to other information interventions in education (e.g. Bergman & Rogers, 2017; Kraft & Rogers, 2015). These interventions, discussed in detail in Chapter 2 (see Table 2.1, p. 40), inform parents about their child's behaviour, such as attendance or homework completion. Parents may otherwise not be aware of the occurrence of absences, low grades or missed assignments. These messages therefore primarily serve to overcome a parent-child communication gap.

The aim of the current intervention is distinctive: the text messages are meant to encourage friends and family to support rather than monitor, to enquire rather than demand. In contrast to most parent texting interventions, the present intervention never shares information about grades, in-class behaviour or class-attendance. Additionally, the messages are not specific to individual students. Tutors at participating colleges helped write the messages for cohorts of students on maths or English courses, referring to upcoming class topics, encouraging planning for future tests, and inspiring students to reflect on their progress. The text messages are always

[12] This intervention was conducted as part of the Behavioural Research Centre for Adult Skills and Knowledge (ASK) in collaboration with Professor Todd Rogers (Harvard Kennedy School), and was funded by the Department for Innovation, Business and Skills. Todd Rogers and his team at the Student Social Support Lab were involved in developing the intervention. I helped develop the intervention, led on trial design and implementation, conducted the analyses presented here, and wrote this chapter in its entirety.

forward-looking and positively worded, encouraging influential third parties such as family or friends to be curious and build a habit of providing light-touch but frequent social support.

The intervention is also distinct from other communication interventions for allowing students to choose who should receive the communications about their learning. To date, all personalised communication interventions within educational settings have focused on the parent-child relationship. Recognising that post-16 students may no longer rely on their parents for support, the intervention gives autonomy to students by allowing them to choose anyone who would take an interest in their learning.

I now turn to a more detailed description of the target population, before describing the experimental design, data collection and quantitative analysis plan. The subsequent section provides the primary and secondary analyses. The final section sets out the qualitative research questions, data collection, analysis strategy, and results.

4.2 Experimental setting

The target population was the cohort of maths and English students at Further Education (FE) colleges in England. Both 16-18 year-olds and 19+ year old students are included in this study. This section briefly sets out age-related differences within further education in relation to academic achievement and attitudes towards education. These characteristics are of interest throughout this chapter, and specifically in the section on heterogeneous treatment effects.

Many of the young 16-18 year-olds who previously failed their qualifications, and subsequently enrolled on a programme of study at an FE college, arrive with low motivation and a negative perception of education (Anderson & Peart, 2016). GCSE resit students “vigorously expressed they felt demotivated” (Anderson & Peart, 2016, p. 202) in an in-depth qualitative study. They believed this sense of demotivation led them to be absent more often, and feel withdrawn or uninterested in lessons. Knowing the resit year is ‘a second chance’ did not appear to bolster motivation (Anderson & Peart, 2016). Therefore, learner engagement appears to be particularly challenging, and a pertinent issue as students are required to resit year after year, until they achieve a grade C or turn 18 (Education and Skills Funding Agency, 2018).

Adult learners (i.e. aged 19 or above), on the other hand, are not required to take GCSE subjects as a condition of funding, and report rather different reasons for attending literacy and numeracy classes at FE colleges (Department for Business, Innovation and Skills, 2013; Swain, 2005). Recurring adult learners' motives to attend numeracy classes include (1) proving they have the ability to succeed, (2) being able to assist their children with school work, and (3) to achieve mastery and feel enjoyment (Swain, 2005). In a survey of 4000 adult learners in England, motives for returning to education were personal and intrinsic. Economic benefits such as obtaining a better job or promotion were listed as primary outcomes, as well as self-confidence, self-esteem, and life satisfaction (Department for Business, Innovation and Skills, 2013). A meta-analysis of 38 international adult education studies finds a similar pattern to the one described here, suggesting that motivation to learn is strongly and positively correlated with age (Gegenfurtner & Vauras, 2012).

The distinction between 16-18 year olds and adult learners is also visible in the national post-16 qualification achievement rate statistics. Table 4.1 displays descriptive statistics of achievement in maths and English qualifications in further education, segmented by subject and age. A number of statistics are of interest. First, the achievement rate in GCSE courses lies considerably lower for 16-18 year olds (18.7% across maths and English) in comparison to adult learners (41.1% across maths and English; see column 1). This stark divergence is absent in the data on all maths and English further education qualifications, which is a measure that aggregates GCSE, Functional Skills and English for Speakers of Other Languages (ESOL) qualifications (see column 2). On average, 55.9 per cent of 16-18 year olds achieve any maths or English qualification, and 45.7 per cent of 19+ learners do so. These patterns are likely a reflection of different student populations. Adult learners taking GCSE courses do so voluntarily, whereas 16-18 year olds are required to do so. Adult learners on other qualifications (e.g. Functional Skills, ESOL) are often speakers of other languages, whereas 16-18 year olds on such courses are taking these courses as a stepping stone towards the GCSEs (i.e. they achieved lower than a D grade on their GCSE at age sixteen).

Table 4.1: National statistics on maths and English achievement at further education colleges

		(1)		(2)	
Subject	Age	GCSE qualifications		All maths/English qualifications	
		N	*Achievement A* - C (%)*	*N*	*Achievement Pass (%)*
Maths	All ages	115,000	23.4	884,400	49.1
	16- 18	86,310	17.6	333,900	54.9
	19+	28,690	41.0	550,500	45.5
English	All ages	131,900	23.8	885,500	49.7
	16- 18	106,770	19.8	310,800	56.9
	19+	25,130	41.2	575,700	45.8

Notes: The achievement rates reported in the final column are calculated using the Further education and skills geography data tool published by the Department for Education ("FE and skills participation: all ages demographic summary 2015/16"), by dividing the total number of students who started a learning aim (participation) by the total number of students who achieved their learning aim (achievement). The data is not segmented by qualification type, and includes GCSEs, Functional Skills, Adult Basic Skills Certificates, Credit Framework Certificates, and Awards in English and maths. Tool last accessed on 05/02/18, retrieved from https://www.gov.uk/government/statistical-data-sets/fe-data-library-further-education-and-skills

The issue is not necessarily that the cohort of 16-18 year-olds are not motivated to learn or progress. Eighty per cent of young people who failed English or maths go on to study at a further education college to study a vocational subject alongside English and maths. The vast majority (90.1 per cent) of FE college students are retained until the end of their vocational study programme and 79.1% of these learning aims resulted in a qualification (Education and Skills Funding Agency, 2017). However, they often struggle to do well in the English and maths qualifications they have been mandated to retake.

4.2.1 Motivation for a focus on class attendance

Attendance data is not routinely collected from post-16 institutions, so national comparative statistics are not available. Having collected attendance data from 33 colleges, Ofsted (2013) reports that low attendance and punctuality is an issue across all sampled colleges, and that attendance declines over the course of the academic year. Unfortunately, Ofsted was only able to report anecdotal evidence that there is a link between attendance and attainment (2013). As recent empirical evidence from further education is lacking, higher education data may be able to provide relevant evidence on the importance of class attendance for academic achievement. Newman-Ford and colleagues (2008) examine the relationship between attendance and attainment within a higher education context, and find that the more students attend class, the less likely they are to fail assignments and the more likely they are to attain good grades. Similarly, class attendance is a better predictor of grades in a class and overall grade point average (GPA) at college than students' study habits and scores on standardised tests (Credé, Roch & Kieszczynka, 2010).

National measures of school attendance rates also emphasise the importance of class attendance. Using a longitudinal dataset of student attendance and attainment between Key Stage 2 (KS2, 11-year-olds) and Key Stage 4 (KS4, 16-year-olds), the Department for Education (2016a) finds that students with no absence are 2.8 times more likely to achieve five good passes for their GCSE exams than their peers who miss between 15 – 20% of KS4 classes. After controlling for other factors, such as having special educational needs or being eligible for free school meals (i.e. a proxy for being disadvantaged), absence was still significantly and negatively related to attainment (Department for Education, 2016a).

The above correlational but large-scale studies have shown that attendance is predictive of academic achievement. It is challenging to test in a field experiment if better attendance results in improved achievement, as this would require the researcher to manipulate attendance by randomly assigning students to attend or skip class. In the absence of experimental conditions, those who attend all classes may have stronger intrinsic motivation to learn or may face fewer practical obstacles than those who miss a significant number of classes. A recent randomised evaluation of a student mentoring program shows that these barriers to attendance can be targeted and broken down by personalised interventions (Guryan et al., 2016). Helping those who miss class regularly to attend class more often is a primary aim of the intervention discussed in this chapter.

At a practical level, attendance is an important and relevant outcome measure because it can be routinely and precisely measured. A wealth of information on student behaviour is available from college administrative datasets. The use of administrative datasets has recently become more popular in impact evaluation, as it allows the researcher to gather information about real-life behaviours of large numbers of people (Figlio, Karbownik & Salvanes, 2015), without overburdening support staff at participating organisations.

4.2.2 Motivation for a focus on attainment in maths and English

The primary outcome measure remains academic attainment, as measured by students' final grade on high-stakes maths and English exams. There is good evidence that the economic and social returns to achieving one's basic maths and English qualifications are significant (Machin, McNally, & Ruiz-Valenzuela, 2018; McIntosh & Vignoles, 2001).

A recent study in the UK found that attaining GCSEs at 16 is a threshold for subsequent progress both in further education and the labour market (Machin et al., 2018). Achieving a good pass in maths and English is a prerequisite for most higher-level courses at further or higher education institutions. Machin and colleagues (2018) use a regression discontinuity design to examine whether *just* failing has a disproportionate cost in comparison to *just* passing the threshold. This study concerns students at the end of Key Stage 4, at the age of 16. The authors find that students who narrowly fail to achieve their GCSE at age 16 are more likely to drop out of education before they turn 18, by approximately 2.3 to 3.8 percentage points compared to students who narrowly achieved a passing grade. Students at either side of the threshold were only a few points apart, yet had different educational trajectories (Machin et al., 2018).

The conclusion that narrowly failing to obtain GCSEs is detrimental to young people's educational achievement (Machin et al., 2018) does not transfer to post-compulsory maths and English outcomes directly, as educational trajectories are relatively unrestricted at age 16 and less so once students embark on vocational qualifications. Even so, the returns to achieving one's literacy or numeracy qualifications at post-16 are pronounced. The Department for Business, Innovation and Skills (BIS) published a major study of returns to further education, using FE learner information, benefits information (Department for Work and Pensions), and employment data (HM Revenue & Customs). The researchers compared individuals who obtained English and maths (Level 2) alongside their vocational qualification to individuals who also obtained a vocational qualification at further education but without maths or English at Level 2. The wage premium of achieving maths and English at post-16 was estimated at 3.5% - 5% (Department for Business Innovation and Skills, 2015). Additionally, the Net Present Value (NPV) of maths and English qualifications delivered at further education colleges is estimated at £17 for every pound of government funding (Department for Business Innovation and Skills, 2015).

4.2.3 Motivation for the use of text messaging

Though perhaps not the latest trend, text messages are a popular mode of communication. Last year, 94 per cent of people in the UK personally owned a mobile phone (Ofcom, 2017), making text-messaging technology an inexpensive and scalable method to motivate, inform, and remind individuals at key moments. A systematic review of text-messaging interventions in health (addressing physical activity, weight loss, smoking and medication adherence) found that the majority of the reviewed interventions were effective (Hall, Cole-Lewis & Bernhardt, 2015). However, the authors warned that longer-term text-messaging intervention effects have been rarely studied. Similarly, no consensus exists about the effectiveness of individual components of the interventions (e.g. personalisation versus texts as timely reminders).

The educational text messaging interventions reviewed in Chapter 2 vary considerably in design and implementation. Some text-messaging programs ran for only four weeks (Kraft & Rogers, 2015) whereas others ran for up to eight months (York, Loeb, & Doss, 2018). All interventions relied on a degree of personalisation. Most communications included the child's name (e.g. Rogers & Feller, 2018), and some provided fine-grained information about the child's missed attendances or their punctuality (Bergman & Chan, 2017).

Although scholars agree that text messages can effectively be used to deliver behaviour change interventions (Fjeldsoe, Marshall, & Miller, 2009; Hall & Cole-Lewis, 2015), there is limited evidence on the optimal design of text-messaging interventions. A few recent studies have begun to untangle the effects of such variations in design and delivery. Cunha et al. (2017) varied the frequency and interactivity of the messages they sent to parents, and found that high frequency and interactivity were associated with larger effects. This study used qualitative inquiry to further explore optimal design from the students' perspective, adding to the academic discussion on how to develop acceptable and evidence-based text messaging programmes (see for example Ranney et al., 2018).

4.3 Experimental design

4.3.1 Sample representativeness

The field experiment described in this chapter was carried out across nine further education colleges in England during the 2015-16 academic year. The sample colleges varied in student population size, Ofsted rating, geographic area and achievement rates, as shown in Table 4.2. College recruitment was opportunistic rather than pre-determined; colleges approached the research team about participation in the study. Issues of self-selection associated with this approach are discussed elsewhere, section 3.4.3 (p. 60).

Two metrics of achievement are included below. First, qualification achievement rates (QARs) are provided for GCSE achievement. The achievement rates vary considerable between colleges in our sample. A second metric, that of progress scores, was introduced recently and is the government's attempt to articulate the rate of improvement between grades at the end of key stage 4 (age 16), and grades at the end of 16-18 study. As can be seen in Table 4.2 all colleges in the sample have a negative progress score, indicating that, on average, students lowered their point scores between KS4 and completing further education. Few colleges in England have positive progress scores,[13] underlining the challenge of requiring young people to re-take qualifications they are unlikely to obtain by the time they turn 18 years of age. Finally, average class attendance in maths and English also varies considerably between colleges, see the final column of Table 4.2.

The colleges are slightly larger in terms of cohort size than the national average. As colleges with larger cohorts were prioritised during recruitment (i.e. at least 250 students on maths/English courses), this is not surprising. The average progress scores in the sample are similar to national averages. The colleges represent a spread of geographic regions, although this is descriptive rather than purposively sampled.

[13] The government's interactive dataset on school and college progress scores can be viewed at https://www.compare-school-performance.service.gov.uk/

Table 4.2: Descriptive statistics for sample colleges

College	Students (all ages)		A*-C (%)		Progress score		Ofsted rating	Region	Attendance
	Maths	*English*	*Maths*	*English*	*Maths*	*English*			*M (SD)*
A	570	570	40.9	43.2	-0.33	-0.17	Good	South East	44.3 (20.7)
B	350	160	38.0	8.1	-0.36	-0.53	Good	Greater London	54.7 (38.9)
C	340	390	14.8	30.3	0.06	0.03	Requires improvement	East	41.0 (41.7)
D	380	440	31.9	32.1	-0.15	-0.30	Outstanding	Greater London	74.1 (34.9)
E	780	1070	20.1	21.0	-0.39	-0.41	Requires improvement	East Midlands	48.8 (43.5)
F	180	280	15.3	23.8	-0.30	-0.20	Requires improvement	North West	59.7 (36.4)
G	1370	1500	20.8	19.5	-0.44	-0.43	Requires improvement	Yorkshire	65.4 (35.8)
H	1060	1180	20.6	17.5	-0.22	-0.17	Outstanding	Greater London	51.8 (27.5)
I	590	750	20.0	27.1	-0.41	-0.44	Good	East	62.9 (24.9)
Sample avg.	624	704	27.4	24.7	-0.28	-0.29	-	-	57.8 (36.9)
England avg.	558	640	23.4	23.8	-0.29	-0.27	-	-	-

Notes: GCSE achievement rate statistics and student numbers for 2015/16 are obtained via the National Achievement Rate Tables published by the Department for Education, https://www.gov.uk/government/statistics/national-achievement-rates-tables-2015-to-2016 on February 5th, 2018. Progress scores are calculated by comparing prior attainment scores (at the end of KS4) against grades at the end of 16-18 studies. All prior attainment and progress scores only include students who are at the end of their 16-18 studies. Reported progress scores are from 2015/16. retrieved from https://www.compare-school-performance.service.gov.uk/ on February 5th, 2018.

4.3.2 Ethical approval

Ethical approval was gained from Harvard's Institutional Review Board on September 25th, 2015.[14] The research design and analyses described below were registered before outcome data collection with the Open Science Framework.[15]

The primary ethical issue of interest was that of selecting an appropriate control to compare the programme of text messages against. It was argued in the ethics application that, at the time of writing, it was unknown whether the programme of supportive text messages was a good use of scarce resources. It was therefore decided that the supportive text messages should be compared against no text messages at all. None of the participating colleges had started implementing comparable communications with students' family, therefore the control condition constitutes the 'business as usual'.

The study does not involve deception or incomplete disclosure. Although all students in the trial were invited to nominate a study supporter before randomisation, they were informed that these nominated individuals may or may not receive text messages. Informed consent was obtained within the classroom, under the supervision of maths and English tutors. Finally, students individually opt in to take part in the trial and are never automatically defaulted into taking part.

Finally, the experimental design was affected by ethical issues surrounding the automatic study supporter opt-in. Study Supporters are not informed that they are participating in research before they receive the first text message (essentially, they were opted in by the student who nominated them). The two conditions may be more comparable if all study supporters, regardless of treatment status, receive an introductory text message about their enrolment into the programme. However, we were concerned that if students assigned to control were informed of this explicitly, they may be more likely to disengage with their learning. A waiver of study supporter opt-in consent was requested and granted. Study supporters could still opt out of receiving the text messages by replying 'Stop' to any text message.

---

[14] Protocol No. IRB15-3360

[15] The pre-registered analysis plan was registered on November 7th, 2016, before the outcome data was collected. The analysis plan is available on https://osf.io/h87ps/

The approved experimental design therefore involved gaining informed consent from students only and features a pure control condition. Finally, all drafted communications were signed off by the project lead at participating colleges to ensure the information contained was relevant and accurate.

4.3.3 Recruitment

Colleges were recruited as part of a nationwide recruitment campaign and were eligible to participate if they had a cohort of 250 or more students on maths and English qualifications. Nine colleges signed up to take part. All maths and English tutors were instructed to introduce the intervention in class in the fourth week of Autumn term once class sizes had settled down. Tutors received an instruction booklet and a link to the online survey, administered via Qualtrics online survey software. Students were guided to college computer rooms to complete the survey. Once tutors introduced the project using a short script, students were invited to sign up. They were informed that participation is voluntary. The sign-up form required names and mobile phone numbers of up to two individuals who would be interested in and suitable for supporting them in their learning throughout the year (see Appendix 4, p. 247).

There is little indication that tutors deviated from these instructions. They received the information booklets just before the sign-up window opened and were informed that the survey would close two weeks later. Sign-up numbers were monitored throughout this period and college administrators communicated progress with tutors on a daily basis.

Due to constraints in access to IT equipment, five out of nine colleges were unable to implement the online survey procedure. These colleges received paper-based versions of the survey, which was otherwise identical. Recent studies have shown that the mode of data collection for a questionnaire survey, such as paper-based versus online, can influence response rates (Hohwü et al., 2013) and shape the degree to which people respond favourably (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003). Additionally, recent evidence from the Programme for International Student Assessment (PISA) shows that students randomly assigned to sit the computer-based test performed considerably worse than their peers using pen and paper (Jerrim, 2018). Therefore, mode effects cannot be ruled out.

As the survey did not include sensitive items, performance-related questions or timed elements, it is not anticipated that the mode influenced responses strongly. The survey administered for this experiment only required students to complete a small number of questions about their relationship with nominated supporters and provide contact details. Additionally, the intervention itself was not delivered via the survey; only once students nominated a study supporter they would be randomised to receive the treatment or control. Nevertheless, the colleges that requested paper-based surveys because they experienced difficulties with logistics may be different from the other colleges on a number of dimensions. It is not inconceivable that the well-organised colleges were able to introduce the project in a more organised fashion, and that this influenced student decision making. This is explored in Appendix 2, which displays differences in demographic characteristics between students completing the paper-based versus online survey (p. 245).

A number of differences between the groups are immediately apparent. The proportion of consenting students that are young, white and on GCSE courses is considerably higher in colleges that offered the survey online, in comparison to colleges that implemented the paper-based version (*normalised differences* > 0.30)[16]. The proportion of male student is also higher in the online survey colleges (*normalised differences* > 0.05). The study supporter choice also appears to be somewhat different: those who completed the survey online appear to have nominated people they were closer to, as evidenced by the greater frequency of self-reported communication (M_{online}= 5.8, M_{paper}= 5.1).

Qualification achievement rates and assignment to treatment do not differ between the two types of colleges. These differences would pose serious threats to the intervention if the randomisation were carried out at the college level. As described in more detail in Section 4.3.5, individual-level randomisation was carried out (see p. 95). Additionally, all main specifications in this chapter control for college-level fixed effects.

[16] The calculation of normalised differences takes the differences in means between the two experimental groups, divided by the square root of the average of the two conditional within-group sample variances. Normalised differences are further discussed in sections 4.4.3 (p. 108) and 5.3.3 (p. 173).

Figure 2: Welcome screen of the student sign-up survey

Welcome!

Your College would like you to give us names and contact details for two 'Study Supporters'. They should just be people that you think might be good at supporting your learning. You can choose friends, family, colleagues, or anyone else.

We may send your Study Supporters messages about what you're learning, study tips or important dates (but not your grades). Research has found that Study Supporters really help learners succeed!

Tell us your Study Supporters and you could win a £250 Amazon voucher in a lottery! 6 learners will win a £250 Amazon voucher each, you will be told if you have won on Friday 23rd October. Entry is automatic, you don't need to do anything except complete this survey.

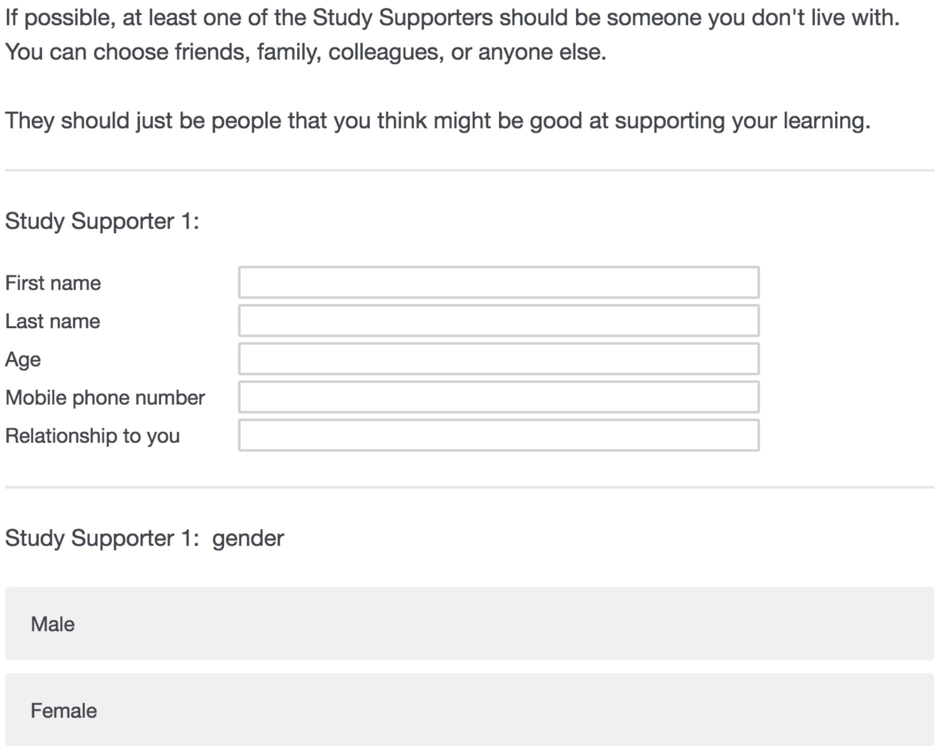
Click 'Next' to tell us your Study Supporters.

You can find more information on the Study Supporter programme here.

You can find more information on the rules of the lottery here.

The opt-in procedure allowed students to choose their study supporter freely. Students were told that they should nominate up to two people who were aged 16+ and who "you think might be good at supporting your learning", as shown in Figure 2. The complete information sheet and survey text are available in Appendix 3 (p. 246) and Appendix 4 (p. 247), respectively. Additionally, as a guide, students were asked to choose one person they did not live with, and one person they did. It is important to note, however, that these guidelines were not enforced. All students who completed the opt-in survey were subsequently randomised; no students were excluded on the basis of the inferred suitability of nominated study supporters. Students' choice of study supporter, and specifically, students' decision to nominate someone within the college, is discussed further below.

Figure 3: Sign-up screen, Qualtrics survey


If possible, at least one of the Study Supporters should be someone you don't live with.
You can choose friends, family, colleagues, or anyone else.

They should just be people that you think might be good at supporting your learning.

Study Supporter 1:

First name

Last name

Age

Mobile phone number

Relationship to you

Study Supporter 1: gender

Male

Female

4.3.4 Power calculations

Since this chapter concerns a communication intervention that relies on people to discuss the content of weekly text messages with each other, potential spillover effects are considered. For example, if participating students speak to each other about the text messages, resulting spillover effects between conditions will bias estimates of treatment effects (Duflo, Glennerster, & Kremer, 2008). When the risks of spillover are severe, randomisation should occur at a higher level. In case of the above example, the researcher may decide to randomise at a cluster level. For example, they may decide to randomise at the class level because they hypothesise that students are less likely to discuss the intervention with non-classroom peers. However, cluster randomisation has an important drawback. Cluster assignment generates more sampling variability, which leads to a considerable increase in the standard error (Gerber & Green, 2012, p. 80).[17] Statistical power, the probability of detecting a treatment effect on the mean

[17] The degree to which the standard error increases with clustering depends on the variability of cluster-level means; if group means are relatively similar (i.e. variability is lower), the standard error is less inflated.

of the outcome variable when there is a true effect, is partly dependent on the size of the standard error, and clustered designs therefore typically suffer from a loss of power in comparison to individually randomised designs (Gerber & Green, 2012). Additionally, if the number of clusters is low, differences in student population between colleges may threaten the integrity of the experiment. Therefore, a researcher would need a large number of participating colleges to be able to randomise at the college level. England only counts 306 colleges, of which some are already involved with other research projects (see for example a recently EEF funded project, Maths-for-Life, which aims to recruit 100 settings; Education Endowment Foundation, 2018).

The present intervention is based on an individual randomised design due to aforementioned statistical power considerations. A total of nine colleges agreed to participate in the trial. Since it was unknown how many students would decide to opt-in, the loss of power arising from class level randomisation was considered to be more severe than potential spillover effects. It was anticipated that the risk of spillover is less acute in this text messaging intervention in comparison to many educational interventions delivered in the classroom. The text messages are sent outside of school hours and are sent only to third parties (i.e. students do not receive texts themselves which they could discuss with classmates). The programme was not referred to within classrooms, except at sign-up which occurred before randomisation. Teachers did not know students' allocation to groups and were asked not to discuss the text messages throughout the academic year. Lastly, students were asked not to nominate a classmate in their maths or English class. They were free to nominate classmates outside of these subjects.

Colleges were unable to provide class-level data at the start of the academic year (i.e. before randomisation) so using a randomisation procedure which simulates the probabilities of exposure to spillover effects was unfeasible (see for example Berlinski et al., (2016), who randomly allocated classes to receive a high or low share of SMS treatment, see p. 8). This study does not have access to class codes, so it is not possible to assess natural variation between classrooms in terms of high of low treatment density. Further, Student IDs were not collected for nominated study supporters, even if they were also students at the same college. This study therefore attempts to minimise the potential for spillover but is not able to eliminate it or assess its impact. As discussed in Chapter 3, the likely impact of spillover on estimates is attenuation of treatment effects. These limitations to the current study are further addressed in the chapter's discussion section.

Sample size calculations are reported for an individually randomised design with a binary outcome variable, see Table 4.3. For these calculations, the required sample size per arm was estimated using the approach taken by Campbell, Julious and Altman (1995). Several assumptions are made. First, the number of students on full-time maths and English courses, per college, was multiplied by the estimated proportion of students opting in to take part, which was estimated at 40%. The resulting estimate numbers of consenting students are displayed in Column 1. The primary outcome variable is whether students pass the qualifications at the end of the year, a binary pass/fail variable. Using national achievement statistics from 2014/15, the baseline proportion of achievement was set at 33.4% (Department for Education, 2016b). Power was set at 80%, and alpha at 0.05. The minimum detectable effect size (MDES) was calculated for several scenarios of opt-in numbers. The MDES (δ) corresponds to the difference between the proportions expected in the control (P1) and treatment group (P2), respectively.

Table 4.3: Ex ante power calculations, individual randomisation

Scenarios: N opt-in	Control N	Treatment N	P1 - Baseline attainment	P2 - Intervention attainment	Estimated MDES (δ)
750	375	375	33.4	43.3	9.9%
1000	500	500	33.4	42.0	8.6%
1250	625	625	33.4	41.1	7.7%
1500	750	750	33.4	40.4	7.0%
1750	875	875	33.4	39.9	6.5%
2000	1000	1000	33.4	39.4	6.0%

Notes: Power was set at 0.80 and alpha at 0.05. The allocation proportion was set at 50:50.

As can be seen from Table 4.3, achieving the most optimistic sample size estimate (N = 2000) would power us to detect an increase in achievement rates of 6 % points. If 1500 students sign up, the estimated MDES is 7 % points. Previous studies have found effects of a similar magnitude. For example, Chande et al. (2017) implemented a text-messaging experiment in further education colleges and found an 8 % point increase in students' likelihood of passing the course.[18] These power calculations show that a minimum of 1500 students would need to opt in for the study to be moderately well-powered.

4.3.5 Randomisation

Once the student opt-in window closed, all completed surveys were gathered. Three research assistants entered the paper-based survey data onto a shared spreadsheet. Four colleges were involved in the implementation of two other interventions, Values Affirmation and Grit, discussed elsewhere (Hume et al., 2018a, 2018b). Within these 'mixed colleges', 40% of classes were allocated to receiving the Study Supporter sign-up survey. Within these classes, students who provided active written consent were subsequently assigned to treatment or control groups.

The remaining five colleges only implemented the Study Supporter programme. For these colleges, all maths and English learners on Functional Skills and GCSE qualifications were eligible to take part. Colleges introduced the programme to all learners, and those who opted in were also assigned to treatment and control in one batch across all colleges.

Each student was treated only in their maths or English course, regardless of whether they took either or both courses. The subject assignment was determined at a college level: some colleges received the intervention only in maths courses, and some colleges only did so in English courses. If a student at College A (assigned to maths) took both maths and English, they were only treated in their maths class. Therefore, each observation in the dataset corresponds to an individual in the course (either maths or English) their study supporters were informed about.

Randomisation was stratified by a number of pre-treatment covariates, namely gender, age (16-18 vs. 19+) and qualification type (FS versus GCSE). Table 4.4 considers the split between groups, and the subject treated for all

[18] The authors sent direct text messages to students, with reminders about upcoming exams, planning prompts, and general motivational content. The content of the text messages is similar to the intervention discussed in this chapter and the mode of delivery is identical, but the recipient is the FE college student, rather than a third party.

participating colleges. 868 students are allocated to the control group (50.9%), and 838 to the treatment group (49.1%).

Table 4.4: Distribution of sample across colleges and treatment groups

College	Control	Treat	Total	Subject treated
College A	73	58	131	Maths
College B	136	145	281	Maths
College C	30	25	55	English
College D	172	165	337	English
College E	47	47	94	English
College F	135	127	262	English
College G	108	98	206	Maths
College H	113	132	245	Maths
College I	54	41	95	Maths
Total	868	838	1706	-

4.3.6 The social support intervention

The intervention was intended to be a light-touch exercise for college staff and administrators. I created the sign-up materials for teachers and students, collected colleges' schemes of work (i.e. curriculum planning documents), and drafted text messages. One lead teacher at every participating college was then asked to provide feedback on the drafted messages, and sign-off once they were happy with the full-year schedule. Text messages were sent to study supporters' mobile phones at weekly intervals during the academic year. The content was developed using the following guidelines, informed by prior studies (Chande, 2017; York & Loeb, 2014):

1. Describe class material (both what was taught last week, and upcoming topics) in a non-technical manner;
2. Reference a question prompt or interesting fact, in order to stimulate genuine curiosity in the study supporter;
3. Use a positive tone, supporting positive study-related behaviours (rather than telling supporters and students what *not* to do);

4. Help navigate the education system (e.g. what to do when students had not received feedback on their assignment, or how to prepare for an exam);
5. Remind supporters about upcoming exams and assessments and encourage supporters to help students plan and organise study behaviours both at college and at home.

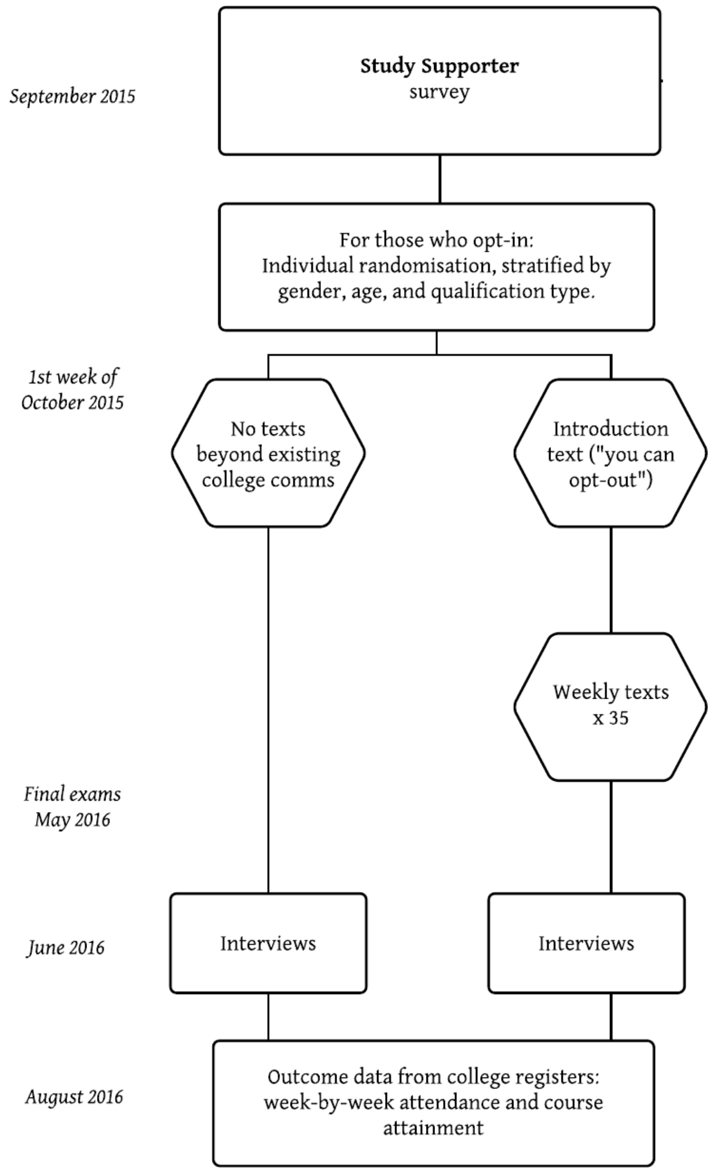
Study supporters received a balanced mix of text messages. For example, if a message in week 1 referred to an upcoming exam, week 2 would focus on course content or academic resources. The resulting text messages included the following information, (1) course content; (2) advance notice of upcoming exams; (3) academic resources available to the student; and (4) general positive reflective conversation prompts. For example:

> Hi [supporter forename][19], [student forename] has recently learnt about percentages. Ask [him/her] to calculate the final price of a £250 TV after adding 20% VAT (tax on things you buy) and show you how [he/she] worked it out. Thanks, [College]

Additional example text messages are provided in Appendix 5, p. 250. All colleges received a unique schedule of text messages, as text content was tailored to college exam and term dates and the course curriculum. However, the length and content of messages were comparable across colleges. A total of 35 weekly messages were sent out to study supporters: one message per week. There was no variation in dosage between colleges or types of courses. The stages of the intervention are graphically presented in Figure 4.

[19] The fields in brackets '[...]' were automatically merged with contact information of students and their study supporters stored on the text messaging platform, FireText, ensuring that all recipients received personalised text messages.

Figure 4: Timeline of the Study Supporter intervention



4.3.7 Quantitative analysis plan

Throughout this thesis I use regression-based methods,[20] in line with the empirical strategy used in other supportive information field experiments (e.g. Rogers & Feller, 2018, Bergman & Chan, 2017, York, Loeb, & Doss, 2018). As discussed in greater detail in Chapter 3, random assignment allows the identification of the causal impact of communicating with nominated third parties on students' class attendance and achievement. Intention-to-

[20] With the exception of the duration modelling of students' likelihood of drop-out from college, performed in Chapter 5, Section 5.6.4, p. 190.

treat (ITT) analysis is performed, which reflects intended treatment assignments, rather than whether the intended recipient actually did receive all text messages. The ITT effect therefore is a measure of the average effectiveness of a programme, regardless of compliance with assignment. ITT provides a lower bound estimate of the impact of supportive text messages on student success, in situations where compliance is incomplete. Given that non-compliance (e.g. lack of uptake, spillover) is an issue in most policy contexts, the ITT estimate provides a good basis for cost effectiveness calculations (Gerber & Green, 2012, p. 150). ITT is the recommended strategy in the CONSORT guidance (Schulz, Altman, & Moher, 2010) and was pre-registered as the analysis strategy for this study.

An ordinary least squares (OLS) approach is taken for all regression models, including those with binary dependent variables. This choice was motivated by the following observations. First, a linear probability model is more easily interpretable than a logistic regression model (Hellevik, 2009) as the former produces coefficients that can be directly interpreted as percentage point changes between the control and treatment groups[21] whereas the latter involves logged odds or predicted probabilities. Additionally, as the modelled probabilities are not expected to be extreme (either close to 0 or 1 where everyone fails or passes), but rather between .20 and .50 (see power calculations; Table 4.3, p.94), the logistic model is unlikely to fit considerably better than a linear model would (Hellevik, 2009, p. 67). Lastly, unobserved heterogeneity through omitted variables affects logistic regression coefficients, making it more problematic to compare coefficients across models with different independent variables, across samples or over time (Mood, 2010).

It should be noted that Kuha and Mills (2018) challenge the widely-held view that logit coefficients or odds ratios cannot be compared between groups. They argue that empirical researchers should instead focus clearly on exactly what it is they are estimating in a binary response model (Kuha & Mills, 2018). If researchers know the target quantities of their analysis, and their target populations in general, comparisons between different groups or models should pose no fundamental problems (Kuha & Mills, 2018). As a

[21] When β_1 is 0.06, for example, it can easily be interpreted that a one-unit increase in T_i is associated with a 6 % point increase in the probability that Y is 1. In the present chapter, a β_1 of 0.06 indicates that students in the treatment group (T= 1) are 6 % points more likely to achieve their qualification.

robustness check, logistic binary regressions of the Study Supporter intervention on qualification achievement rates are conducted, and further discussed in section 4.7, p. 112.

4.3.7.1 Average treatment effects

The ordinary least squares (OLS) regression specifications are provided below:

[8a] $$Y_i = \alpha + \boldsymbol{\beta_1} T_i + \epsilon_i$$

[8b] $$Y_{ij} = \alpha + \boldsymbol{\beta_1} T_i + \beta_2 \boldsymbol{X}_i + \beta_4 \boldsymbol{C}_j + \epsilon_i$$

where:

Y_{ij} represents the outcome of interest for student *i* in college *j*. In case of the primary outcome of interest, achievement, Y_{ij} represents a binary variable of the learner's final grade. For students on GCSE courses, Y_{ij} is equal to 1 if individual *i* scored A* - C and 0 if individual *i* scored below a C (i.e. D - U). In Functional Skills courses, Y_{ij} is equal to 1 if individual *i* scored a P (i.e. pass) and 0 if individual *i* scored an F (i.e. fail). For attendance Y_{ij} represents the average attendance rate for individual *i* in their treated subject, calculated as the total number of attended classes divided by the total number of scheduled classes.

α is the regression constant;

$\boldsymbol{\beta_1}$ is where the estimate of the intent-to-treat effect of the supportive text messages is captured. A positive and statistically significant estimate of $\boldsymbol{\beta_1}$ suggests that simple prompts sent to third parties improved student success at college;

T_i is the treatment indicator, equal to 0 for control group participants, and 1 for treatment group participants;

$\boldsymbol{X}_i$ is a vector of student-level covariates including gender, age and qualification type. Randomisation was stratified by these covariates;

$\boldsymbol{C_j}$ is a vector of college-level fixed effects; and,

ϵ_i is an individual-level error term. Robust Huber-White standard errors are calculated as they are typically more conservative than conventional standard errors (Angrist & Pischke, 2008, Ch. 8, p. 221).

4.3.7.2 Heterogeneous treatment effects

The above regression specification assesses the average effect of the intervention on student outcomes. To test whether the intervention has a differential effect on different subgroup of students, the regression model is run separately for the subgroups of interest. The variability in treatment effects is of interest, both in terms of policy implications and furthering understanding of mechanisms (Gerber & Green, 2012). It is important to understand which individuals benefit most from a given intervention, and under what conditions (Duflo et al., 2008; Gerber & Green, 2012). Similarly, an intervention could have an opposite (and unintended) effect for a subgroup (Bamberger et al., 2016). Treatment effect heterogeneity discussed in this chapter focuses on treatment-by-covariate interactions, where covariates of policy relevance are selected: age, gender and qualification type. Exploring such variation in treatment effects should be paired with cautious interpretation, as the experiment is powered to detect average treatment effects (ATE) rather than heterogeneous treatment effects (HTE). The exploratory subgroup analyses are included because they could potentially inform the research design or focus of the follow-up field experiment (Chapter 5).

As national achievement statistics introduced in Table 4.1 show clear age-related differences (p. 82), I consider how treatment effects varied by age: 16-18 year olds versus 19+ learners. Second, heterogeneity of treatment effects by qualification type (GCSE versus FS) is examined. Thus far, studies on attendance and attainment at further educations have not explored the effectiveness of interventions by qualification type. Finally, differences between male and female students are examined, as a large literature points towards gender differences in the sources of support they turn to and the type of support they utilise (Day & Livingstone, 2003) and how their well-being is affected (Rueger, Malecki, & Demaray, 2010). Heterogeneity of treatment effects by college and subject are not examined, due to small sample sizes and college-level assignment, respectively.

4.3.8 Effect sizes

Hedges' *g* is reported in the main tables in this chapter, and Cohen's d and Glass' Δ are reported in the appendices. These standardised mean difference (SMD) effect sizes are scale-free and allow researchers to quantify the magnitude and direction of the difference between two groups (Durlak, 2009), and assess the relative effectiveness of different interventions (Higgins & Katsipataki, 2016). Hedges *g* is used by the Education Endowment Foundation, the What Works Clearinghouse and the Campbell Collaboration, and therefore allows for comparison across education trials. Hedges *g* is a modification of Cohen's *d* (Hedges, 1982), but both are reported for ease of comparability. Glass' Δ was developed specifically for experimental studies, in recognition that the treatment may affect the homogeneity of variances in the treatment group. It therefore takes the standard deviation of the control group, rather than the pooled standard deviation (as is the case with Cohen's *d* and Glass' *g*). I use the unconditional sample variance for all calculations. All formulae as used in this chapter are reported in Appendix 6, p.252.

4.4 Outcome data

4.4.1 Attendance and achievement data

All outcome data was collected from participating colleges directly. Attendance data was collected in a week-by-week format. Actual and missed attendances are recorded for every maths or English course consenting participants attended. Average attendance rates are calculated by dividing actual attendances by the maximum number of possible attendances for a given class.

Some students switched classes during the academic year, so a small number of students appear in the dataset several times (N = 77). For this subset of students, the following steps were taken. First, data for the subject they were not treated in are discarded. Second, attendance rates are calculated for every student-course pair. Attendance rates are then merged into a single measure of attendance (i.e. if a student appears to take 3 different maths classes within the same qualification, their final attendance rate is the average of the three). The final dataset contains one observation per student (N = 1638). The distribution of the attendance variable is visualised in Appendix 7, p. 254.

Achievement data consists of students' GCSE or FS final exam grades. GCSE grades range between U and A*, whereas Functional Skills grades are recorded as either pass or fail. Students' interim assessment scores were not collected.[22] As described in the section on regression specifications (Section 4.3.7), GCSE grades were distilled into a binary pass/fail measure where C was the cut-off for a passing grade. The rationale for using this outcome measure is as follows: if a student does not obtain a C at the end of the year, they will still be required to resit the examination, even if they improved their grade between secondary school and post-16 study. This dataset also contains one observation per student (N = 1451).

Colleges provided attendance and achievement data separately, and the resulting datasets are kept separately. Both datasets were merged with consent form data, which specifies who consented to be part of the study supporter intervention, as well as their random assignment to treatment or to control.

4.4.2 Attrition and non-response

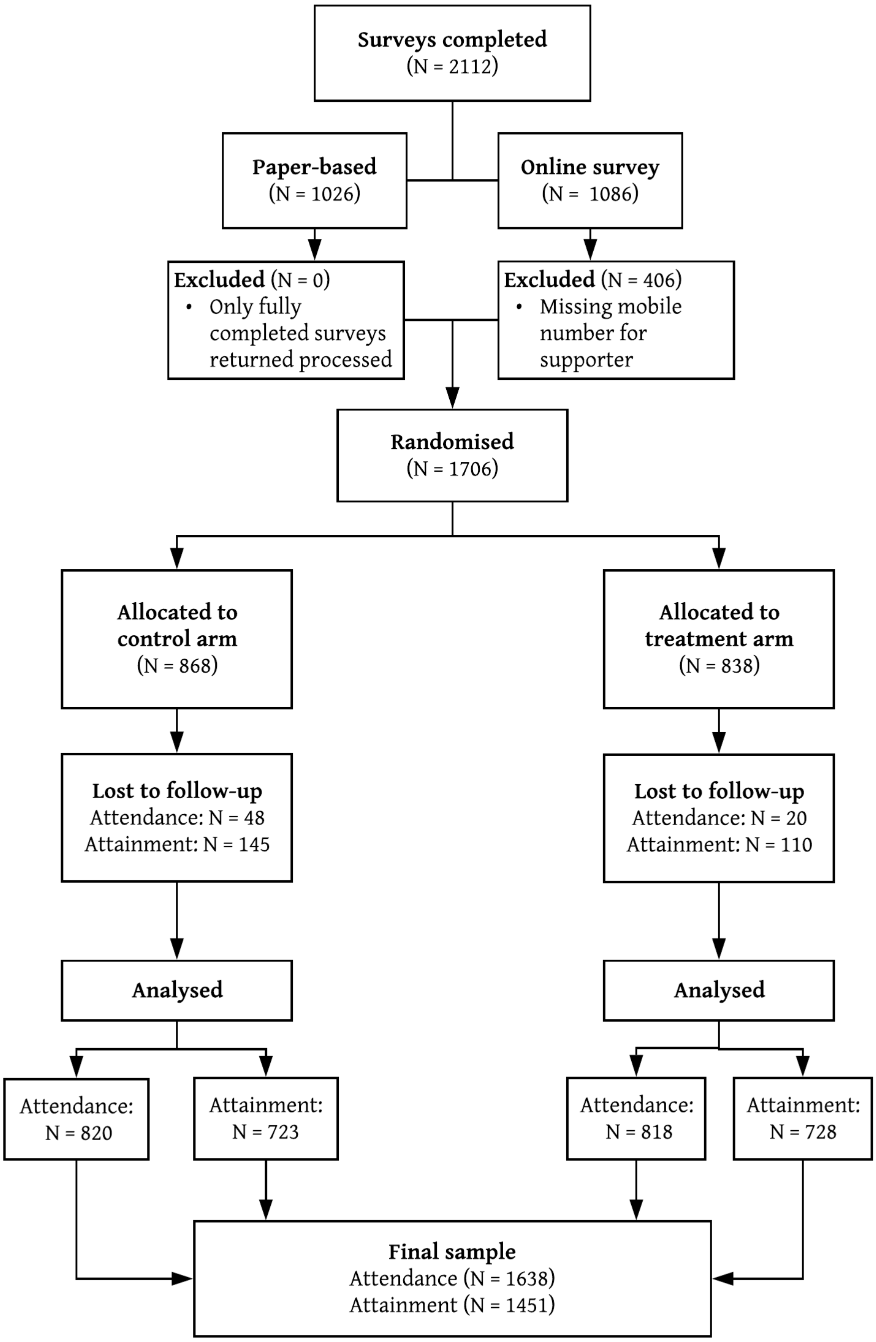
There are two sources of attrition in this field experiment: missing achievement scores and missing attendance data. One college was unable to provide attainment data due to a recent switch to a new management information system and they had issues with the data migration. All colleges provided attendance data. Additionally, a small number of students (N = 68) signed up to take part with a Student ID that was not recognised by their respective colleges. Upon inspection, some errors were likely due to spelling errors, and some were deliberately incorrect. Where the issue was likely a small spelling error, college staff attempted to match based on student name. Due to data protection considerations, only the first initial of students' last names was requested. For 68 students, no match was obtained. See Appendix 8 for the final sample, by college and outcome measure, and Figure 5 below for a visualisation of the recruitment, randomisation and analysis stages.

Differential attrition by treatment status impairs the researcher's ability to make unbiased causal inferences (Gerber & Green, 2012, p. 211). As shown in Figure 5, a slightly larger number of students were assigned to control than

[22] For Functional Skills courses, interim assessment grades may count towards the final grade. Only the final grade was collected from colleges, however, as this grade incorporates progress made throughout the year, and ultimately determines whether students achieve their qualification.

treatment (50.9% vs. 49.1%) and a larger number of students in the control group were lost to follow-up. The final sample is balanced across the two conditions. Since baseline data was not collected, potential correlates of attrition such as baseline grades or baseline attendance cannot be assessed. This limitation is addressed in Chapter 5, where baseline attendance and achievement information are collected for every participating student.

Figure 5: Study Supporter flow chart (CONSORT)



Finally, a number of students decided not to sit the final exam or left college during the trial period (N = 156). In line with the Intention-to-Treat analysis used in this chapter these individuals were retained, and their outcomes are assigned to a fail (i.e. no distinction is made between performing poorly on the exam versus not taking it). Retaining dropped-out students helps avoid potential selection effects that could arise from removing such cases from the primary analysis. All students starting the experiment are included in the analysis under the intention-to-treat estimator.

4.4.3 Balance checks

Balance checks were carried out using the learners who could be matched back to the randomisation data. To assess the degree of (im)balance in the covariate distributions between students in the control and treatment group, normalised differences are reported in Table 4.5. This calculation uses baseline data to assess covariate balance prior to analysis, and is advocated by Imbens and Rubin (2015). It takes the differences in means between the two experimental groups, divided by the square root of the average of the two conditional within-group sample variances.[23] This approach does not test whether the covariate means are significantly different from one another at baseline. Testing for baseline differences is now widely discouraged because it is not clear how the resulting p-values help assess whether statistically significant baseline differences are 'real' or are simply unavoidable due to the multiple significance tests (De Boer, Waterlander, Kuijper, Steenhuis, & Twisk, 2015, p.3).

Naturally, there are techniques to overcome these flaws; a relatively straightforward option is to discard the t-statistics in favour of alternatives that are less easily influenced by sample size. Calculating normalised differences allow the researchers to assess whether adding the covariates to the regression model will adequately remove most biases in estimated group differences (Imbens & Rubin, 2015). Normalised differences are scale-invariant which allows for straightforward comparisons between groups or studies.

[23] The formula provided by Imbens & Rubin (2015, p. 311) is as follows: $\hat{\Delta}_{ct} = \frac{\overline{X}_t - \overline{X}_c}{\sqrt{(s_t^2 + s_c^2)/2}}$

Table 4.5: Balance between treatment and control groups, attendance dataset

Covariates	Control M (SE)	Treat M (SE)	Normalised difference $\hat{\Delta}_{ct}$
Female (proportion)	0.469 (0.019)	0.488 (0.019)	0.026
Age (years)	18.667 (0.197)	18.787 (0.196)	0.011
White (proportion)	0.477 (0.019)	0.467 (0.019)	-0.013
GCSE (proportion)	0.529 (0.017)	0.546 (0.017)	0.016
Maths (proportion)	0.545 (0.017)	0.546 (0.017)	0.002
Observations	820	818	

Notes: Mean values reported with robust standard errors in parentheses for all continuous and binary variables. Age and belonging are continuous variables, and all other variables are binary. Data on student demographics was merged with the attendance dataset, therefore the sample size used to assess balance corresponds to analyses reported for attendance.

When $\hat{\Delta}_{ct} > 0.10$, the differences between the covariate distributions are potentially more challenging to correct for in analyses (Austin, 2009). Student demographics at baseline are comparable across the control and treatment groups for all baseline covariates I was able to collect. For both attendance and achievement rates, the addition of covariates changes the point estimates only slightly, and ultimately does not change the inferences made. Appendix 9 provides the balance between treatment and control for the attainment dataset. The only notable difference in balance checks of the attainment datasets is that the proportion of students on GCSE courses is higher in the treatment than control group ($\hat{\Delta}_{ct}$ = 0.08; p. 255).

4.5 Descriptive statistics

4.5.1 Student demographics

Data on gender and age is collected through data provided by the college. Demographic information is missing for a number of students in the attendance dataset, but the attainment dataset does not contain missing variables.[24] Of the full sample of 1638 students (attendance data), gender and

[24] These datasets cannot be combined because they are stripped of the students' unique reference number, which is the only variable these datasets can be merged on.

age are missing for 14.1% and 5.3% of students, respectively. Where it is missing, age and gender are assigned to an arbitrary category and indicators for missing are added to the regressions.[25] Binary indicators for missing data are entered into the main specification models. As can be seen in Table 4.6, missingness of covariates is more prevalent in the control group. Table 4.6 also shows that the rate of student drop-out from college is larger in the control group. Unlike the covariates in the table (gender, age and qualification type), this may have been affected by the treatment itself.

Table 4.6: Missingness rates of covariates by treatment status

Missing (proportion)	Control M (SE)	Treat M (SE)	Normalised difference $\hat{\Delta}_{ct}$
Missing: gender	0.161	0.121	-0.081
	(0.013)	(0.011)	
Missing: age	0.060	0.046	-0.042
	(0.008)	(0.007)	
Missing: qualification	0.129	0.094	-0.079
	(0.012)	(0.010)	
Attrited (drop out)	0.112	0.078	-0.082
	(0.011)	(0.009)	
Observations	820	818	

Notes: Mean values reported with robust standard errors in parentheses for all continuous and binary variables. Age and belonging are continuous variables, and all other variables are binary. Data on student demographics was merged with the attendance dataset, therefore the sample size used to assess balance corresponds to analyses reported for attendance.

Within the group of students where demographic variables are not missing, gender is well-balanced, and the majority of participants is young: 52% are female and 76% are between 16 and 18 years old. The second largest age group is those aged 19 to 25 (17%), and only 7% of the sample is older than 26.

As described in section 4.4.1, observations of student attendance and achievement in non-treated classes are discarded. Subject treatment (i.e. maths, English) is assigned at the college level. For example, in college D, only maths classes were treated, and so English classes are excluded from the dataset. The qualification studied (FS, GCSE) is derived from class codes, which vary between colleges but typically include clear identification of

[25] Gender is coded as 1 if female, and 0 if male or missing. Age is converted to a binary 16-18 versus 19+ measure. Missing age is grouped with the 19+ cohort of students.

course type. Qualification type is missing for 11.2 % of our sample and the remaining 53.5% of students in the sample are on GCSE courses and 35.3% are on Functional Skills courses.

4.5.2 Relationship with nominated Study Supporters

During the sign-up procedure at the start of the academic year, students were asked to describe their relationship with the person(s) they nominated. Their free-text responses were subsequently coded and categorised into broad types of relationships. A small proportion (6%) of students left this question blank. The majority of students nominate either a member of their nuclear family (40.6%)[26] or a peer inside or outside of college (41.4%) to be their first study supporter. 80% of the sample also nominated a second study supporter, where peers were the most popular category at 39.4 % overall. See Appendix 10 for the full breakdown of supporter categories.

Of the students who nominated a peer as a study supporter, 82% of participants indicated that this individual worked or studied at the college. Student IDs of nominated supporters were not collected; thus, I cannot check whether they attended the same maths or English class. Only 21 students indicated they had nominated a classmate (1.5%) in their maths or English class, thus it appears likely that students have nominated classmates from their vocational classes or previous year group. Nevertheless, the risk of spillover is evident, and cannot be assessed empirically. Limitations of the data will be discussed in Chapter 7, section 7.2, p. 236.

Students also completed a number of questions about their closeness to the nominated study supporters, and their frequency of contact. The average age of the study supporters was 28.1 (SD = 13.6). Average age varied considerably between types of study supporters, naturally, as peers were generally of the same age as students (M = 19.5, SD = 7.2) whereas parents (M = 43.8, SD = 7.1) and other relatives are older (M = 33.0, SD = 18.7). A more detailed breakdown can be found in the final rows of Appendix 11 (p. 257). Students also indicated that, on average, they spoke with the nominated individuals five out of seven days (SD = 2.2). Forty-four per cent of students indicate they co-habit with their first nominated supporter. The self-reported data also indicates that students felt emotionally close to the third parties they nominated, as 84.3 per cent felt either very close or close to their first study

[26] Of which parents were the most popular choice, at 29.7%.

supporter, and 68.8 per cent felt (very) close to Study Supporter 2. Appendix 12 provides additional descriptive statistics on nominated study supporters (p. 258). I now turn to the primary results.

4.6 Primary analyses

Table 4.7 presents the primary analysis for attendance and attainment. Column 1 conducts a basic regression model, regressing only the treatment indicator on the outcome variable. Column 2 is the primary regression of interest for attendance, controlling for student age, gender, qualification type, subject, and college-level fixed effects. The primary analyses address the primary research question, which asks if a programme of supportive text messages can help improve student attendance and attainment. [27] It was hypothesised that having access to a supportive third party may help students feel more engaged at college. This change was predicted to motivate students to attend class more often and do better in their final exams. I now turn to the treatment effects on attendance and attainment in turn.

First, students who were randomly assigned to having a study supporter who received weekly texts messages attend their maths or English class significantly more often. I find statistically significant effects (p = 0.009) in the basic regression, and slightly smaller, but still statistically significant effects when controlling for student-level covariates and college fixed effects (p = 0.034; see Figure 6). The comparison of class attendance across experimental groups (Column 1 and 2) suggests that students whose supporters were texted attended on average 3.1% to 4.8% more classes than participants in the control group depending on model specification.

Standardised mean differences between the two groups post-intervention are reported in the table. In keeping with conventions, Hedges' g is reported in the table below; Cohen's d and Glass' Δ are reported in Appendix 13 as they are similar to Hedges' g values. Covariate adjusted means are used for effect size calculations in Column 2. The effect size (Hedges' g) of the intervention on attendance rate ranges between 0.11 and 0.13, depending on model specification. These effect sizes are moderate and comparable to effect sizes

[27] The secondary research question is addressed in Chapter 5 since it concerns the effectiveness of text messages delivered to directly to students versus supportive texts delivered to their nominated supporters.

reported in evaluations of other information interventions (see for example Chande, Luca, Sanders, Soon, & Borcan, 2017). Comparisons with similar interventions in terms of effect sizes will be further addressed in the discussion section of this chapter, section 4.9.1 (p. 132).

Table 4.7: Average treatment effects of the intervention on attendance and achievement

	(1)	(2)	(3)	(4)
	Attendance rate		Achievement rate	
	Basic model	*Inc. covariates*	*Basic model*	*Inc. covariates*
Treatment	0.048**	0.031*	0.071**	0.060**
	(0.018)	(0.015)	(0.023)	(0.022)
Constant	0.555**	0.619**	0.223**	0.159**
	(0.013)	(0.039)	(0.015)	(0.056)
Student-level covariates	*No*	*Yes*	*No*	*Yes*
College fixed effects	*No*	*Yes*	*No*	*Yes*
Control mean	0.555	0.563	0.223	0.229
N observations	1638	1638	1451	1451
R-squared	0.041	0.356	0.007	0.070
Hedges g (ES)	0.129	0.105	0.163	0.136

Notes: The columns report the intent-to-treat (ITT) estimate and robust standard errors (in parentheses) of individual-level random assignment on outcomes measured by the end of 2016. Attendance is recorded on a scale between 0 and 1, as the proportion of classes attended throughout the full academic year. Whether the students passed the course is recorded as a binary variable, and therefore the achievement rate (see Column 3 and 4) denotes the proportion of students passing the course. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) and missingness dummies as pre-specified. Sample size between attendance and achievement results varies due to missing values. Effect size calculations use unconditional standard deviations, and covariate adjusted means are used for the calculations in column (2) and (4).

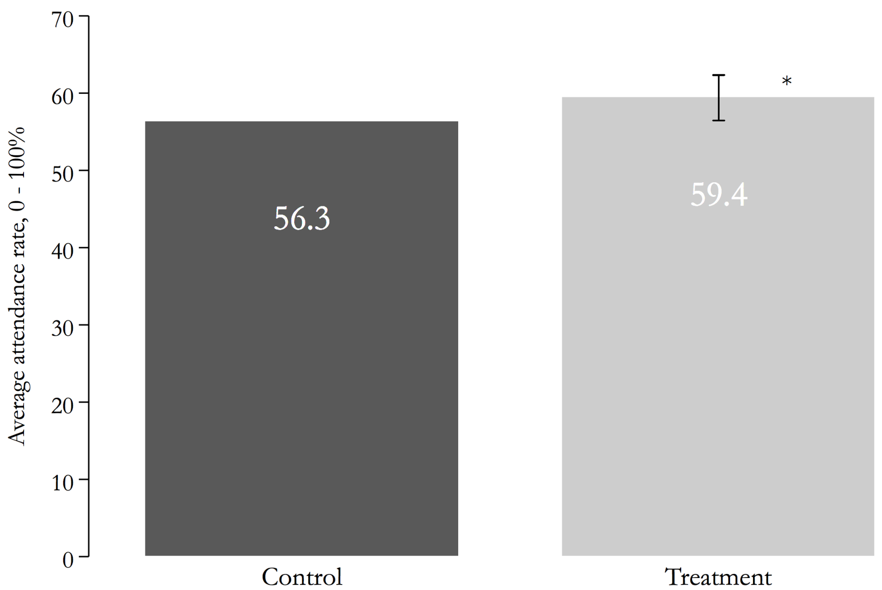
+ = $p < 0.10$, * = $p<0.05$, ** = $p<0.01$.

Second, treatment effects are detected on the probability that students achieved their qualification, see Table 4.7 (Column 3 and 4). Similar to the first set of regression specifications, the 'crude' effect is estimated first, and then adjustments are added for the same set of baseline variables in the final column. Students' probability of achieving the qualification increases by 7.1

% points in the basic model (p = 0.002) and treatment estimates remain similar at 6 % points when the same set of baseline covariates as above are introduced (p = 0.008; see Figure 7). The mean proportion of students in the control group who achieved their qualification was 22.8%, and raised to 28.8% for students whose supporters were texted. The effect size is larger for attainment than attendance; Hedges' g varies between 0.16 and 0.13 depending on whether unadjusted (Column 3) or adjusted means (Column 4) are used for the calculations. These findings are in line with previous studies that found positive impacts of text messaging interventions on student achievement (see for example Bergman & Rogers, 2017).

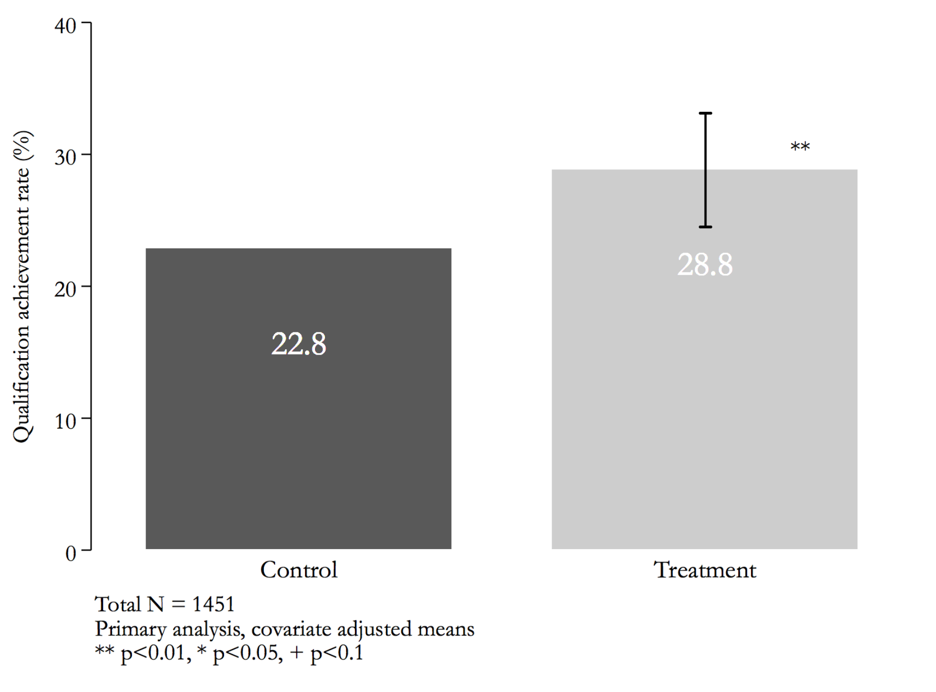
The results are promising. Positive treatment effects from the supportive text messages imply that nominated study supporters discussed the content of the text messages or provided extra support in other ways. Students never received text messages themselves, therefore the observed treatment effects must have occurred through interaction between students and their study supporters. Whether these interactions do indeed occur was not observed for the full sample, which is noted as a limitation of the research design. A small number of participants were interviewed in order to explore the types and frequency of supportive interaction between both parties. The qualitative component is introduced in Section 4.8, p. 116.

Figure 6: Average treatment effects, attendance rate



Total N = 1638
Primary analysis, covariate adjusted means
** p<0.01, * p<0.05, + p<0.1

Figure 7: Average treatment effects, achievement rate



4.7 Secondary analyses

In Table 4.8, heterogeneous treatment effects on attendance rates are examined. This is an exploratory analysis. First, the analysis is partitioned depending on whether a participant is taking GCSEs (column 1), or Functional Skills (column 2) qualifications. Neither estimate of the effect of the intervention on the subgroup is statistically significant ($p > 0.05$), which may be primarily attributable to a lack of power since the point estimates are similar to the primary results in Table 4.7.

Second, the regression specification is run separately for young students (aged 16 – 18; column 3) and adult learners (aged 19+; column 4) to explore whether the intervention primarily benefits traditional students or returning students. The treatment estimate is statistically significant for younger students ($p = 0.012$, Hedges' $g = 0.12$), suggesting that this cohort may be driving the effect. Younger students, at post-compulsory education institutions, have been found to have lower baseline motivation to learn than older learners (Gegenfurtner & Vauras, 2012), so there may be more scope to have an impact on their level of motivation. Additionally, young students' reasons for absences may be related to poor motivation, whereas older learners often face more situational hurdles including family commitments and work conflict (Osam, Bergman, & Cumberland, 2017). A supportive third party may be more able to address low motivation than to resolve structural or situational hurdles for their supportee. To this end, Chapter 6 explores the types of support study supporters tend to provide (See Section 6.4, p. 221).

Finally, differential effects of the text messages by gender are examined. The analysis is partitioned by female (column 5) versus male and non-identified combined (column 6). The exploratory analysis finds no effect of the intervention on attendance rates for female participants but reports significant (p = 0.008, Hedges' g = 0.13) effects for non-female participants. The mean attendance rate of male/non-identified students is 49%, considerably lower than females (M = 65%), so there may be more room for improvement in the former group. Appendix 14 reports Cohen's d and Glass' Δ for the subgroup analyses (p. 259).

Table 4.8: Heterogeneous treatment effects on attendance

	(1) GCSE	(2) FS	(3) 16- 18	(4) 19+	(5) Female	(6) Male
Treatment	0.030	0.029	0.042*	0.001	0.004	0.053**
	(0.020)	(0.022)	(0.017)	(0.030)	(0.022)	(0.020)
Constant	0.671**	0.609**	0.644**	0.542**	0.662**	0.600**
	(0.052)	(0.055)	(0.035)	(0.114)	(0.072)	(0.048)
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.634	0.483	0.554	0.587	0.653	0.490
N observations	876	762	1180	458	733	905
R-squared	0.151	0.456	0.385	0.310	0.169	0.436
Hedges g	0.095	0.070	0.116	0.001	0.001	0.134

Notes: All analyses are OLS regressions, including fixed effects at the college level. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) and missingness dummies as pre-specified. Huber white standards errors, clustered at the student-level, in parentheses. Effect size calculations use unconditional standard deviations, and covariate adjusted means are used for all calculations.

+ = p < 0.10, * = p<0.05, ** = p<0.01.

Heterogeneous treatment effects on attainment are examined next. Table 4.9 provides subgroup analysis on achievement rates by qualification type (column 1 – 2), age (column 3 – 4), and gender (column 5 – 6). The primary treatment effect is no longer statistically significant when estimated for GCSE and FS courses separately ($p = 0.086$ and $p = 0.056$ respectively). The intervention improves pass rates significantly ($p = 0.005$, Hedges' $g = 0.17$) for the younger cohort of students and does not impact adult learners ($p > 0.05$), which is in line with the attendance results for this particular subgroup. Finally, the intervention significantly improved pass rates for female students ($p = 0.040$, Hedges' $g = 0.14$), but not for male students ($p = 0.068$).[28]

Table 4.9: Heterogeneous treatment effects on achievement

	(1) GCSE	(2) FS	(3) 16- 18	(4) 19+	(5) Female	(6) Male
Treatment	0.051+	0.064+	0.073**	0.021	0.065*	0.058+
	(0.030)	(0.033)	(0.026)	(0.043)	(0.032)	(0.032)
Constant	0.207**	0.232*	0.163**	0.106	0.036	0.221**
	(0.068)	(0.105)	(0.057)	(0.129)	(0.080)	(0.074)
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.252	0.199	0.206	0.301	0.254	0.199
N observations	852	599	1114	337	755	696
R-squared	0.072	0.153	0.021	0.354	0.125	0.032
Hedges' g	0.115	0.152	0.171	0.044	0.144	0.138

Notes: All analyses are OLS regressions, including fixed effects at the college level. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) and missingness dummies as pre-specified. Huber white standards errors in parentheses.
+ = p < 0.10, * = p<0.05, ** = p<0.01.

At first sight, the treatment estimates for male and female subgroups for achievement rates run counter to the treatment estimates on attendance for these two subgroups. The subgroup analysis on class attendance (presented in Table 4.8; Columns 5 and 6) shows that the effect of the intervention was significant for male (and non-identified) students, but not for female

[28] The achievement dataset does not contain missing values for covariates – gender, age, and course type are known for the full sample.

students. The opposite is true in the regression results for attainment: females experienced greater treatment effects (p = 0.04, Hedges' g = 0.14) than their male/non-identified counterparts (p > 0.05; Table 4.9; Columns 5 and 6).

As discussed in the introduction to this chapter, class attendance is hypothesised to lead to improvements in pass rates (Credé, Roch, & Kieszczynka, 2010). Especially at further education colleges where much of the learning occurs within the classroom (as opposed to self-study; Higton et al., 2017), improvements in attendance should lead to improvements in attainment. However, attending class is not equal to paying attention in class. Nevertheless, it is perhaps surprising that the positive effects on attendance for male students do not flow through to statistically significant treatment effects on achievement for this subgroup. Similarly, the positive treatment effects for female students do not flow through attendance, so other mechanisms may be at play. Since these subgroup analyses are not well-powered, these results should be further explored in larger-scale follow-up studies.

As a robustness check, logistic binary regressions of the Study Supporter intervention on qualification achievement rates are reported in Appendix 16 (p. 260) and Appendix 17 (p.261). These estimates are consistent with the primary findings reported above, with the exception of the subgroup of students taking Functional Skills courses. The treatment effect is statistically significant at $p < 0.05$ in the logistic regression model while just failing to reach statistical significance in the primary regression model. The size of the effect is comparable, however. In summary, I find that the results are consistent across the linear probability and logistic regression models reported.

The above results also call for critical reflection on assumptions made when sample size calculations were performed. The baseline achievement rate was estimated at 33.4%, based on national statistics from 2014/15 (Department for Education, 2016b). In this experiment, baseline achievement is considerably lower at 22.3%. To assess whether this lower average pass rate was a feature of the specific intervention context or data cleaning procedure (i.e. attrited students were null-imputed), the achievement rates are compared to national qualification achievement rates (NQARs). The national 2015/16 results across further education colleges were very similar to the results reported in this chapter. Among students who started the learning aim at an FE college, 23.8 % achieved their GCSE English qualifications and 23.4% achieved maths (qualification achievement rates; Department for Education, 2017a). See Skills Funding Agency, 2017, p.4 for the qualification achievement rate formulae. In conclusion, the achievement rate of students in the control group was comparable to the national averages, strengthening the external validity of this study's findings.

4.8 Qualitative inquiry

Qualitative inquiry is wrapped around the two field experiments reported in this thesis in order to provide a richer account of students' experience of learning at an FE college and how they and their nominated supporter interact with the text messaging intervention. Chapter 3, Section 3.6.3 sets out how qualitative inquiry in general, and in-depth interviews in particular, may be used to gain a deeper understanding of participants' lived experiences of the intervention (p. 69). Particularly since the intervention is mediated by human behaviour - in the sense that nominated supporters need to engage with the text messages for the intervention to ultimately affect student outcomes - qualitative methods help facilitate the interpretation of trial results. In order to gain a rich and triangulated description of the themes of interest, the different viewpoints of tutors and students are considered in turn.

4.8.1 Qualitative research objectives

Table 4.10 lists the qualitative objectives, formulated to help contextualise quantitative findings and gain a better understanding of the process of the intervention.

Table 4.10: Overview of qualitative objectives

Qualitative objective	*Qualitative research questions*
Gather data on adherence to and acceptability of the intervention.	What burden does recruitment for the trial impose on tutors? To what degree is students' openness to the intervention influenced by the way the programme was introduced and delivered throughout the year?
Develop good recruiting and consent practices	What information do tutors and students need to engage with the intervention?
Understand the conditions that need to be satisfied for greater intervention effectiveness.	How does the strength of the existing relationship with the study supporter influence the degree to which texts are discussed? What suggestions for further improvement do students and tutors propose?

Few interventions have been implemented in a further education context (some exceptions are Anderson et al., 2001; Dalby & Noyes, 2015; Swan, 2006), and even fewer using a field experimental approach (e.g. Chande et al., 2017). The qualitative component of the first field experiment was thus characterised by a focus on the feasibility and acceptability of the trial within a further education context. The interview schedule was developed to gather detailed information on the most important factors for successful implementation. The topic guide was modelled on the research questions displayed in Table 4.10.

4.8.2 Site selection and sampling approach

Most purposive sampling strategies require a priori knowledge about the range of variation. In this study, the range of variation of interest (quality and quantity of communication with supporter) was unknown because no survey data was collected on students' actual interactions with their study supporters. The literature on social support suggests that parents and peers affect student motivation in different ways (Meeus, Oosterwegel, & Vollebergh, 2002; Wang & Eccles, 2012), so a balanced spread of relationship types was prioritised. To ensure the information-rich cases were selected, students from the treatment group were oversampled.

A quota sampling strategy was used to ensure variability in key variables such as (1) gender, (2) age, and (3) type of relationship with nominated supporter. The quotas were monitored throughout the recruitment and interview phase. Three colleges were selected to maximise variation in geographic location, subject treated, and implementation effectiveness (see Appendix 18 for college characteristics, p. 261). Within each college, four students were selected to participate in the study. Students were oversampled to allow for dropout or refusal to participate in the interview. The list of selected students was shared with college administrative staff, who approached the students to confirm the interview date and time. Project leads at the selected sites were also approached for an interview.

Forty-two per cent of invited students agreed to participate. None of the invited students at College C agreed to participate and suitable replacements were not found. The administrative staff unfortunately only found replacements who were not involved in the experiment. Tutors at two out of three sites agreed to take part in the interview, and a third tutor who was involved in a pilot study which tested the sign-up survey also took part. Participants provided verbal and written consent at the start of the interview and agreed to being recorded. See Appendix 19 for the consent form (p. 262). Student participants were offered a £10 gift card as a token of appreciation for their time, and teachers were offered a choice of popular education books.

4.8.3 Topic guide

The data were collected through a combination of short self-rating questions and open-ended questions in semi-structured interviews. Interviews with tutors primarily focused on aspects of the process of implementation. The interviewed tutors provided input on the text messages at the start of the year and were also involved at the student recruitment stage. Student interviews focused on understanding the lived experience of nominating and interacting with a study supporter. The student interviews explored if and how the text messages contributed to communication between student and supporter. Interview schedules for student and tutor interviews can be found in Appendix 20 and Appendix 21, respectively (both p. 263). All interviews were recorded and transcribed verbatim.

It is important to be mindful of the ways in which the participants may be influenced by the questions, interview settings, and researcher's behaviour (Maxwell, 2009). Interviewees may be influenced by the interviewer's body language or tone, or they might want to please the researcher by giving 'the right answer'. Naturally, there are no right answers, but it is conceivable that students described their experience of the intervention more positively than they felt in actuality. Care was taken not to disclose that I, the interviewer, designed and delivered the text messages. At the start of the interviews, all interviewees were invited to be open about their positive, neutral or negative experiences.

4.8.4 Sample characteristics

As shown in Table 4.11, a greater proportion of female students agreed to participate. The sample is relatively more varied in terms of ethnicity and both 16-18 year-olds as well as 19+ learners are represented. Unfortunately, the college where English students were treated, Great Yarmouth College, was unable to recruit students. The student sample therefore consists of maths students only.

Selection issues in terms of representativeness in purposive qualitative samples are discussed in Chapter 3. Selection issues are unlikely to have been overcome by the current approach; merely 42% of sampled students agreed to participate. Those who responded to our request may be more likely to be motivated to learn as they had not dropped out of college (the interviews were carried out at the end of the academic year). These limitations preclude us from generalising from the sample to the wider population of FE college students. However, the primary aim of these interviews was not to faithfully describe the student experience at colleges, but rather to gain fresh insights on the experiences of students taking part in the intervention, identify challenges to uptake, and further adapt the intervention into an effective text-message intervention.

Table 4.11: Interview sample characteristics

College	Interviewee type	Group	Study supporter(s) relationship	Gender	Age	Years at college	Ethnicity	Geographical Area	Qual.	Subject	Ofsted rating
College A	Learner	Control	Father and classroom friend	Male	20	4	Arab - British	South East	GCSE	Maths	Good
College A	Learner	Treat	Mother and now-ex-boyfriend	Female	17	1	White - British	South East	GCSE	Maths	Good
College H	Learner	Treat	Father and college tutor	Female	20	2	Arab	Greater London	FS	Maths	Outstanding
College H	Learner	Treat	Father and mother	Female	18	1	Arab	Greater London	GCSE	Maths	Outstanding
College H	Learner	Treat	2x college tutor	Female	18	2	Black/Black British	Greater London	FS	Maths	Outstanding
College H	Tutor	Tutor, involved in trial	-	Male	-	6	White - European	Greater London	GCSE	Maths	Outstanding
College A	Tutor	Course Leader, involved in trial	-	Male	-	1	White - British	South east	GCSE	Maths	Good
Pilot college	Tutor	Course Leader, involved in pilot	-	Female	-	2	Arab - British	Greater London	GCSE and FS	English	Good

4.8.5 Analysis and interpretation

A thematic analysis approach was used to analyse the interview data. This approach allows for an in-depth exploration of respondents' views, motivations and experiences through a systematic coding process and identification of themes and patterns (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2008) A largely inductive approach was used to code the student interviews, where themes that arise inductively from the data are added to the coding framework. The inductive element of the analysis process is important. Few studies have directly examined the way students seek and receive social support within the mandatory resit context, thus little was known a priori about common themes.

The tutor interviews focused primarily on implementation. These interviews were coded using codes developed by the consolidated framework for implementation research (CIFR). This framework was developed by Damschroder et al. (2009) to provide a comprehensive taxonomy of constructs that are likely to influence intervention implementation.[29] Although CIFR was originally developed for a health research context, the framework's programme implementation constructs are applicable to all forms of implementation research (Kirk et al., 2016). The CIFR codebook was used to guide data coding and analysis. Additional text passages that cannot be coded with the existing coding scheme were added to the scheme.

4.8.5.1 The analysis sequence

This thesis applies the following step-wise approach. First, the qualitative research questions and the CIFR coding framework are used to develop the *a priori* coding framework which are then entered into NVivo. Second, the resulting coding framework is applied to three transcripts. Text passages that do not fit with existing codes are identified and added to the coding scheme. After coding of the initial transcripts is completed, the coding scheme is reviewed. Coding categories that overlap are collapsed into higher order categories, and categories that upon inspection contain multiple different topics are given separate codes. The resulting coding scheme is then applied to the remainder of the transcripts. Throughout this process, each new transcript is used to confirm support for existing codes or document counterpoints, in order to critically assess the strength of evidence for each

[29] The full set of constructs can be found online: http://cfirguide.org/constructs.html

code. Once coding is completed, resulting coding categories are compared to raw data to ensure the codes are representative of the data. Subsequently, codes are grouped into themes. Themes are identified through an iterative process of comparing and contrasting similarities and differences. The emerging themes then receive a description of how and when the theme occurs. After coding of all transcripts is completed, the resulting coding framework is rearranged into categories or 'themes'. This data reduction process ensures that only the most relevant text passages are applied to address the research questions.

For the tutor interviews specifically, the CIFR constructs addressing implementation climate were subjected to a rating process.[30] A score between -2 and +2 was assigned to each construct for each of the colleges (see Damschroder et al., 2013, p. 5 for the criteria used to assign ratings to CIFR constructs). If comments are equally positive and negative, a mixed (X) rating is applied. A neutral (0) rating is applied when the construct is addressed but the valence of the data is neutral. Finally, constructs with insufficient data are indicated as 'missing'. The ratings are then compared across sites (i.e. colleges) to identify cross-cutting and college-specific patterns of barriers to successful implementation.

4.8.6 Tutor interviews: a focus on implementation

A matrix with ratings for each CIFR construct for each of the sites can be found in Table 4.12. The qualitative aim of this section is to explore ways recruitment and implementation may be improved in the subsequent experiment, rather than the contextualisation of treatment estimates.

[30] Implementation climate is a sub-category of the CIFR framework. Other constructs, such as 'innovation characteristics' and 'outer settings' are less relevant to this process evaluation because the intervention was implemented by the research team. The participating colleges had no influence over the design of the programme, nor timing, frequency or focus of the text messages.

Table 4.12: Implementation climate at sampled colleges, CIFR construct ratings

Construct	*College A*	*College H*
Tension for change	Missing	Missing
Compatibility	+1	+2
Relative priority	Missing	Missing
Organisational incentives & rewards	-1	-1
Goals & feedback	+1	+2
Learning climate	-1	X
Leadership engagement	-2	+2
Available resources	-1	+1
Access to knowledge and information	-2	-1
Total	-5	5

Several findings of interest emerge from the CIFR coding exercise. First, it was clear that both tutors regarded the Study Supporter intervention as an externally developed programme they merely implemented because they were instructed to. Neither mentioned a need for innovation or change. Ways to involve college staff more closely at planning stage are discussed below. Second, of the nine CIFR constructs assessed, two constructs strongly distinguished between college sites. Two constructs were not addressed, and five were rated similarly across the two sites. Since these ratings are based only on two out of nine colleges that participated in the field experiment, the discussion below focuses on exploring the practices that contributed to positive ratings and those that had a negative influence on the implementation climate. These findings are then used to craft recommendations for implementation of future interventions in further education colleges.

4.8.6.1 Leadership engagement

Leadership engagement was a strongly distinguishing construct. A positive organisational climate and the availability of administrative support is essential to successful implementation, as well as the perception that senior leadership values the project (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). If the leadership team is not knowledgeable about the research project, not able to articulate the importance of implementation, or not proactive in problem solving, successful implementation is inhibited (Lyon et al., 2018).

In college H, senior leadership connected the importance of the research project to the strategic goals of the college. Additionally, they provided the tutors with the necessary administrative support. For example, the booking of computer classrooms for the sign-up survey was not left to the individual tutors, as was the case in college A, but was organised centrally. Tutors received clear instructions and were "reminded to take it very seriously" *[Tutor, college H, Maths].*

In contrast, senior leadership was scarcely involved in the other college. The interviewed tutor had a particularly strong negative view of the way the intervention was communicated to staff. He said that the project "was sort of very much dropped on me and the team unexpectedly and very short notice." This college was recruited relatively late, and as a result tutors were briefed at the start of the academic year rather than over the relatively quiet summer period. This lack of planning time resulted in poor uptake by the tutors. Low initial engagement persisted throughout the academic year. Building positive relationships before implementation appears to be crucial; especially because tutor feedback is enlisted at regular intervals. In this case, a false start resulted in limited engagement with the text-message writing process:

> "We could probably do better because there was, you know, a sort of bad feel to the start of the programme. It wasn't the sort of great buy in and enthusiasm, or could really come up with an inspiring message for this particular week. [...] So a better introduction would improve the quality of message." *[Tutor, college A, Maths]*

4.8.6.2 Access to knowledge and information

Both sites scored relatively low on the CIFR construct 'access to knowledge and information'. The training for college staff was limited to a one-hour training session delivered by the research team.[31] This session may not have prepared tutors well enough for implementation. It is therefore recommended that research teams develop additional resources for tutors they can take away after the introductory meeting. The data indicated there was a tension between on the one hand receiving limited information about

[31] The team consisted of six members the Adult Skills and Knowledge (ASK) team. I presented to tutors at 8 colleges, before the trial launched.

the project, and on the other hand, having to introduce it to students in a convincing manner. One tutor reflected on the lack of information, saying "Students want to know what it's all about and we couldn't say anything because we didn't know anything, so it's just a mystery" *[tutor, maths, college H].* Tutors were not informed of the study hypotheses, primarily to avoid bias. If tutors were aware of our hypotheses, they may have encouraged specific individuals to sign up.

On reflection, the decision not to disclose relevant background information was a miscalculation. Tutors were frustrated by the lack of information and felt this hindered their communication with students. Although tutors received an information booklet which included information about choosing study supporters, they felt at times unable to guide their students in the decision to sign up and nominate a suitable study supporter.

4.8.6.3 Available resources

The second distinguishing construct was the availability of resources. Although the resources required for implementation were limited, namely access to a computer, implementation led to disruptions to the classroom. In college A, all tutors booked computer classrooms and shepherded their students to these rooms which could take up to 30 – 45 minutes of class time. When reflecting on the resources required for implementation, the interviewed tutor said that "the way it went this year… it's just been a burden rather than an asset". In college H, on the other hand, portable laptops were brought into each classroom and students could easily log on and complete the short survey. Tutors did not perceive implementation to be intrusive.

A recent study found that teachers' willingness and ability to implement new interventions or programmes is strongly influenced by their work load and work-related stress (Larson, Cook, Fiat, & Lyon, 2018). All interviewed tutors expressed a degree of work-related stress, due to the strict schedule of exams and assessments, and limited class instruction time (often no more than 2 hours per week). Tutors lamented that they had to teach the two-year curriculum of GCSEs within less than 40 weeks, often to students with heightened academic support- and socio-emotional needs. For example, when asked whether her work environment is conducive to implementation of interventions, the Head of English at one of the participating colleges said "it's just… we have other pressures. You know, meeting deadlines and controlled assessments, and making sure that students [...] have learnt what they need to learn before we progress." *[Tutor, pilot college].*

4.8.6.4 Suggestions for improvement

The interviewed tutors could not remember discussing the strategic goals with their teams, did not receive feedback about implementation, and did not know what the purpose of the intervention was. The interview data therefore suggests that strong organisational structures may not have been in place to support the implementation of the programme. Additionally, organisational incentives for successful implementation were not in place. Indeed, the compressed teaching schedule and frequent assessments allowed little room for non-teaching activities. The senior leadership team could reward successful implementation more. The research team could facilitate this by sharing frequent progress updates. Additional, small but meaningful incentives could be awarded to tutors who introduced the project to the highest proportion of their classes, or who championed text-message writing ideas.

Tutors requested a longer lead-in period to implement the study and suggested it would be beneficial if student were given more time to source supporters. It was suggested that implementation in the first few weeks of the academic year is too hectic. If the intervention were introduced at the beginning of October, students might have overcome initial worries and be more comfortable asking for the support of a third party. Tutors suggested the student recruitment could take place in tutorial lessons rather than maths or English class, for logistical ease (i.e. computers are often already available in tutorial lessons).

It was also felt that the student sign-up survey was short and did not contain enough background information about the aims of the project. Future iterations of the intervention should better address questions such as 'Who is a good study supporter?', 'How often will they receive texts?' or 'Will they know if I skipped class?'. These questions remained unanswered in the current iteration and led to confusion and disengagement at the point of sign-up. More information provision for tutors may address this issue in future iterations.

Additionally, it was suggested that students should be given more time to think about their study supporter choice. If the project could be introduced in week 1 and ask students to complete the sign-up survey in week 3 or 4, they might have been able to dedicate more effort to choose a suitable individual. This would allow students to check with these individuals if they would be happy to be their study supporter before signing up. Related to this point was the tutors' concern that students selected suboptimal study supporters. Because students had not been given time to think about their choice, it appears that a number of students simply nominated the person sitting next to them at the time. It was also suggested that the nominated supporter should have sufficient background knowledge of maths and English. Future iterations should include more guidance as to what a 'good study supporter' might look like and allow more time for the sign-up process.

Finally, concerns about data security were brought up by tutors. Within the colleges we have online survey data for, on average 35% of students opted in to take part. Tutors suggested that students' reluctance to sign up may be due to lack of trustworthiness of the sign-up survey:

> "They are very protective of those people that they know who care about them or have those relationships with, and they didn't want to give telephone numbers. [...] Again though, I think it may be to do with the way it's delivered." *[Tutor, English, pilot college]*

The transcripts did not provide clear suggestions as to why students might be protective of these details. Since the student interviewee sample consisted of those who opted in, future qualitative studies should sample students who decided not to opt-in in order to better understand perceived barriers to sharing contact details.

4.8.7 The learning experience

The student interviews were coded primarily inductively and were not guided by the CIFR constructs introduced in the previous section. The interview questions focused on students' perceptions of the college environment and their experience of nominating and communicating with study supporters. Several themes emerged from the thematic analysis. First and foremost, the interview data underlines the importance of creating a positive learning environment.

4.8.7.1 Fragile learner identities

Several students had previously negative school experiences (e.g. bullying, lack of belonging and failed academics), suggesting an increased need for emotional support in college. Three out of five interviewees discussed having negative experiences with learning, while none reported positive learning experiences in secondary school. A maths tutor described this general sense of disengagement with learning:

> "One of the students went up and said "you know, we have been sort of trampled on throughout our GCSE. [...] I don't know if we will pass sir, but I think I have a chance of passing." I can imagine them at school and from year seven, they haven't progressed and that's sort of how they felt. They felt sort of ignored in a way. I don't really know how to say it and say, well... "If they don't get a C now, they'll get a C at the college I guess" (referring to KS4 teachers, 'giving up' on students too far away from the C-threshold). *[Tutor, college A, maths]*

Self-determination theory proposes that optimal learning can only occur when students' basic psychological needs are satisfied (Deci & Ryan, 2000; Ryan & Deci, 2000). The above excerpt illustrates that teacher support of student competence is crucial for autonomous motivation. People experience competence when they feel they can achieve a positive outcome and master environmental challenges (Reeve, 2012). By definition, the students in our sample did not achieve the C-threshold at age 16, when they first sat the GCSE exams. Research shows that students' subjective experiences of 'being labelled a failure' in maths and English GCSEs can result in a lack of motivation and "expectations of continuing failure" (Higton et al., 2017, p. 32). Commissioned by the Department for Education, Higton and colleagues (2017) completed in-depth fieldwork in 45 colleges to explore effective practices in the delivery of English and maths to 16-18 year old. Across the board, they found that the motivations of students to study these subjects are strongly affected by their prior learning experiences and examination outcomes (Higton et al., 2017).

Cultivating more positive attitudes to learning and bolstering self-esteem (i.e. psychological need satisfaction) is a central feature of further education teaching. The lack of autonomy to choose whether or not to continue taking maths and English also emerged as an important theme across the student and tutor interviews. The need for autonomy is satisfied by having opportunities to make choices, and understanding why the less interesting activities are meaningful and useful (Guay, Ratelle, & Chanal, 2008). Especially now that students who obtained a D grade at age 16 are required to repeat their GCSE(s), reluctance to re-engage with the subject is a common theme:

> Their interest in maths and the reason for doing maths is because they have to [...] there's a few that wanted to do it. [...] At the start of the year it's very much "I don't want to be here anyway, so I'm not really worried about now you know getting a mentor (i.e. study supporter) to talk about maths when I don't even want to do maths." *[Tutor, college A, maths]*

The above passage also emphasises an important potential barrier to student engagement with the Study Supporter intervention. If the most disengaged students are least likely to sign up to take part, those who need it most are left out. Future intervention studies may benefit from incorporating incentives and interactive sign-up materials to encourage less intrinsically motivated students to take part.

4.8.7.2 Student experiences of support

Out of the five interviewed students, three students could recall signing up. Their responses suggest that they were more attentive to programme purpose than their tutors thought. For example, they theorised that the programme was developed "to encourage people to get more involved with maths and English", "to encourage maths outside the classroom", or "have support because not many teenagers [...] really speak about their problems that much". Those who remembered nominating a study supporter (N = 3; 60%) were positive about the programme. For example, one student said:

> "It was helpful.. [...] I don't think I would have got through the year without having someone to support because she'd get a text about if I had an exam. She'd be like you've got your exam and revise it and like bring it up which was good because I used to didn't talk about my exams." *[05LE03, Female, 17-year old, nominated mother, Maths GCSE]*

The data suggests that successful study supporters tended to be people whom the student had a close relationship with; those who actively asked questions and gave emotional rather than practical support.

4.8.7.3 Overt and covert interactions with the intervention

The text messages were not an overt component of the supportive process. Two out of three interviewed students in the treatment group referred to actual text messages:

> They [parents, both nominated] didn't even know if I had exams, or if I had a control system. They didn't know it until the text messages started. *[04LE01, Female, 20-years old, nominated parents, Maths GCSE]*

Only one of the interviewed students who remembered signing up felt their relationship with the nominated supporters had changed as a result of the frequent text messages. She suggested her relationship with her mother (i.e. nominated study supporter) "became closer" because she previously "didn't actually speak about my work". Taken together, the interviewed students unanimously felt the existing relationship was strong. Overall, the programme of texts may not have increased the frequency of communication, instead, the conversations may have become more informed as a function of the texts:

> "I had a controlled assessment that I didn't even know [about]. So, they get their text message saying that "your daughter has a controlled assessment, please help her". So they always make time for me, like, you have a control system, sit down and do your work." *[04LE01, Female, 20-years old, nominated parents, Maths GCSE]*

The students whose supporters were already actively engaged in their college life reported that the text messages had a positive impact relationally and academically. Drawing on this data, it appears that the study supporter intervention effectively reinforces support. However, where a supportive network is less strong, it is unclear whether the text messages help to foster a supportive relationship in the same way.

4.8.7.4 Future iterations

Two students did not remember the sign-up survey. Both mentioned that they had to complete several surveys at the start of the year. The sign-up survey should be memorable and visual and provide clear information. In this project, an effort was made to keep the sign-up materials as short as possible in order to make it a light-touch experience for tutors and students. In hindsight, the introduction to the project should have been more detailed. Two out of five students mentioned that they were not aware their nominated supporters would receive text messages throughout the academic year, which suggests tutors did not introduce the project in sufficient detail. Since tutors were informed that the project would revolve around weekly text messages, it is surprising that some students were not aware.

The next chapter describes a follow-up study to the Study Supporter intervention. Its design and implementation was informed by the interviews described here. A number of changes were made to the sign-up process and tutor on-boarding, as the interviews made clear that both students and tutors were insufficiently informed of the programme aims. The specific changes made to the intervention protocol are further discussed in Chapter 5, section 5.2.3 (p. 145).

4.9 Discussion

This chapter contributes to the thesis by addressing research question 1. The supportive information intervention set out in this chapter aims to spark conversations between students and their key relationships. This study adds to the growing number of information interventions showing that informative text messages can be a light-touch but effective way to harness individuals' existing social networks to take a greater interest in their learning which is in turn hypothesised to improve student outcomes.

The analysis finds that simply informing (close) relationships of students of topics studied in class, reminding them about upcoming deadlines, and encouraging them to reach out to the student improves attendance rates by 3.1 % points and achievement rates by 6 % points. The heterogeneous analyses are exploratory in nature and resulting estimates of treatment effects warrant independent replication. Subgroup analyses show that the average treatment effects on attendance were primarily driven by young and male students. The poorest attenders, namely 16 to 18-year-olds and male students, are more positively affected ($p < 0.05$) than older learners (19+) and female students where the treatment effect was indistinguishable from zero. This study does not have the statistical power to distinguish whether students on Functional Skills or GCSE courses benefited more.

The attainment results paint a somewhat different picture. Treated female students are significantly more likely to achieve their qualification at the end of the academic year than female students assigned to the control condition ($p < 0.05$). Whereas the point estimates for male students was similar in size, they do not reach statistical significance. Again, the effects are driven by young rather than adult learners. Additionally, logistic regression models were fitted, and these estimates were consistent with the primary findings.

4.9.1 A closer look at information intervention effect sizes

The treatment estimates on attendance and achievement correspond to effect sizes of Hedges' $g = 0.11$ and $g = 0.14$, respectively. To put these effect sizes into perspective, I now turn to a brief review of effect sizes found in similar randomised controlled trials that evaluated the potential of supportive and informative text messages.

The effect sizes of the intervention on attendance are comparable to or larger than those found in similar interventions. For example, Chande et al. (2017) who sent weekly text messages to further education college learners, find a 6.6 % point increase in attendance from a control mean of 42.1%, which corresponds to $g = 0.06$. Kraft and Rogers (2015) tested a similar intervention where teacher sent weekly informative text messages to parents. They found that treated students were 2.5 % points less likely to be absent than their control peers ($M_{control} = 12\%$) during a four-week summer course.[32] Finally, Rogers et al. (2017) sent personalised postcards about the importance of attendance to over 50,000 students' homes and found a 2.4 per cent reduction in student absences ($M_{control} = 3.45$, $M_{treat} = 3.29$), which corresponds to $g = 0.02$.

The effect size of the intervention on attainment was more pronounced than that on class attendance, which is promising. Alternative pathways of the intervention to student success, through homework completion, improved confidence or better exam preparation, could be alternative mechanisms

[32] Kraft and Rogers (2015) did not include sufficient information for the calculation of Hedges g.

through which the intervention improves student outcomes. It may also be true that the conversations students and study supporters engaged in helped students do better in their exams but were not as effective at motivating them to attend class more often. The effect size on achievement is comparable to or larger than other reported effect sizes of text messaging interventions on achievement. For example, Berlinski et al. (2016) find a positive impact of 2.8% points on the proportion of students achieving the cut-off for passing the subject, and Bergman and Rogers (2017) find a 10 per cent reduction in the number of courses failed for treated students ($M_{control}$ = 2.43, M_{treat} = 2.20). The 6 %-point increase in pass rates found in this chapter falls slightly short of Chande et al.'s (2017) findings, where being assigned to the treatment group improves pass rates by 8.7 % points.

4.9.2 Cost-effectiveness of the intervention

It may be useful also to compare the effect sizes of this light-touch intervention to effects of more intensive and costly interventions to improve student attendance and attainment. The Accelerated Study in Associate Programs, developed by City University of New York (CUNY) provides selected community college students[33] with tutoring, weekly seminars, free use of text books, employment support, mandatory advising sessions, and free tuition. Using a propensity score matching approach, the researchers found that 55% of students on the programme graduated within three years of enrolment, whereas only 26% of comparison students had done so (Kolenovic, Linderman, & Karp, 2013). One-year retention in college, a more comparable statistic to the present study, was on average 12% higher for programme participants than comparison-group students. Although the programme was found to be highly effective, its cost of approximately $14,000 per student is likely a barrier to widespread implementation.

Especially in the context of year-on-year funding cuts to the further education sector (Hupkau et al., 2016), such a costly programme is not likely to receive the necessary funding. In contrast, the supportive intervention

[33] Community colleges in the United States are comparable to Further Education colleges in England, both in terms of student demographics and courses offered. Community colleges primarily offer lower-level tertiary degrees and similar to further education colleges, may serve as the stepping stone towards a bachelor's degree at university.

described in this chapter cost less than £5 in text message credits for the full academic year. All research tasks, including development, delivery of the intervention, and data collection added up to 101 days of researcher time, which amounts to approximately £45 per student. It should be noted that if the programme were to be implemented by college staff, associated costs would be reduced considerably. Additionally, the above calculation includes funding for research staff, which amounted to the majority of incurred costs.

4.9.3 Exploring interaction between students and their study supporters

The qualitative evidence presented in this chapter indicates that students experienced the intervention positively, but that the existing strength of the relationship was key. Student engagement with the sign-up exercise was a key prerequisite for selecting an appropriate study supporter. Interestingly, students spoke about the benefits of having a study supporter in global terms, rather than pin-pointing specific conversations about the text messages.

Few noted the impact of the messages themselves, suggesting that support transactions may have been implicit rather than overt. In a study of couples experiencing acute stress, Shrout, Herman and Bolger (2006), find that the optimal support patterns is characterised by 'invisible support' where "the supporter provides assistance without making the provision obvious to the recipient" (p. 131). Noting the difference in context, this idea that support can come at a cost if it is provided too overtly is was first introduced in Chapter 2 (see section 2.3.3, p. 30). Additionally, a number of suggestions for improving the intervention design and implementation were offered. Increasing tutor-buy in through timely on-boarding emerged as a key component, as well as the need to help students think through their study supporter choice carefully. Chapter 5 starts with a discussion of the iterated study design.

4.9.4 Limitations and future research

Although limitations to the research design are discussed in more depth in Chapter 7, I briefly touch upon a number of study-specific limitations here. The findings presented in this chapter show that, for students who are willing to opt in, texting friends and family can be a powerful way to increase attendance and achievement. However, the current intervention design does not allow us to disentangle whether it was driven by an increase in perceived social support on the student's part, or an increase in monitoring behaviour on the supporters' part, or both.

We may see a positive effect of the intervention because the study supporters simply passed on the information contained in the messages, rather than actively engaging in supportive behaviours such as helping with assignments or providing emotional support. Given the policy goals to improve attendance and achievement in maths and English courses at FE Colleges, it is essential to better understand the mechanisms through which personalised messages help leverage students' social support networks to improve academic motivation.

This evaluation of the supportive text messaging intervention was also limited by the lack of baseline information on student attendance and achievement. Additionally, as all outcome data was collected routinely by colleges, it necessarily was limited to college behaviours. It would be interesting to explore how nominated study supporters perceive the intervention and to what degree they interact with its content. Future studies may benefit from surveying students and study supporters to gather information on the quantity and quality of communications.

4.10 Conclusions

This chapter provides evidence in support of research question 1, which asks if supportive text messages can result in improved learning outcomes. Short weekly text messages to study supporters, touching upon a variety of topics such as revision resources, upcoming deadlines and lesson content, improved attendance rates by 3.1 % points (Hedges g = 0.11) and achievement rates by 6 % points (Hedges g = 0.16). Notably, the effects are stronger on final attainment than class attendance.

These results provide evidence that enlisting support from outside the classroom can help improve student success. Additionally, these primary findings are in line with the literature discussed in Chapter 2, which sets out the importance of positive communication with close relationships. The literature suggests that supportive communication and encouragement from family and peers may help improve recipients' sense of self-efficacy, self-confidence, positive mood, and coping strategies (Collins & Feeney, 2004; Taylor, 2011). Social support may also benefit recipients by strengthening the relationship between provider and recipient (Burleson, 2003).

This chapter does not examine these potential mechanisms experimentally due to limitations of the research design, but the qualitative interview data suggests that encouragement and support from study supporters helped students to feel more confident in their ability to succeed in their maths or English studies. Further, this psychological benefit of the intervention may have primarily occurred through an increase in perceived support rather than enacted (i.e. received) support. Relatively few examples of actual support were recounted, in contrast to more frequent accounts of generally being able to rely on the study supporter. This notion is supported by the literature on perceived and received support, which consistently shows more beneficial outcomes for individuals who report high perceived support but mixed outcomes for those who report high received support (Gottlieb & Bergen, 2010; Thoits, 1995).

It should be noted that the design of the field experiment and schedule of text messages were informed by the literatures on social support and supportive communication, but that it was not feasible to explore contextual factors that determine the effectiveness of social support experimentally. Such studies require systematic manipulation of characteristics of the supporter-recipient interaction, including gender of supporter, relationship closeness, type of support provided, and stressfulness of the situation. These characteristics are then varied systematically in scenarios read by participants, but this does not translate easily to real relationships(see for example Cutrona et al., 1990). In this chapter, students nominated existing ties who may or may not have been proficient support providers. Additionally, as study supporter choice is endogenous I was unable to explore if parents, peers or other types of social connections provide more effective support. Future studies may address this issue by introducing exogenous variation in study supporter choice, although it should be noted that this may reduce the external validity of the experiment.

In relation to the literature on information interventions, this chapter provides additional support for the idea that light-touch and practical text messages can help instigate greater involvement from students' social networks. A growing number of empirical studies assess the importance of parent-teacher and parent-child communication, and most studies do indeed find that informing parents of their child's behaviour significantly reduces absenteeism (Robinson & Lee, 2017; Smythe-Leistico & Page, 2018), improves work habits (Bergman, 2015) and increases achievement rates (Kraft & Rogers, 2015; Rogers & Feller, 2018).

Finally, the results showed that the development and testing of dyadic interventions may be a promising avenue for further research. This study adds to the literature by providing further support for the powerful role of friends and family in young people's lives. Most social support interventions focus on introducing new ties, such as mentors or peers with similar experiences (see for example May & West, 2000). Albeit new relationships may be formed with mentors or similar peers, it may be especially promising to target existing relationships for sustainable behaviour change. After all, introducing a habit of positive communication and a social norm of the importance of maths and English may have self-reinforcing effects beyond the trial period. After all, a supportive home environment is created and maintained by several actors; and involving them in a targeted intervention can help empower them to support students' learning process where they might otherwise have lacked information or resources.

5 FIELD EXPERIMENT 2 - PROJECT SUCCESS[34]

5.1 Introduction

Results from the Study Supporter intervention introduced in the previous chapter indicate that a programme of supportive text messages sent to influential third parties, such as parents, other relatives or peers, can improve attendance and achievement at further education institutions. Subgroup analyses further showed that the treatment effects were driven particularly by the younger cohort of students, aged 16-to-18. Gender-differences were apparent too, where the improvements in attendance were driven by male students but the increases in qualification achievement rates were driven primarily by female students.

5.1.1 Exploring heterogenous treatment effects

The moderating effect of gender has not been explored within the context of supportive text-messaging interventions. The topic has received attention in the literatures on social support and supportive communication, however. Scholars have suggested that giving and receiving social support is moderated by gender and gender roles (Barbee et al., 1993; Song et al., 2015). A meta-analyses of fifty studies on coping behaviour and gender differences found that women are consistently more likely to actively seek social support (Tamres, Janicki, & Helgeson, 2002). According to this study, women do not only seek emotional support more often but also seek specific and concrete help from family and friends (Tamres et al., 2002). Gender role expectations may play a significant role. For example, Barbee et al. (1993) propose that the male role, which traditionally emphasises "success and emotional inexpressiveness" (p. 179) may make it more difficult for men to seek support. If soliciting help from others is a more natural coping strategy for

[34] This intervention was conducted as part of the Behavioural Research Centre for Adult Skills and Knowledge (ASK) and was funded by the Department for Innovation, Business and Skills. I developed the intervention, led on trial design and implementation, conducted the analyses presented here, and wrote this chapter in its entirety. Todd Rogers and his team at the Student Social Support Lab were involved in initial conversations about intervention design.

women than men, it could be hypothesised that the supportive text messaging intervention breaks down the barriers men may experience. After all, the weekly text encourages the study supporter to enquire about their learning and offer support. In extension, this leads to the hypothesis that the intervention might benefit male students especially.

On the other hand, researchers have argued that adolescent girls are more sensitive to social relationships and benefit more from social support than adolescent boys (Helsen, Vollebergh, & Meeus, 2000). It could therefore also be hypothesised that the women in our experimental sample benefit most, because they are more sensitive to social relationships and receiving support. A recent meta-analysis found that social support is a critical resource for both genders in adolescence (Rueger, Malecki, Pyun, Aycock, & Coyle, 2016). This chapter therefore further explores whether gender moderates the effectiveness of the intervention.

5.1.2 Exploring mechanisms

Although the study introduced in Chapter 4 showed that a low-cost and low-touch intervention can meaningfully improve student outcomes, its design prevented us from exploring its potential mechanisms. For example, it is unclear whether the informational content of the text messages benefited students, or perhaps that their nominated study supporter knew more about their learning and could therefore provide more skilful and targeted support. Study supporters could simply relay factual information or forward the text messages to the student. On the other hand, study supporters could engage in a more detailed conversation with the students that nominated them and build a habit of supportive communication over time. This chapter seeks to address this question of whether the informational versus interactive elements of the intervention produce treatment effects, through the addition of two trial arms. First, rather than communicating with a study supporter only, a copy of the text message is sent to the student directly. The second variation removes the study supporter from the equation entirely as the text messages were directed at the student only. This direct text message informs students about upcoming exams and course content, and does not activate students' social networks in the process.

Aside from the addition of two trial arms, the intervention design is similar to the Study Supporter intervention in Chapter 4, see section 4.3.6 for a detailed exposition of the intervention content (p. 96). Informative and positively-worded weekly text messages are sent to nominated study supporters. The texts suggested discussing a topic the student has recently learnt about, or encouraged the student to do a certain task, such as preparing for a mock test or attending a class. Students were individually randomised to one of the following arms: (1) control, where no one receives text messages; (2) study supporter only receives text messages; (3) student only receives text messages; and (4) both study supporter and learner receive the messages. This trial design allows us to test whether the text messages help improve learner success through direct college communication, or whether having a supportive individual who receives updates about college is beneficial over and above learners receiving this information directly. The experiment was carried out in further education colleges, with students on GCSE maths and English courses.

5.1.3 Similar interventions

Similar parent-school communication interventions were reviewed in detail in Chapter 2 (Table 2.1, p. 40). The commonalities between the intervention set out in this thesis and other text messaging interventions (e.g. Bergman & Chan, 2017; Berlinski, Busso, Dinkelman, & Martinez, 2016; Kraft & Rogers, 2015; Smythe-Leistico & Page, 2018) are manifold. First, the interventions aim to communicate relevant information about the student's learning in a timely and tailored fashion. Second, the text messages implicitly communicate the importance of attendance, homework completion and exam preparation. Finally, these texts are typically sent to parents or guardians, who care about their child's progress and may wish to get more involved in their learning. The intervention presented in this thesis gives the student more autonomy over the choice of the text-message recipient, recognising that a parent may not unanimously be the most supportive adult in the lives of post-16 students.

It was hypothesised that having a study supporter may be beneficial to students' well-being and motivation to learn through a sense of belonging and companionship (Thoits, 2011). Alternative pathways are explored in this chapter. For example, the intervention may benefit students primarily through its informational content. Indeed, many behaviour change interventions are delivered directly to the intended recipient, rather than a third party. A brief overview of such directly-delivered information interventions follows in Table 5.1. These studies all use text messages to deliver personalised and relevant content to students with the aim to improve persistence and attainment in college.

Table 5.1: Overview of similar interventions, direct text-message communication

Paper and intervention	N	Results
Castleman & Page (2016). College freshmen received a series of 12 texts with information about financial aid resources, remind them of aid-related deadlines, and signpost support.	808	First-year community college students who received the texts were 12 % points more likely to persist in their second year in comparison to students in control.
Castleman & Meyer (2016). Text messages provided lower-income college students with encouragement, simplified information about campus resources, and reminders of aid-related deadline.	1,198	No random assignment. Students who received the text messages completed more courses than their non-treated peers, but this did not translate into improvements in grades.
Chande et al., (2017). Further education college learners on maths and English courses received weekly supportive texts, including planning prompts, texts to foster a growth mindset, and feel belonging with college.	1632	The text messages led to an 8.7%-point increase in qualification pass rates in comparison to a control group, as well as a 7.3%-point increase in attendance for the full academic year.
Oreopoulos & Petronijevic (2017). University students received several texts per week about study preparation, information about available resources, and general motivation and encouragement.	3844	The intervention did not result in positive changes in student grades.

The first two studies listed in Table 5.1 primarily focus on nudging students to complete a one-time action, such as completing a financial aid form (Castleman & Meyer, 2016; Castleman & Page, 2016). There is good evidence that text messages can help break down the complexity of a task (Bird, Castleman, Goodman, & Lamberton, 2017), helping people to complete it successfully. The latter two studies, in contrast, delivered a text-messaging programme designed to support student persistence throughout the academic year (Chande et al., 2017; Oreopoulos & Petronijevic, 2017).

Within the field of education, few text messaging interventions have begun to focus on altering behaviour over a prolonged period of time. Chande et al. (2017) were among the first to test a full-year text-messaging programme to increase student persistence at college. Their results are encouraging, especially because they text messages were general in nature. The texts were not personalised to individual students but instead referred to the general student experience. Additionally, they were written by an external research team, and implemented remotely using an online texting platform. These features make Chande et al.'s (2017) intervention more scalable and light-touch to implement than the study by Oreopoulos and Petronijevic (2017) who encouraged students to reply to the tailored texts sent by a team within the participating university. For example, students were encouraged to initiate contact about a topic of their choice, mimicking student-coach interaction. This type of intervention may predominantly benefit engaged students rather than those who are less pro-actively seeking out college support.

All four studies discussed above have in common that they sent frequent text messages to students directly, instead of disseminating the information to third parties. The interventions were also all targeted at young adults at post-16 education institutions. It is encouraging that Castleman and Page (2016) and Chande et al. (2017) have both found positive results with further education college students. If the informational value of the texts has greater benefit to student motivation than the potential support of a third party, direct text messages may prove to be more effective than messages communicated via a study supporter. After all, these texts are delivered to the intended recipient without delay, and the recipient can choose to re-read the messages at their leisure. This chapter will test the relative effectiveness of direct communication with students, versus communication through someone in their social network whom is prompted to play a support role.

5.2 Experimental design

5.2.1 Sample representativeness

This chapter reports the follow-up of the Study Supporter intervention set out in Chapter 4. Colleges within easy traveling distance were prioritised, and they were recruited through an email to senior college staff I had an existing relationship with. The sample is less representative of further education colleges in England than those in Chapter 4, which is a limitation of this study. As can be seen from Table 5.2, participating colleges had positive Ofsted ratings and were primarily London-based. The colleges were also medium- to large-sized. Qualification achievement rates at baseline (2015/16 results) were comparable to national averages, although college C scored exceptionally well on the qualification achievement metrics. The final column reports attendance rates during the first eight weeks of the academic term, before trial launch.

5.2.2 Ethical approval

Ethical approval was gained via UCL Institute of Education student research ethics procedures on September 23rd, 2016, and the trial was also registered on AEA online registry (AEARCTR-0001644) before randomisation.[35] The consent procedures are identical between Chapter 4 and 5. Ethical concerns relevant to the trial design were previously discussed in section 4.3.2, p.88.

This study adds two treatment arms to the experimental design but continues to compare outcomes across treatment arms against a control arm where study supporters do not receive text messages. A pure control group was maintained because any single intervention study does not provide sufficient proof of a true effect. Furthermore, small-scale proof-of-concept studies such as the study presented in Chapter 4 require multiple iterations of experimentation (Banerjee et al., 2017). Therefore, a pure control condition is warranted. It should also be noted that only 25% of consenting students are assigned to control due to the greater number of treatment arms.

[35] See online trial registration here: https://www.socialscienceregistry.org/trials/1644

Table 5.2: Descriptive statistics for sample colleges

College	Students		A*-C (%)		Progress score		Ofsted	Region	Eligible students
	MA	EN	MA	EN	MA	EN			N
A	360	440	22.9	18.1	-0.33	-0.23	Good	South East	304
B	630	720	19.2	17.2	-0.27	-0.35	Good	Greater London	800
C	380	440	31.9	32.1	-0.15	-0.30	Outstanding	Greater London	750
D	1060	1180	20.6	17.5	-0.22	-0.17	Outstanding	Greater London	1987
Sample Avg.	608	695	23.7	21.2	-0.24	-0.26	-	-	-
England Avg.	558	640	23.4	23.8	-0.29	-0.27	-	-	

Notes: National achievement rates, for 2015/16 calculated by the Skills Funding Agency (SFA) are reported here (Department for Education, 2017). Progress scores are calculated by comparing prior attainment scores (at the end of KS4) against attainment at the end of 16-18 studies. A negative progress score indicates that students lowered their point scores at the end of 16-18 phase of education. All prior attainment and progress scores only include students who are at the end of their 16-18 studies and who achieved a C or lower in the subject at age 16. Reported progress scores are from the 2015/16 academic year. Data retrieved from https://www.compare-school-performance.service.gov.uk/ on February 5th, 2018.

5.2.3 Intervention refinement through qualitative inquiry

The iterative refinement of the intervention implementation and content was guided by the expert opinion of tutors and students. The fieldwork conducted at the end of the intervention phase of the Study Supporter intervention informed the further development of the text messaging programme. In the qualitative component of Chapter 4, four students and three tutors were asked to describe their experience signing up for and participating in the intervention (see Section 4.8, p. 116). Students and tutors provided a number of suggestions to improve take-up of the intervention, listed below:

1. Introduce the project as a college-wide initiative. The intervention should be integrated into the maths or English lesson, rather than introducing it as 'something extra' [tutor suggestion].
2. Allow students to complete the survey in their own time, rather than during class time. Alternatively, it was suggested that students could complete the survey in tutorial lessons (e.g. personal development and study skills lessons) or induction week rather than the maths/English lessons since tutors felt uneasy about giving up teaching time [tutor suggestion].
3. Personalise the survey further, by including examples of students who nominated study supporters before and how benefited from it. Additionally, it was suggested that the survey could be more interactive. [student suggestion].
4. Ensure students think carefully about a suitable study supporter. Related to point 3, the sign-up survey should guide students through an exercise that explains what a good study supporter looks like [tutor suggestion].
5. Allow students to indicate their preferences for mode of communication. A student suggested some study supporters may appreciate letters or emails better than text messages [student suggestion].
6. Open the sign-up survey for two to four weeks, so (1) tutors have ample time to introduce the project in their classrooms and (2) students can overthink their choice before having to complete the survey [tutor suggestion].

7. Share more resources with study supporters. The texts could include more frequent links to the online learning environment, helpful instruction videos, or other resources used in class [tutor suggestion].
8. Finally, students felt they would like to be able to change their study supporter throughout the year when relationship with the nominated study supporter(s) breaks down. Students felt they would not know who would "continue to be there" and should have the choice to change who receives texts throughout the academic year [student suggestion].

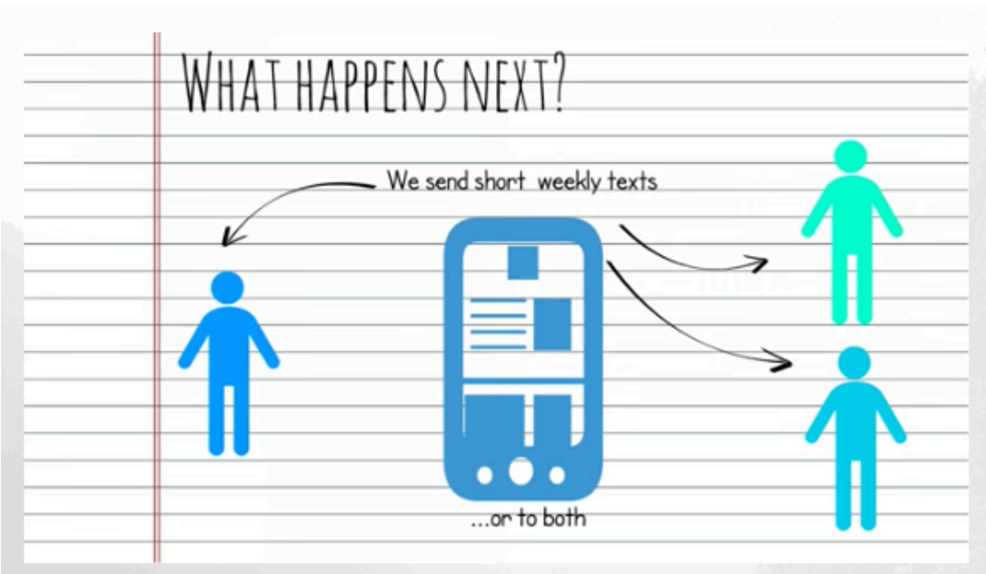
Three of the above suggestions were deemed infeasible. First, allowing students to complete the sign-up survey outside the classroom could lead to a drastically lowered opt-in rate. It was felt that tutor guidance throughout the sign-up procedure would benefit students' study supporter choice. It was decided that tutors should continue to introduce the project in their own classrooms and guide their students through the sign-up exercises. Tutors were provided with detailed instruction sheets, as well as a script. Second, preferences for email, letter or text communication were not solicited due to increased administrative costs of administering communications via two separate platforms. A few years ago, 82.1% of US survey respondents indicate they opened and read every SMS (Anderson, 2015) and the primary means of accessing the internet (and therefore reading emails) is now the smartphone for many people (Anderson, 2015). Finally, changing study supporters throughout the year would complicate the scheduling of text messages and complicate analysis of delivery statistics. It was decided that students and supporters could opt-out of receiving the SMS at any point throughout the academic year, but that they could not request other changes. The remaining suggestions informed the development of the intervention content and its recruitment phase, as set out below.

5.2.4 Recruitment

Tutors were briefed about the project in the weeks leading up to the academic year and were invited to share their feedback on the intervention materials before introducing them in the classroom. Their feedback was solicited to ensure the content of the recruitment materials was suitable to the literacy levels of GCSE resit students. In contrast to the Study Supporter trial, tutors were introduced to the theoretical foundations of the intervention. The tutor interviews conducted towards the end of the Study Supporter trial showed that tutors felt a lack of information about the programme at the start of the year prevented them from introducing the project in an enthusiastic manner. It was also felt that all tutors should be present at the briefing to ensure uniform take-up between the classes (see section 4.8.6, p. 122).

After the tutor-briefing workshop, two tutors at every participating college were appointed as subject leads of the intervention. These individuals helped distribute instruction materials to all teaching staff and coordinated the sharing of course curricula. These curricula informed the schedule of 35 text messages. Tutors then introduced the project to their GCSE English and maths students during a three-week window, from the last week of September until mid-October 2016. They were asked to introduce the project by showing a short informative video to their students.[36]

Figure 8: Screenshot of introductory video, student recruitment phase



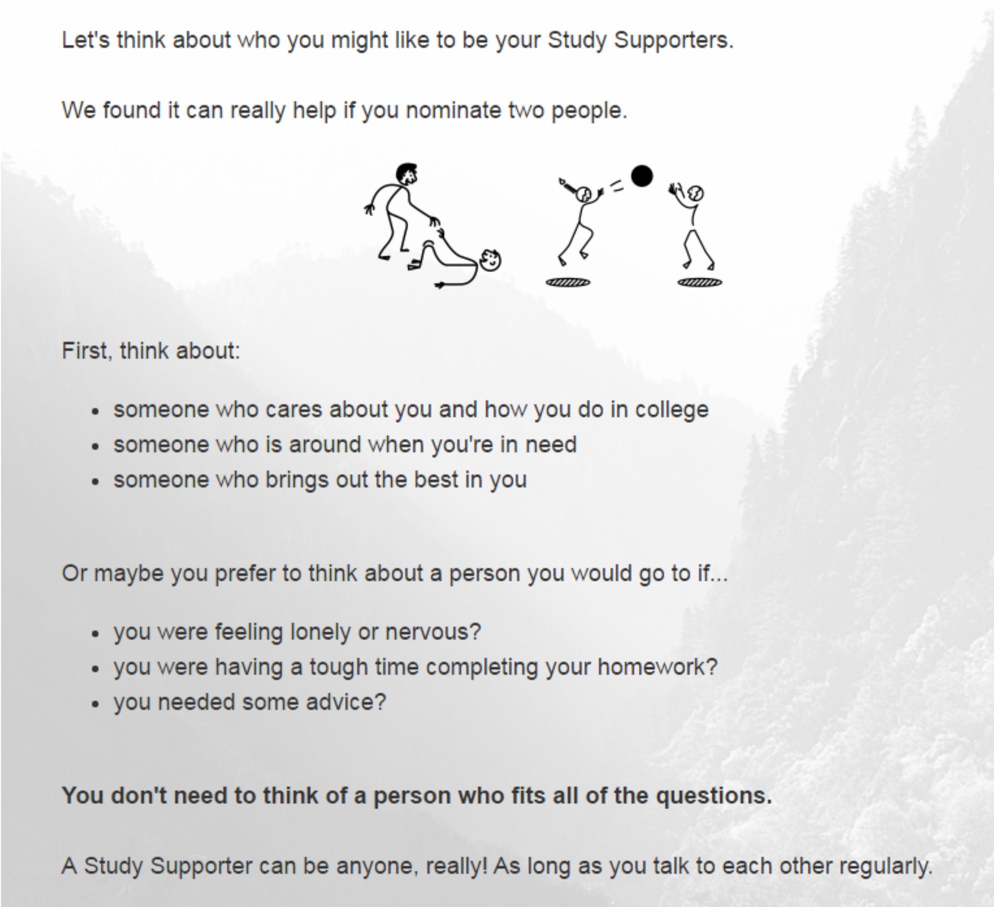
Both students and tutors were informed about the random allocation procedure in detail. It was anticipated that this information video would satisfy students' questions more adequately than the one-page sign-up survey students signing up to the previous study (Chapter 4) were presented with. Additionally, students and their nominated study supporters could read more about the project online.[37]

Students were then provided with the online Qualtrics survey link and were asked to indicate their consent. Appendix 23 lists the full survey text (p. 265).

[36] The survey and video can be accessed via https://uclpsych.eu.qualtrics.com/jfe/form/SV_exGaoH7MPz9bdjf
[37] The website is available at www.projectcollegesuccess.co.uk/video

Those who did not provide consent were asked to list reasons for declining to participate. Students who provided active consent were first guided through a reflection exercise on choosing a suitable supporter (see Figure 9). In this exercise, students learn "what good study supporters look like" (see Appendix 24 for the full guidance text, p. 269). This exercise was developed in response to student feedback to the previous version of the sign-up survey, which did not help guide student choices. Many students nominated their classmates or peers (41.4%) in the previous trial. This time, however, students nominated a greater proportion of family members than classmates or peers (56.1% versus 32.6%). After this short exercise, students were asked to provide contact details for up to two individuals. To help incentivise participation, students could win one of six £25 Amazon vouchers.

Figure 9: Guidance on nominating a suitable study supporter



5.2.4.1 Understanding reasons for non-participation

As introduced in the section above, students who declined to participate completed a short survey. Their reasons for opting out were of interest for two reasons: (1) to further develop engaging recruitment materials, and (2) to assess the evidence of differences in demographic profiles between those who self-select into the trial and those who decline to participate.

Three groups were constructed; (1) students who declined to participate; (2) students who provided consent but subsequently failed to provide valid phone numbers either for themselves or their study supporter(s), and (3) students who provided consent and all necessary contact information (i.e. participants in the trial). In total, 378 students completed the opt-out survey. 189 students provided opt-in consent but failed to provide the necessary information. The demographic profiles are displayed in Table 5.3.

Table 5.3: Demographic profile of consenting and non-consenting students

	(1) Opt out	(2) Opt in, in-complete	(3) Opt in, complete	(1) - (2)	(1) - (3)	(2) - (3)
	M (SE)	M (SE)	M (SE)	*p* - value	*p*- value	*p*- value
Gender: male	0.675	0.674	0.511	0.989	0.000	0.000
	(0.024)	(0.035)	(0.016)			
Age: 16-18	0.648	0.598	0.575	0.247	0.013	0.565
	(0.025)	(0.036)	(0.016)			
Resitting GCSEs after KS4	0.737	0.689	0.680	0.268	0.039	0.819
	(0.023)	(0.036)	(0.015)			
Expecting > C, maths	0.881	0.836	0.939	0.250	0.008	0.003
	(0.020)	(0.034)	(0.009)			
Expecting > C, English	0.906	0.941	0.954	0.242	0.018	0.593
	(0.019)	(0.023)	(0.008)			
N	378	189	975			

Notes: All variables are binary dummy variables, and values in this table are displayed as proportions. Students were asked to indicate if they were resitting their GCSEs at college for the first time (after KS4), or whether they had already taken the GCSE course at further education. In this table, the proportion of students who self-reported taking their first GCSE resits at college is reported.

Pairwise comparisons between the groups show that those who opt out are distinct from those who participated in the trial in a number of ways. The former group is significantly older, predominantly male, more likely to report taking GCSEs at a further education college for the first time, and less likely to expect a pass grade both in maths and English than students who successfully completed the sign-up procedure (all $p < 0.05$). The students who provided consent but failed to share valid contact details are similar to the opt-out students on some dimensions (i.e. gender and expectation of success), and similar to our trial participants on others (i.e. age and resit status). Unfortunately, I cannot assess whether the opt-out students are less likely to benefit from the programme than those who self-selected into the intervention since academic outcomes of opt-out students were not collected. If the new General Data Protection Act (GDPR) permits, future studies may benefit from collecting administrative data for all students at participating colleges regardless of whether students opt in to receive the text messages.

Reasons for non-completion were of interest for further development of the intervention. The most popular reason for opting out was not wanting to participate in research (50%), followed by not being able to identify a suitable study supporter (31.5%) and not wanting to receive college communications via SMS (26.7%). Especially the proportion of students who report they could not think of a suitable individual to nominate is cause for concern. Future iterations of the intervention could further explore how to help students without strong existing social networks. Several college tutors suggested these students could nominate a member of the wellbeing support team at college, for example.

5.2.5 Power calculations

Basic power calculations for binary outcomes were carried out, with varying numbers of treatment arms. The binary outcome variable of interest is students' pass rate on the final GCSE exam. The minimum detectable effect size is calculated as the difference in means between the control and intervention group. As set out in Chapter 4, the MDES is the smallest true effect that can be detected from an experimental study for a specified level of statistical power, statistical significance, and sample size (Bloom, 1995). To calculate the minimum detectable effect size (MDES), the statistical significance level was set at 5% and statistical power at 80%.

Baseline attendance and attainment are collected for all consenting students before randomisation. As baseline and final achievement data were strongly correlated ($r = .70$) in a similar study by Chande et al. (2017) with maths and English students at further education colleges, I assume that baseline outcome data and student characteristics explain 30% of the variation in final attendance and achievement outcome data. The inclusion of covariates in the regression models reduce the residual variance of the outcome variable and therefore lower the needed sample size (see McConnell & Vera-Hérnandez, 2015, p. 12 for a discussion of covariate-adjusted power calculations). Covariate-adjusted power calculations are therefore used here. The MDES is therefore downward adjusted by 30%. The unadjusted power calculations are available in Appendix 25 (p. 270).

Table 5.4 displays the power calculations for several scenarios, by number of students who provide consent and number of trial arms. Based on the 2015/26 qualification achievement rates for post-16 GCSE, I assume a baseline pass rate of 28.2% (Department for Education, 2016b)[38]. The total number of eligible students at each of the four participating colleges is displayed in the final column of Table 5.2.

[38] The provisional qualification achievement statistics were published by the Department for Education in 2016, and this data (28.2% averaged across GCSE maths and English) was used for the power calculations. The revised data was published in Spring 2017 and saw a small reduction in GCSE qualification pass rates averaged across maths and English, at 26.8% (Department for Education, 2017).

Table 5.4: Ex ante power calculations, individually randomised design

	Scenarios: N opt-in			
	500	750	1000	1250
Trial arms	MDES δ (N per arm)	MDES δ (N per arm)	MDES δ (N per arm)	MDES δ (N per arm)
2 Arms	0.083 (250)	0.067 (375)	0.058 (500)	0.052 (625)
3 Arms	0.103 (166)	0.083 (250)	0.071 (333)	0.064 (416)
4 Arms	0.120 (125)	0.097 (188)	0.083 (250)	0.074 (313)

Notes: Power was set at 0.80 and alpha at 0.05. The allocation proportion was set equal group sizes. MDES estimations were adjusted downward by 30% as I assume outcome variation can be partly explained by controlling for baseline achievement and attendance.

As only four colleges participated in the experiment the maximum total number of opt-ins was estimated at 1250 students, and the minimum at 500 students. The adjusted power calculations show that reaching 1250 sign-ups would allow us to retain all four trial arms. In this scenario, we would be able to detect an increase in achievement rates of 7.4 percentage points. This effect size corresponds to Chande et al. (2017) who found an 8.7 %-point increase in students' likelihood of passing the course.[39] The Study Supporter intervention, set out in Chapter 4, improved maths and English pass rates by 6 % points. Retaining four treatment arms would allow for a direct comparison between the direct informational value of the text messages versus the socially-mediated effect. It was thus decided to maintain four trial arms, with the aim to recruit approximately 1250 students across the four colleges.

The four trial arms were constructed as follows:

1 A control group, where neither student nor supporter receive SMS communications;

[39] The authors sent direct text messages to students, with reminders about upcoming exams, planning prompts, and general motivational content. The content of the text messages is similar to the intervention discussed in this chapter and the mode of delivery is identical, but the recipient is the FE college student, rather than a third party.

2 The nominated study supporter receives text messages;
3 The student receives text messages, and;
4 Both study supporter and learner receive the messages.

5.2.6 Randomisation

In total 1164 students provided active consent. Phone number validation was performed once the sign-up window closed, using an online mobile verification service.[40] Phone numbers were missing, incomplete or non-connected for 189 students. The three eligibility criteria for students were that (1) they studied towards a GCSE English and/or maths, (2) they were studying full-time and (3) that they provided a valid phone number for themselves and their nominated study supporters(s). The first two criteria had been validated through the online survey. The 189 students who did not satisfy the third criterion were not randomised into treatment groups.

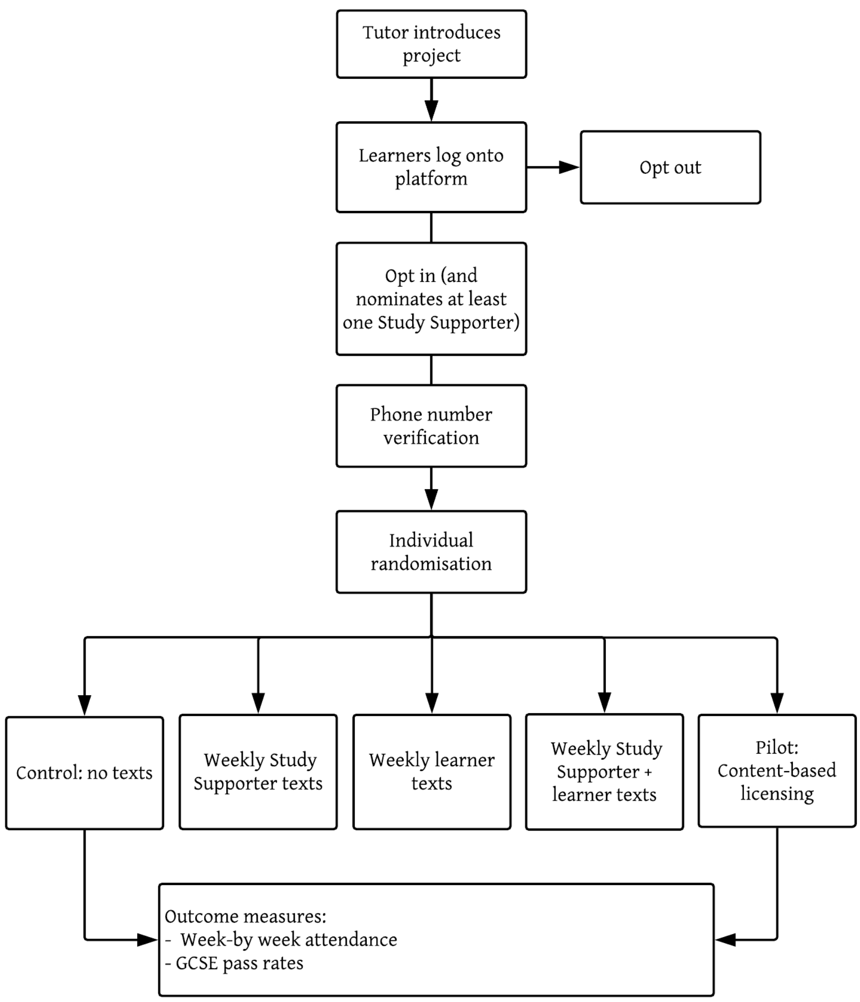
In total, 975 students were randomised into treatment groups. Individual-level randomisation was performed. Sixty-three students (6.5% of the sample) were assigned to a small pilot ('content-based licensing', described in more detail in section 5.2.7.1, p. 157), and the remaining 912 students to one of the three treatment arms or the control arm. Individual randomisation was stratified by gender, baseline attendance (quartile split) and age (binary; 16-18 vs. 19+). Students taking both maths and English were randomly assigned to be treated in one of the two subjects. Of the 975 randomised students, 473 are treated in their English GCSE and 502 are treated in maths. Table 5.5 displays the distribution of treatment and control numbers by college and subject. Finally, students who nominated two study supporters (N = 367) were randomly assigned to having either their first or second contact receiving the schedule of text messages. Figure 10 displays the intervention procedure as outlined above.

[40] The verification software was accessed via https://www.datasoap.co.uk/ on October 22nd, 2016.

Table 5.5: Treatment assignment by college and subject

	Control	Supporter only	Student only	Supporter + student	Pilot (CBL)	Total
	N	*N*	*N*	*N*	*N*	*N*
College 1	11	19	8	22	5	65
College 2	47	43	48	51	14	203
College 3	60	62	64	66	19	271
College 4	109	103	110	89	25	436
English	102	113	113	117	28	473
Maths	125	113	117	111	35	502
Total	*227*	*227*	*230*	*228*	*63*	*975*

Figure 10: Flowchart of recruitment and randomisation procedure



5.2.7 The intervention

The intervention is an iteration on the Study Supporter intervention (Chapter 4). All students who provided active consent nominated up to two study supporters. Nominated study supporters and the learners themselves assigned to the 'student only' arm were contacted once a week throughout the 2016-17 academic year via text messages. Recipients in any of the three the treatment groups received an additional text message 3 days before the final GCSE exams. The text messages were scheduled via the FireText texting platform, for Thursday evenings at 7PM. This trial design allows a rigorous test of whether the text messages help improve learner success by their informational value, or whether having a supportive individual who receives updates about College is beneficial over and above receiving this information oneself.

The texts were written to ensure they were specific to the curriculum being taught, but the type of topics and positive wording were similar to the Study Supporter intervention in Chapter 4. The messages focused on upcoming deadlines or exams, course content, booking extra tutorial sessions, academic resources, and general motivational content. Appendix 26 provides a breakdown of text message categories and example text messages (p. 271). In line with Chapter 4, the aim of the programme is to encourage study supporters to talk to the student about the student's educational experience at the college. Study supporters are not informed about the learners' academic performance or in-class behaviour.

In addition to the communications directed at nominated study supporters, students could receive direct text messages. Students assigned to the 'Study supporter + learner' group received text messages at the same time as their study supporters, so both parties knew they were encouraged to speak to each other. These messages informed students that their study supporter might ask them specific questions about their maths or English class over the coming days. As Table 5.6 shows, the messages are identical for the study supporters in the 'supporter only' arm and the 'supporter + student' arm. The texts directed at students do (supporter + student group) and do not (student only group) refer directly to the study supporter they had nominated. In order to ensure equivalence across the treatments, all texts included identical information and were similar in length.

Table 5.6: Example text messages across treatment groups

Group	Recipient	Text message content
Supporter only (1)	Supporter	Hello {supporter first name}, {student first name} will have another controlled assessment in English class next week. They will be assessed on the newspaper article ("What would I get rid of to improve life in the 21st century?"). Ask {student first name} how they're planning to prepare. #SUCCESS
Student only (2)	Student	Hello {student first name} just a reminder that you will have another controlled assessment in English class next week. You'll be assessed on the newspaper article ("What would I get rid of to improve life in the 21st century?"). Think about how you plan to prepare. #SUCCESS
Supporter + Student (3)	Supporter	*Identical to row 1.*
	Student	Hello {student first name} just a reminder that you will have another controlled assessment in English class next week. You'll be assessed on the newspaper article ("What would I get rid of to improve life in the 21st century?"). It's a good idea to chat to {supporter first name} and let them know how you are preparing for it. #SUCCESS
Content-based licensing (pilot) (1)	Supporter	Hi {supporter first name}, this week we have set {student first name} a maths puzzle. To solve it, you'll need to tell them that the ticket price is £6. Ask them to explain what this means! If they text us with the right answer to the puzzle they could win 2 cinema tickets. #SUCCESS
	Student	Hi {student first name}, imagine you own a cinema with 50 seats. How much do you earn when the cinema is full? Hint: talk to {supporter first name} to find out the ticket price. Text us the answer (the amount you earn when the cinema is full) and we'll enter you into a lottery to win 2 cinema tickets! #SUCCESS

The text messages were written using colleges' schemes of work. College subject leads were invited to provide input on the topic and wording of the messages to ensure that the text messages were specific and applicable to each college, as there is substantial variation between colleges in terms of key dates and course content. Additional example texts are displayed in Appendix 27. (p. 272).

5.2.7.1 A pilot study

Alongside the four arms, a small pilot study was conducted to explore whether an interactive component could increase engagement with the programme. This pilot, henceforth named 'content-based licensing' (CBL), tests whether meaningful conversations between learners and their study supporters can be stimulated through the provision of an opening topic. The methodology and outcome measures are identical to the four main trial arms. The only difference between the 'content-based licensing' arm and the 'learner + supporter texts' arm is that the CBL arm includes a specific call to action. This trial arm was added to aid the generation of new hypotheses for future research and exploration of the potential of interactive text messaging interventions. It is hypothesised that an explicit reason to talk to each other about the received text message may help the learner-supporter pair further build a supportive relationship, and ultimately improve learner outcomes. Example text messages are displayed in the final row of Table 5.6.

Students allocated to the pilot received the specific content-based licensing prompts every few weeks, rather than every week. This is to ensure that the prompts remain interesting and curiosity-inducing. In the remaining weeks, learners and their study supporters received the same text message prompts as recipients in the 'supporter and learner' group.

5.2.8 Quantitative analysis plan

An ordinary least squares (OLS) approach is taken for all regression models (with the exception of the duration model introduced below), including those with binary dependent variables. As robustness checks, logistic regressions are performed on the binary achievement outcome data. These are further discussed in section 5.6.

Primary analyses adjust for student-level covariates including gender, age, subject, resit status, baseline attendance and achievement, and college-level fixed effects. The analysis plan was pre-registered via the AEA online registry (AEARCTR-0001644) before randomisation.

5.2.8.1 *Average treatment effects*

I estimate the following models to uncover the Average Treatment Effects (ATE) on an intent-to-treat basis:

[9a] $$Y_{ij} = \alpha + \beta_1 \boldsymbol{T_i} + \epsilon_i$$

[9b] $$Y_{ij} = \alpha + \beta_1 \boldsymbol{T_i} + \beta_2 M_i + \beta_3 \boldsymbol{D}_i + \beta_4 \boldsymbol{C}_j + \epsilon_i$$

where:

Y_{ij} is a binary variable of the learner's final GCSE grade, equal to 1 if individual i in college j scored C or above (or 9 - 4 in the new system), and 0 if individual i scored lower than a C (3 - 1). For attendance Y_{ic} represents the attendance rate averaged over the trial period (35 weeks) for individual i in the treated subject;

α is the regression constant;

$\boldsymbol{T_i}$ is a vector of treatment indicators, equal to 1 if participant i is assigned to one of the four treatment groups (including the pilot), and 0 if assigned to the control condition. β_1 is the main parameter of interest, as this is where the treatment estimates are recovered;

M_i is a student-level binary variable, equal to 1 if individual i is treated in a maths class, and 0 if they are treated in an English class;

$\boldsymbol{D}_i$ is a vector of learner level pre-treatment covariates of gender, age, resit status, baseline attendance and baseline achievement;

$\boldsymbol{C}_j$ is a vector of college-level fixed effects; and

ϵ_{ij} is the individual-level error term. Robust Huber-White standard errors are calculated.

The pre-registered regression model was amended slightly in light of changes made during the randomisation process. First, the pre-specified model in the trial protocol included an indicator of opt-in status. As this study only randomised learners who had opted-in, this indicator is set to 1 for all participants. It has therefore been taken out of the model. Second, initially it was planned to stratify randomisation by class, but a number of tutor groups contained too few consenting students. This over-stratification error was overcome by stratifying at the college level instead, so that sufficient numbers of students could be distributed across all bins. This updated strategy is reflected in the model. Rather than controlling for class-level effects, college-level fixed effects are included in the model. Finally, separate indicators for maths and English were removed as they are perfectly collinear. Students who take both subjects were only treated in either of the two subjects.

5.2.8.2 Heterogeneous treatment effects

The final sample size of 975 students divided over four treatment arms limits what heterogeneous treatment effects can be examined. For example, due to the changing demographic profile of further education college students on GCSE courses, only 203 participants (20.8%) were aged 19 or above; limiting the feasibility of further partitioning the data by treatment group. I therefore explore patterns of heterogeneous treatment effects for a limited number of subgroups. Exploratory subgroup analyses are performed on subject treated (maths versus English) and gender, both of which maintain a roughly even split in the full sample. As in chapter 4, subgroup analyses are performed by restricting the sample to each subgroup and applying the regression model used in the estimation of ATEs (see section 5.2.8.1).

5.2.8.3 Estimating the treatment effect for compliers

The Complier Average Causal Effect (CACE) is estimated using an Instrumental Variables approach in order to better understand treatment effects for students who received a meaningful dosage of the schedule of text messages. This statistical technique focuses on the subgroup of participants who would always have complied with their treatment allocation (Gerber & Green, 2012, p. 147). As a function of randomisation, it is assumed that that the proportion of compliers is roughly equal between control and treatment groups.

Compliance is defined as having received the full schedule of thirty-five text messages. In essence, this analysis estimates the effect of full compliance. The analysis is repeated for participants who had received at least 25%, 50% and 75% of the messages, which includes participants who were fully compliant (see Table 5.8). An Instrumental Variables (IV) is considered more rigorous than per-protocol or on-treatment analyses (Gerber & Green, 2012; Tilbrook et al., 2014). A two-stage least squares (2SLS) model is estimated in which the compliance indicator (e.g. receiving all 35 texts) is instrumented with the treatment indicator.

The first stage involves a model of the outcome variable:

[10] $$Alerted_i = \alpha + \beta_1 \boldsymbol{T_i} + X_i + \epsilon_i$$

where;

$Alerted_i$ is a binary indicator for whether the student, supporter or both,[41] received the full schedule of texts during the intervention;

T_i is a vector of treatment indicators,[42] determined at random assignment;

X_i is a vector of the student-level covariates available for the full sample and college fixed effects used to improve the precision of the primary analyses; and;

ϵ_i is an individual-level error term.

[11] $$Y_i = \beta_0 + \beta_1 Alerted + u_i$$

where;

Y_i is the outcome of interest (i.e. attendance and achievement rates), and;

β_1 is the parameter of interest, the CACE (Gerber & Green, 2012, p. 159). The second equation regresses actual treatment on assigned treatment:

Standard errors in the second stage are adjusted to take account of the instrumented nature of the predictor. The three treatment groups are instrumented separately, to allow straightforward comparison to ITT

[41] In the treatment arm where both student and supporter are alerted, participants are classified as compliers only when both parties receive the full schedule of texts.
[42] The content-based licensing pilot arm is omitted, it was included primarily to explore the potential of a two-way communication intervention, and was not powered to detect moderate treatment effects.

estimates. It was assumed that students assigned to the control group had the same probability of non-compliance as students assigned to one of the three treatment groups, and that their treatment assignment does not have an effect on the outcomes of interest apart from the effect of treatment itself (exclusion restriction; see Schochet & Chiang, 2009, p 11). The results are reported in section 5.6.2 (p. 182).

5.3 Outcome data

5.3.1 Attendance and achievement data

College administrative data was collected from all four participating colleges. They supplied register data (day-by-day attendance registers) and attainment data (final GCSE achievement scores). This data is routinely collected by colleges as part of their business-as-usual operations. Due to subtle differences in the ways attendance keys were defined in each college, the four datasets were individually cleaned and subsequently merged.

Student attendance was collected for the treated maths or English GCSE class. Attendance data was collected in a daily format; all possible and missed attendances are registered over the full academic year. Colleges provide quite detailed information about students' attendance, including whether the absence was (un)authorised, absent due to other college activities or religious reasons, or whether students had transferred or dropped out of the class. Attendance rate was calculated by dividing the total number of attended classes throughout the academic year by the maximum number of classes they could have attended. For example, if a student attended 42 out of 48 lessons, their attendance rate was 87.5%. Attendance is coded strictly: if the student was absent (for any reason), they are coded as non-attending. A number of students transferred between classes within the same subject, and therefore appeared in the dataset several times. In these cases, attendance rates are merged into a single measure of attendance. Additionally, if a student dropped out of college altogether, the weeks in which they were no longer in class count towards their final attendance rate (e.g. if a student attends 100% of classes in the first half of the year and then drops out, their final attendance rate is 50%).

The primary outcome of interest is a binary indicator for whether students achieved their final GCSE exam administered in summer 2017. Final GCSE grades, ranging from A* - U (and 9 – 1 in the new scoring system introduced in 2016), are recoded into a pass/fail dummy. Students also sit a diagnostic assessment at induction. This assessment, which was provided by bksb®[43] for all four participating colleges, assesses students' individual skill gaps at the start of the academic year. Students received a point score between 0 and 1.

5.3.2 Attrition

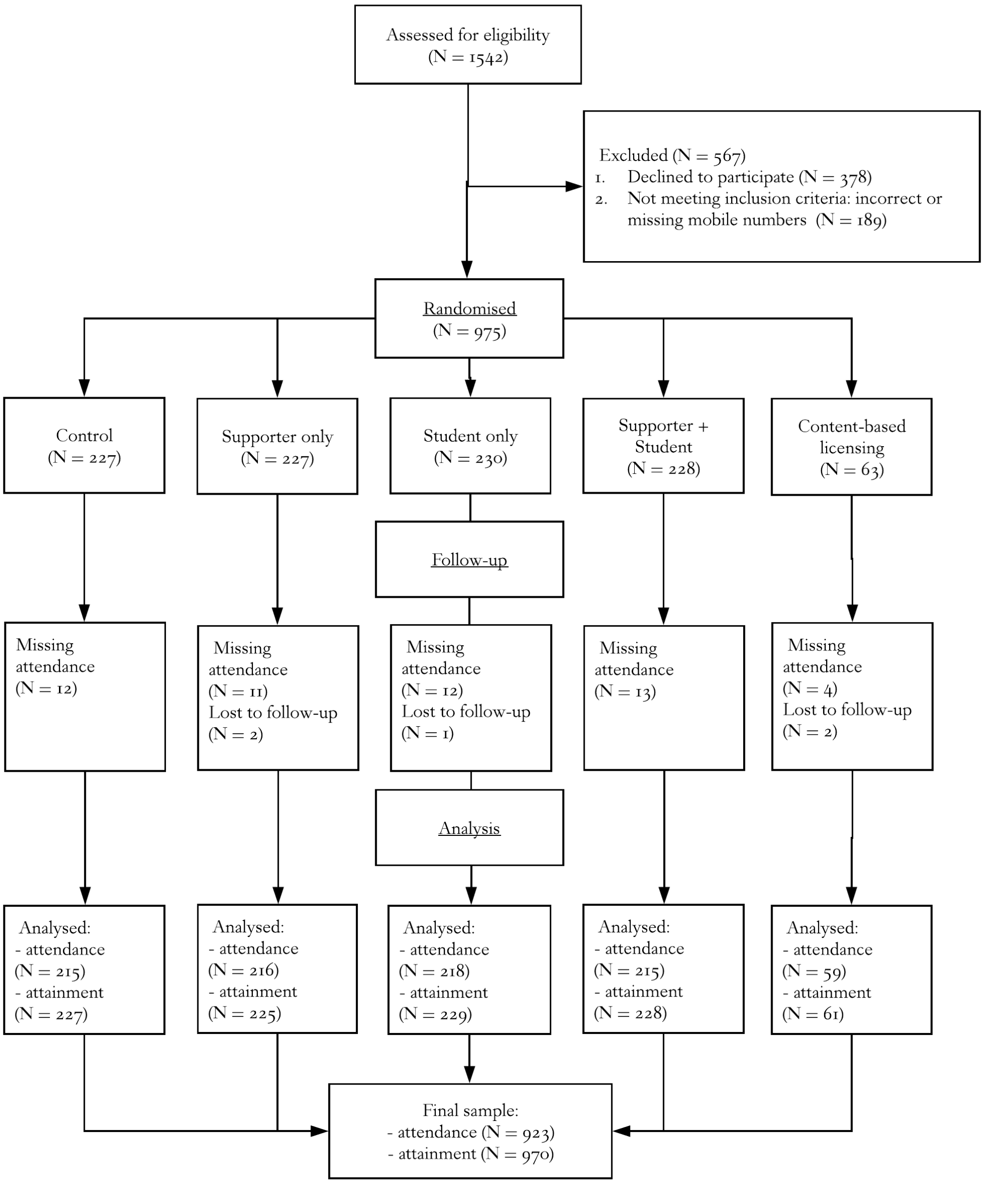
As reported in section 5.2.6 on randomisation (p. 95), 975 students actively consented to take part and provided the necessary details to be randomised. Attendance data was provided for 923 students (94.7%), and attainment data for 944 students (96.8%). The discrepancy between these numbers arose from imperfect college reporting, where for some students the correct attainment results were provided (N = 19), but attendance data was provided for their non-treated subject. Overall, attrition was low at 3.2% of the total sample.

Pre-trial data did not merge successfully with the college outcome datasets for 31 participants. Loss to follow-up is attributable to two issues. First, pre-trial data for five students did not merge to outcome data due to misspelled student ID and colleges were not able to match these students based on their demographic characteristics. These students are discarded for the primary analyses as both pre-trial data and outcome data is missing. Sensitivity checks, where outcome data is imputed for this small group of students, are reported in the analysis section.

The remaining non-merging students (N = 26) dropped out of the subject or college altogether. Their attainment was coded as a 'fail', but their attendance rate is coded as missing. Attendance data could not be recovered for these students. Student drop-out as an outcome variable in its own right is further explored using a duration modelling approach, Section 5.6.4 (p. 185). See Figure 11 for a graphical representation of the data collection and attrition.

[43] The online diagnostic maths and English tests are provided by bksb, a learning provider that operates across a large share of England's further education colleges, https://www.bksb.co.uk/

Figure 11: Consort flow diagram



In line with intention to treat (ITT) analysis, all students providing data were analysed regardless of protocol deviation or student dropout in the primary analyses. This analysis strategy is regarded as the most robust approach against bias, as those who deviate from assignment or drop out of the trial may be a non-random subset of the trial sample in ways difficult to measure (Gerber & Green, 2012, p. 153). List-wise deletion, where such cases are simply discarded and analysis is performed on the participants with full outcome data, may lead to selection bias (White, Horton, Carpenter, & Pocock, 2011). Finally, the estimated ITT allows for an evaluation of overall programme effectiveness (Gerber & Green, 2012, p. 150) and provides a useful basis for further scaling up of the intervention where deviation would also likely occur. Although ITT is the primary analysis strategy used, the complier average causal effect is explored as a secondary analysis strategy in order to assess treatment effectiveness for compliers, or those who received the intended text messages. See section 5.2.8.3 for the CACE regression specification (p. 159).

5.3.2.1 Exploring differential attrition

The final row of Table 5.7 reports on differential attrition by treatment status. None of the differences between the groups are statistically significant ($p > 0.05$) and the reported normalised differences are modest. Although differential attrition between treatment groups therefore does not appear to plague this trial, the students with incomplete outcome data do appear to be a non-random subset of the experimental sample.

Appendix 29 displays the covariate balance table for students with and without missing outcome data (p. 274). The missingness of outcome data is related to covariates in the dataset. In comparison to complete cases, a higher proportion of students with missing outcome data are female, older, and with lower baseline attainment. While there is no clear consensus on the cut-off for meaningful imbalance between covariate distributions, scholars have proposed that normalised differences greater than 0.10 (i.e. 10 per cent) are cause for concern (Austin & Austin, 2009). Baseline attendance was only collected for two students in the missing-data group and is therefore of limited interest. Since students aged 19 or above are no longer required to take GCSE qualifications to receive funding for their vocational courses (Education and Skills Funding Agency, 2018), the incentives to stay on the course are likely weaker.

5.3.2.2 Missing data handling methods

As Figure 11 shows, attendance data is missing for 47 participants. To ensure that missing data does not hinder the generalisability of our findings to the target population, results are reported both for the complete cases and the total sample including multiply imputed data. Multiple imputation using chained equations (MICE) was performed, which fills in missing values iteratively by drawing on auxiliary variables present for all students (i.e. gender, college, age). These variables are associated with the dependent variable and related to the missing mechanism. Thirty imputations were conducted. MICE is a more sophisticated approach to handling missing data than mean or null imputation, as it simulates natural variation in missing data and is less prone to producing biased (smaller) standard errors (Cheema, 2014).

Additionally, a number of relevant covariates were missing for a subset of our sample. Baseline achievement data is missing for a substantial subset of students (30.7%), and baseline attendance for a small proportion (4.6%). The baseline achievement data is judged to be missing at random (MAR), as college staff report that the completion of the baseline assessment is not uniformly enforced by tutors. Poor reporting of the assessment scores was suggested to be another reason for the relatively high proportion of missing data (personal communications with College 1, September 2017). In order to preserve sample size in the regression models that include these covariates, missing baseline values for attendance and achievement were imputed. The MICE approach, described above, was used to impute missing baseline values.

In summary, the following approach will be applied to the analysis, in line with the framework proposed by White et al. (2011). The main analysis is performed first on the observed outcome data, and second on the full dataset with multiple imputation. The final two regression models repeat model 1 and 2, but include pre-specified covariates. These sensitivity analyses serve to confirm whether the inferences made from the primary analyses are maintained or changed. The following four models are reported for all primary analyses:

1. Simple model, complete case analysis;
2. Simple model, using the multiply imputed dataset;
3. Additional pre-specified covariates, complete case analysis, and;
4. Additional pre-specified covariates, using the multiply imputed dataset.

5.3.3 Baseline treatment-control balance

Table 5.7 displays the balance between treatment groups on observable covariates. Normalised differences of 0.05 should be noted, and 0.10 or larger are cause for concern (Austin, 2009). The treatment groups are generally well balanced, although differences in attendance at baseline and age are discernable in comparisons between the three treatment groups and the control ($\hat{\Delta}_{ct}$ > 0.05). Gender, ethnicity, baseline score and resit status are balanced, with the exception of the content-based licensing group which is smaller in size due to its pilot nature. In this group, students are slightly older and score lower on the baseline test.

In summary, randomization appears to be relatively balanced on observable characteristics. To ensure that baseline differences do not exert an undue influence on the results, baseline attendance and age are included as control variables in the primary analyses.

Table 5.7: Normalised differences of covariate distributions between treatment groups

	(0) Control M (SE)	(1) Supporter only M (SE)	Norm. difference (0 – 1) $\hat{\Delta}_{ct}$	(2) Student only M (SE)	Norm. difference (0 – 2) $\hat{\Delta}_{ct}$	(3) Student + supporter M (SE)	Norm. difference (0 – 3) $\hat{\Delta}_{ct}$	(4) Content-based licensing M (SE)	Norm. difference (0 – 4) $\hat{\Delta}_{ct}$
Gender: male	0.507	0.507	.000	0.513	0.009	0.509	0.003	0.524	0.024
	(0.033)	(0.033)		(0.033)		(0.033)		(0.063)	
Age	18.796	18.521	-0.042	19.288	0.060	19.221	0.052	20.282	0.147
	(0.361)	(0.245)		(0.405)		(0.405)		(1.061)	
Ethnicity: white	0.317	0.366	0.072	0.326	0.013	0.351	0.050	0.317	0.001
	(0.031)	(0.032)		(0.031)		(0.032)		(0.059)	
Baseline score	0.640	0.639	-0.004	0.637	-0.008	0.620	-0.065	0.577	-0.204
	(0.017)	(0.017)		(0.017)		(0.017)		(0.034)	
Attendance at baseline	81.876	78.612	-0.098	79.152	-0.085	80.422	-0.049	81.373	-0.016
	(1.470)	(1.709)		(1.585)		(1.385)		(2.888)	
First time resit	0.425	0.436	0.016	0.450	0.036	0.410	-0.022	0.429	0.005
	(0.033)	(0.033)		(0.033)		(0.033)		(0.063)	
Outcome missing: attendance	0.053	0.040	-0.043	0.048	-0.016	0.057	0.013	0.033	-0.070
	(0.015)	(0.013)		(0.014)		(0.015)		(0.023)	
N	227	225		229		228		61	

Notes: Sample means, robust standard errors in parentheses. Normalised difference scores are calculated as the sample averages of the covariate values for the control and treatment group, normalised by the square root of the average of the two within-group sample variances (Imbens & Rubin, 2015; see p. 311 for formula). In this table, normalised differences are calculated for baseline variables for the control group and contrasted against each of the treatment groups. Age is measured in years, attendance is measured from 0 – 100%, and baseline score is a point score between 0 - 1. Ethnicity is reported as the proportion of White participants, and first time resit indicates the proportion of students in the sample that self-reported resitting their GCSEs for the first time this year.

5.3.4 Compliance with treatment

All text messages were sent as intended, 35 weekly text messages in total. Phone numbers were validated prior to randomisation. If students' or study supporters' phone numbers were not connected to a UK phone provider, they were not randomised (N = 189). Nevertheless, not all intended recipients received the full schedule of text messages, which could be observed through detailed delivery statistics (e.g. bounce-backs, opt-outs) obtained from the text-messaging platform. Additionally, a small number of recipients' mobile numbers could not be merged back to the text messaging data (N = 41). These individuals had changed numbers after signing up and before trial launch, and thus did not receive any text messages.

Participants were able to opt out of receiving the text messages at any time. In total, 6.4% of nominated study supporters unsubscribed from the programme of text messages throughout the academic year, and 10.1% of students did so. Second, individuals who switched phone numbers during the trial received fewer text messages, as changes in phone numbers were rarely communicated to the research team. All in all, both students and supporters received a varying number of text messages.

Table 5.8 shows the opt-out rate and successful delivery rate for the schedule of text messages in the treatment groups averaged over the full trial period. Students assigned to the control group did not receive text messages during the trial, and neither did their nominated study supporter. The rate of successful text delivery is high overall, as on average 79 % of text messages were delivered to recipients. This rate of successful text delivery is higher than delivery statistics reported in similar studies. For example, Berlinski et al. (2016) report a 60 to 70 per cent success rate and Castleman and Page (2017) were able to reach 50 per cent of parents assigned to receive their programme of informative messages.

Approximately two third of recipients received all 35 text messages (68.5%), and the vast majority (80.3%) received at least 75% of the scheduled texts. The total number of texts delivered is significantly different for the supporter and student only arms (Table 5.8, column 1 – 2) versus the combined treatment arms (Table 5.8, column 3 – 4, $p < 0.05$). Successful delivery was significantly higher in the 'student & supporter' group than the other treatment groups ($p < 0.05$), but this difference disappears when the stricter definition of the success rate is applied, where both parties are required to have received the full schedule. Further exploration of the data indicates that supporters were more likely to having received all texts than students. Informal conversations with college staff indicate that students change phone numbers relatively often, and perhaps more so than study supporters. Interestingly, the un-subscription rate is considerably higher both for students and supporters in the content-based licensing group which suggest acceptability of the two-way communication messages was somewhat reduced.

Table 5.8: Fidelity of treatment implementation

	(1)	(2)	(3)	(4)	
	Supporter only	Student only	Student + Supporter	Content-based licensing	Total
	M (N)	M (N)	M (N)	M (N)	M (N)
Total SMS delivered (35 max.)	26.42	24.84	30.95	30.69	27.67
	(225)	(229)	(228)	(61)	(743)
Opt-out from texts (proportion)					
Students	-	0.072	0.113	0.164	0.101
	-	(15)	(23)	(9)	(47)
Study supporters	0.067	-	0.049	0.123	0.064
	(14)	-	(10)	(7)	(31)
Delivery success (proportion)					
100%	0.560	0.561	0.803	0.689	0.647
≤ 75%	0.760	0.725	0.904	0.885	0.803
≤ 50%	0.813	0.773	0.934	0.918	0.847
≤ 25%	0.844	0.821	0.947	0.984	0.880
Strict: both receive 100%	0.560	0.561	0.439	0.295	0.499

Notes: Neither students nor supporters in the control group received text messages. Total number of SMS delivered is calculated per individual who was assigned to receiving texts – these statistics include those who never received texts but should have. Of the student-supporter pair, in column 3 and 4. the recipient with the highest number of texts is counted. In column 3 and 4, the student + supporter group and content-based licensing group, delivery success is defined as either or both the student and supporter receiving all 35 text messages. A stricter definition of 100% success rate, where both parties within the pair receive the full schedule, is provided in the last column of this table.

The SMS delivery success data is further used to inform the Instrumental Variables approach to estimate the Complier Average Causal Effect (CACE) in section 5.6.2 (p. 182).

5.3.5 Assessing contamination

To assess the threat of contamination, or more specifically, treatment crossover, students were asked if their nominated peers were also their classmates in maths or English GCSE. Of those who nominated a peer as study supporter, 22.9% of students indicated that this individual is also their classmate. This is a worrying figure, as students were explicitly asked not to nominate a direct classmate. Indeed, doing so introduces the possibility that participating students nominated each other. In total 96 students appear twice, both as a student (i.e. intended recipient of the intervention) and nominated supporter as measured by matching phone numbers. Of this group, 22 students were allocated to control but received texts about their classmate's learning, and 33 students were originally allocated to 'supporter only texts' (thus should not have received messages themselves) but also received texts as their classmate's supporter. The remaining 41 students did not receive (additional) texts since their classmate was allocated to control or 'student only' groups.

Since all texts are tailored at the class- rather than the individual-level, the information provided in these texts is relevant to the intended recipient as well as their nominated classmate. Such treatment crossover threatens the validity of the experiment. A second issue, identified during data cleaning, is that a number of students nominated themselves (N = 72). Although some gave different names than their own, these students provided their own mobile phone number. As a result, those assigned to the 'supporter only', 'supporter + student' or content-based licensing arms received the study supporter texts themselves (N = 41).

To assess the effect on contamination on the inferences made from the primary analyses, two separate robustness checks were performed. First, the contaminated students are re-assigned to their observed treatment, and the second robustness check removes them from the analysis. The results are further discussed in the analysis section.

5.4 Descriptive statistics

5.4.1 Student demographics

Student age, ethnicity and gender are collected via college administrative datasets. Student resit status (i.e. how often they have taken their GCSEs) and living situation were collected via the online sign-up survey. Furthermore, Free School Meals (FSM) status was requested but only collected for a fraction of the sample (15.9%) so not reported here.

Gender, age and resit status are complete for the full sample. These are the student-level covariates controlled for in the primary analyses, and their balance across the treatment and control groups can be inspected in Table 5.7 (p. 167). The median age of students in our sample is 17.7 years old (79.2% of students are 16-18 year olds), 50.9% are male, and the majority are resitting their GCSEs at a further education college for the first time (43.2%) as opposed to a second or third GCSE catch-up year. Thirty-two per cent of the sample report taking their GCSEs for the first time, which may be due to students moving to England only recently or having taken alternative qualifications such as Functional Skills prior to this academic year. Finally, 24.8% of students report taking the GCSE subject again at the same further education college.

The remaining student demographics have missing values due to item non-response on the sign-up survey. These variables are not included in regression models and only used for descriptive purposes. 89.3% of students in our sample report living with their parents or guardians, and the remainder live independently or with their partners. Furthermore 34% of our sample is White, 32% is Asian British, 14% is Black British, and the remaining students are of mixed or other ethnicities.

The sample is relatively homogeneous, with a high proportion of relatively young live-at-home students (89.3%); a higher proportion than the trial sample in Chapter 4. This observed shift in demographics in further education colleges is not unique to our four participating colleges. Since students who obtain a D grade for their GCSEs are required to re-take GCSEs rather than stepping stone qualifications, further education colleges have seen a large increase in numbers of resit students who come straight from secondary school (Impetus-PEF, 2017). As the Study Supporter intervention from Chapter 4 was shown to benefit 16-18 year olds more so than 19+ learners this chapter will test if this remains the case.

5.4.2 Study supporter choice

Consenting students self-reported their relationship and closeness to their nominated study supporter(s), as well as whether they lived together, frequency and mode of communication, the study supporter's level of education, gender, first language, and whether the study supporter understood written English. Students could nominate up to two individuals; but only one was randomly allocated to receiving the programme of text messages. The below statistics are reported for the study supporter who received communications from Project Success only.

The majority of study supporters fall within two broad categories: nuclear family (47.9%) and peers (33.7%). Next, students nominated partners (7.8%), extended family such as grandparents, aunts and uncles (5.7%), professional support including teachers or support workers (1.6%), and colleagues (0.6%). The remaining 2.7% of students did not provide information on the relationship they have with their nominated supporter.

Study supporters were slightly more likely to be female than male (56.3% vs. 43.7%). Study supporter age was not collected because students often did not recall in the previous iteration of the sign-up survey in Chapter 4. There was an even split of cohabiting and non-cohabiting supporters (50% vs. 50%), and students overwhelmingly reported feeling either close or very close to their supporters at the start of the trial (88.8%). This final statistic is corroborated by the reported frequency of communication, as 88.5% of students reported speaking with their supporter at least 4 out of 7 days. Overall, these descriptive statistics highlight that most students nominated someone they spoke to regularly and cared about. See Appendix 30 for all descriptive statistics about the nominated study supporters, collected through the opt-in survey (p. 275).

5.4.3 Patterns of attendance and achievement

As can be seen in Table 5.9, attendance was high during the first nine weeks of the academic year at 80%, averaged across all four colleges (Column 1). Contrasting average baseline attendance of students in our sample shows that there is considerable variation across colleges of up to 10 percentage points. The baseline achievement data tells a similar story as average achievement fluctuates considerably between colleges (Column 3). It should be noted that 17.7% of students at College C failed to sit this assessment so these descriptive statistics may be biased by the higher rate of missing data in one out of four colleges.

Similar heterogeneity between colleges can be observed in the final outcome data (see column 2 and 4, Table 5.9). The proportion of students who pass their qualification is exceptionally high in College 3 (36.9%) and exceptionally low in College 4 (13.1%) in comparison to national average (26.8%; Department for Education, 2016b). The differences in average attendance rates between the participating colleges are less pronounced, however (see column 4).

Table 5.9: Descriptive statistics on attendance and achievement rates

	(1) Baseline attendance	(2) Final attendance	(3) Baseline attainment	(4) Final attainment
	M (SD)	M (SD)	M (SD)	M (SD)
College 1	71.22	65.16	.47	.23
	(26.25)	(27.01)	(.17)	(.43)
College 2	85.22	79.44	.48	.26
	(28.06)	(27.78)	(.18)	(.44)
College 3	72.94	67.42	.66	.37
	(24.76)	(19.40)	(.20)	(.48)
College 4	83.71	72.95	.70	.13
	(16.18)	(20.83)	(.19)	(.34)
Total	80.01	72.04	.63	.23
	(22.69)	(22.69)	(.21)	(.42)

5.5 Primary analyses

5.5.1 Average treatment effects

Table 5.10 presents the model coefficients for the primary analysis for attendance, our first outcome of interest. Column 1 reports a simple regression model, regressing only the treatment indicator on the outcome variable for students with complete outcome data (complete case analysis; CCA). Column 2 provides the model coefficients for the simple model using the multiply imputed (MI) dataset. Column 3 provides the complete case analysis (CCA) and includes additional covariates. The pre-registered covariates are as follows: student age, gender, subject, resit status and college-level fixed effects. Finally, column 4 provides the results from the multiply imputed dataset, including the same covariates.

Overall, the results tell us that the supportive communication intervention had no significant average treatment effect on class attendance. Across all four models, none of the estimates are statistically significantly different from zero. The best performing treatment, where study supporters received text messages, resulted in a 0.05 to 2.5% improvement in attendance rates, depending on model specification. The treatment coefficient for the 'student only' group is negative across models and indicates that the attendance rate for students allocated to this group was lower than that of control students. These estimates are not statistically significant, and imprecisely estimated. The models utilising additional covariates are similar to the simple models, and the multiple imputation of missing attendance does not change inferences made, either. Effect sizes for the four treatment arms in the simple model are reported in Appendix 31 (p. 276). The effect sizes, ranging between Hedges' g = -0.07 and 0.05 for the different treatment arms, are small both in absolute terms and in comparison to the effect sizes observed in Chapter 4. Overall, these results tell us that the supportive communication intervention had no significant average treatment effect on class attendance.

Table 5.10: Average treatment effects of the intervention on attendance rates

	(1) Simple CCA	(2) Simple MI	(3) Inc. covariates CCA	(4) Inc. covariates MI
	β (SE)	β (SE)	β (SE)	β (SE)
Supporter only	0.005	0.006	0.025	0.024
	(0.021)	(0.021)	(0.017)	(0.017)
Student only	-0.017	-0.014	-0.004	-0.003
	(0.022)	(0.022)	(0.018)	(0.018)
Supporter and Student	0.010	0.011	0.016	0.015
	(0.021)	(0.021)	(0.018)	(0.017)
Content-based licensing	-0.006	-0.009	-0.008	-0.007
	(0.034)	(0.034)	(0.025)	(0.025)
Constant	0.721**	0.725**	0.210**	0.210**
	(0.015)	(0.015)	(0.053)	(0.053)
Student-level covariates	*No*	*No*	*Yes*	*Yes*
College fixed effects	*No*	*No*	*Yes*	*Yes*
Control mean	0.721	0.725	0.712	0.717
Observations	923	975	923	975
R-squared	0.002	0.002	0.344	0.350

Notes: The columns report the intent-to-treat (ITT) estimate and robust standard errors (in parentheses) of individual-level random assignment on class attendance. Attendance is calculated using daily attendance registers and is recorded on a scale between 0 and 1 as the proportion of classes attended throughout the full academic year. Student-level covariates include age, gender, subject (maths/English), resit status and baseline attendance (first 9 weeks of the academic year). Thirty imputations were conducted using chained equations (MICE).

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

I now turn to the average treatment effects of the intervention on the probability that students achieved their GCSE maths or English. OLS estimates are presented first. The simple model is reported in column 1 and 2 using complete case analysis (CCA) and multiple imputation (MI), respectively. Column 3 reports only on the students who completed the baseline achievement test (69.1% of the sample) and includes student-level covariates and college fixed effects. Finally, column 4 reports the additional covariates model using the multiply imputed dataset.

Model (3) was pre-specified as the primary model of interest, but it was not anticipated that such a large proportion of students would fail to complete the baseline achievement test. Particularly few students at College C completed the baseline achievement test. Scores are missing for 65.7% for this sample, in comparison to 12.1% – 27.5% in the remaining three colleges. It is not uncommon for tutors to prioritise tasks other than this assessment, so a students' failure to take the test is not necessarily a reflection of their lack of punctuality or motivation. Anders, Dorsett and Stokes (2018) experienced similarly patchy baseline achievement data in a randomised controlled trial in further education colleges, where only 30% of English and 42% of maths learners had completed the pre-assessments (p. 30).

Table 5.11: Average treatment effects of the intervention on achievement rates

	(1) Simple CCA	(2) Simple MI	(3) Inc. covariates CCA	(4) Inc. covariates MI
	β (SE)	β (SE)	β (SE)	β (SE)
Supporter only	-0.012	-0.013	0.032	-0.013
	(0.037)	(0.037)	(0.043)	(0.036)
Student only	0.042	0.041	0.080+	0.041
	(0.039)	(0.039)	(0.044)	(0.038)
Supporter and Student	0.074+	0.074+	0.089*	0.057
	(0.040)	(0.040)	(0.044)	(0.038)
Content-based licensing	0.076	0.067	0.060	0.056
	(0.063)	(0.062)	(0.065)	(0.059)
Constant	0.203**	0.203**	-0.028	-0.095
	(0.027)	(0.027)	(0.105)	(0.137)
Student-level covariates	*No*	*No*	*Yes*	*Yes*
College fixed effects	*No*	*No*	*Yes*	*Yes*
Control mean	.203	.203	.171	.207
Observations	970	975	672	975
R-squared	0.007	0.007	0.114	0.096

Notes: The columns report the intent-to-treat (ITT) estimate and robust standard errors (in parentheses) of individual-level random assignment on qualification achievement rates. Student-level covariates include age, gender, subject (maths/English), resit status and baseline achievement as measured by the bksb diagnostic assessment, and prior GCSE grade. Thirty imputations were conducted using chained equations (MICE).

+ = p < 0.10, * = p<0.05, ** = p<0.01.

Across all four models, the 'student and supporter' arm of the intervention performs best, with improvements in achievement rates from 5.7 to 8.9 % points, depending on model specification. Both simple models in column 1 and 2 estimate a positive treatment effect but fail to meet conventional benchmarks for statistical significance (both p = 0.066). Although the treatment coefficient fails to reach the benchmark of p < 0.05, its effect size is meaningful. Using the simple model in column 1 for parsimony, on average 20.3 % of students in the control group pass their qualification. Those assigned to 'supporter + student arm' were 7.4 % points more likely to achieve their qualification, at 27.6%. The simple model (column 1) estimates an effect size for the 'supporter + student' group of Hedges' g = 0.173, which is comparable to the observed effect size in Chapter 4.[44] Effect sizes for the simple model are reported in Appendix 31, p. 276).

The results of column 3 warrant a closer examination. The third model produces a statistically significant effect for the 'student + supporter' arm (p = 0.043) but its sample is restricted to a non-random sample of the total participant pool. Upon closer inspection, this effect is likely attributable to a college effect. Students at College C are both more likely to be missing baseline achievement scores, and to achieve a good pass. As can be seen from Table 5.2, average achievement rates lie considerably higher for this college in comparison to the other participating colleges (p. 143). In the analysis of complete cases, the average achievement rate of control group students is lower, and treatment effects become more pronounced. Alternatively, the larger treatment estimates reported in Column (3) could be a function of unmeasured student characteristics. Although it cannot be tested empirically, it is possible that students who sit the baseline achievement test benefit more from the intervention (e.g. if sitting the baseline test is a proxy of motivation to learn, and students with greater motivation may benefit more from this type of intervention).

The other treatment groups produce only modest effects. The treatment coefficient of the 'supporter only' arm is negative across models, but not statistically significantly so. The 'student only' group produces modest positive effects that fail to reach statistical significance, potentially due to the underpowered nature of this comparison. The effect of the 'student-only' arm is more pronounced when baseline achievement is controlled for (column 3; p = 0.065). Finally, the content-based licensing arm produces encouraging results as the effects are similar in size as the best-performing treatment. This pilot arm is similar to the 'supporter + student' treatment both in design and estimated effect size.

[44] The study reported in Chapter 4 (which corresponds to the 'supporter only' group) produced effect sizes of Hedges' g = 0.163 (simple model) and Hedges' g = 0.136 (model including covariates).

Although the pilot treatment estimates do not reach significance ($p < 0.05$), the estimated 6 - 7 percentage point improvement in achievement rates warrants further testing. The pilot texts were delivered both to student and supporter, and the only difference with the 'supporter + students' texts was the occasional inclusion of puzzles which could only be solved if both parties spoke to each other about the text message. Students could win small prizes if they replied with the correct answer. This experimental design cannot disentangle the relative impact of curiosity-inducing information versus financial incentives. It is possible that the half-solved puzzles were sufficiently curiosity-inducing for students and supporters to bring up the topic. On the other hand, winning two cinema tickets might be the push students needed to bring up the text messages. Future studies could begin to tackle these questions.

It should be noted that the content-based licensing treatment has a sample size of only 61 students with complete outcome data. This group is small, especially in comparison to the other arms. The robust Huber-White standard errors used in this analysis rely on large samples for their validity (N > 50; Imbens & Kolesár, 2016). To assess potential problems with the conventional robust standard errors reported in this chapter, HC3 standard errors are calculated in line with the recommended strategy when sample sizes are small (Long & Ervin, 2000; Mackinnon & White, 1985). Using the HC3 corrected standard errors our treatment estimates are similar, suggesting that the results are not affected by the small sample size.

Overall, these results tell us that not all versions of the supportive text-messaging intervention produce the intended positive effects on qualification achievement rates. As such, this study provides mixed support for Research Question 1. In line with the secondary research question, however, I find larger average treatment effects for the arm where text messages were sent to students directly as well as supportive others in their social networks. Surprisingly, the arm that produced significant results in Chapter 4, the 'supporter only' arm, fails to have a positive effect in this study. This contrary finding is further considered in this chapter's discussion, section 5.8.

5.6 Secondary analyses

5.6.1 Heterogeneous treatment effects

Due to the limited sample size of this experiment, only two subgroups of interest are examined, namely gender and subject. Chapter 4 showed that receiving the Study Supporter intervention improved attendance only for male students, whereas the intervention only improved achievement rates for female students. Chapter 4 was unable to shed light on heterogeneous treatment effects by subject due to its experimental design: whether students were treated for English or maths was assigned at the college level. This study was able to examine heterogeneous treatment effects for subject as students were randomly allocated to receiving the text messages about maths or English class.

In Table 5.12 the analysis is partitioned depending on whether students are treated in a maths class (Column 1) or English class (Column 2), and whether they are identified as female (Column 3), or male (Column 4). Pre-specified student-level covariates college fixed effects are included in each model. A statistically significant impact (p = 0.009) of the 'Supporter only' arm is found for the attendance rates of male participants; an increase of 6.1% points from a control mean of 69.3%. The other treatment arms do not result in a statistically significant improvement in attendance. Again, I find no effect of the intervention on attendance rates for female participants (p > .05). Further, no significant effects of the treatments on attendance partitioned by subject are observed.

Table 5.12: Heterogeneous treatment effects on attendance

	(1) Maths β (SE)	(2) English β (SE)	(3) Female β (SE)	(4) Male β (SE)
Supporter only	0.021	0.031	-0.012	0.061**
	(0.023)	(0.026)	(0.025)	(0.023)
Student only	-0.008	0.003	-0.024	0.016
	(0.025)	(0.025)	(0.024)	(0.026)
Supporter + student	0.007	0.030	-0.014	0.039
	(0.026)	(0.024)	(0.025)	(0.024)
Content-based licensing	-0.032	0.022	-0.004	-0.011
	(0.034)	(0.039)	(0.037)	(0.035)
Constant	0.076	0.280**	0.193**	0.172*
	(0.081)	(0.066)	(0.067)	(0.081)
Student-level covariates	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes
Control mean	0.699	0.724	0.733	0.693
Observations	481	442	449	474
R-squared	0.365	0.328	0.405	0.310

Notes: All analyses are OLS regressions and include fixed effects at the college level. Student-level covariates include age, gender, subject (maths/English), resit status and baseline attendance. Robust Huber white standards errors in parentheses. + = p < 0.10, * = p<0.05, ** = p<0.01.

Table 5.13 displays the partitioned analysis for the second primary outcome of interest, achievement of the GCSE qualification. As above, Column 1 conducts the analysis for students treated in Maths class, Column 2 for English, Column 3 for students identified as female and finally, Column 4 for male students. Baseline achievement is multiply imputed, and all other pre-treatment covariates were non-missing for the full sample.

The treatment effects differ by subject as well as gender. Column 1 displays a negative effect of the 'supporter only' arm on achievement for maths albeit only significant at the $p < 0.10$ level ($p = 0.062$). Conversely, column 2 displays a statistically significant positive effect of the 'supporter + student' on pass rates in English class ($p = 0.023$). In this arm, treated students' probability of passing the course is relatively high at 30.8%, in contrast to the control mean of 19.3%. The other variations of the interventions do not impact attainment significantly within the subject subgroups. Second, I find no effect of the intervention for female participants, but significant effects for male participants in the 'Student only' (p = 0.005) and Supporter + Student' (p = 0.006) groups. When comparing male and female students directly, it is evident that female students achieve better scores on average. The probability of obtaining a pass grade increases considerably for male students assigned to the 'student only' or 'supporter + student' groups, from 11.9% from the control group mean, to 25.8% and 25.6%, respectively. These estimates correspond to a 115% increase, from a particularly low baseline.

Table 5.13: Heterogeneous treatment effects on achievement

	(1) Maths β (SE)	(2) English β (SE)	(3) Female β (SE)	(4) Male β (SE)
Supporter only	-0.091^{+}	0.066	-0.061	0.025
	(0.049)	(0.055)	(0.057)	(0.045)
Student only	0.032	0.050	-0.075	0.138**
	(0.052)	(0.054)	(0.056)	(0.050)
Supporter + student	-0.003	0.115*	-0.037	0.137**
	(0.052)	(0.056)	(0.057)	(0.050)
Content-based licensing	-0.017	0.154	-0.051	0.163^{+}
	(0.080)	(0.093)	(0.089)	(0.084)
Constant	-0.012	0.016	0.104	-0.036
	(0.126)	(0.120)	(0.116)	(0.130)
Student-level covariates	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes
Control mean	0.222	0.193	0.308	0.119
Observations	498	472	476	494
R-squared	0.090	0.114	0.112	0.095

Notes: All analyses are OLS regressions and include fixed effects at the college level. Student-level covariates include age, gender, subject (maths/English), resit status and baseline attendance. Robust Huber white standards errors in parentheses.+ = p < 0.10, * = p<0.05, ** = p<0.01.

As a robustness check, logistic binary regressions for the average and heterogeneous treatment effects on achievement rate are reported in Appendix 32 (p. 277) and Appendix 33 (p. 278), respectively. These estimates are consistent with the findings reported above, with the exception of the subgroup of male students in the content-based licensing (pilot) arm. The treatment effect is statistically significant (p = 0.032) in the logistic regression model while just failing to reach statistical significance in the primary regression model. The average treatment effects fail to reach statistical significance, but the sign and size of the treatment coefficients are comparable. In summary, I find that the results are consistent across the linear probability and logistic regression models.

5.6.2 Assessing the effect of one-sided noncompliance on treatment effects

The primary intention-to-treat (ITT) analyses focus on the effect of ‘prescribing communication’, rather than the effect of students’ and supporters’ engagement with the text message outreach. Due to detailed delivery statistics obtained from the text-messaging platform, it is possible to observe whether the intended recipients do receive the scheduled messages. Recipients can opt-out of receiving the texts or change their mobile number (and fail to inform us of the change), and in both cases treatment ceases at this point. The dose of treatment therefore varied between participants. Additionally, forty-one intended recipients never received the first text due to having changed phone numbers between sign-up and launch of the experiment. The ITT analysis may therefore underestimate the efficacy of the supportive communication intervention for those who comply with treatment, as the effect is diluted by non-compliance. If this were true, the CACE estimates would be positive and significant for the subset of people who were treated, in line with hypothesis 1 and 2 (see Table 3.1, p. 78). To assess the effect of non-compliance on the two primary outcomes, the Complier Average Causal Effect (CACE) is estimated. The model for the estimation of CACE was introduced in section 5.2.8.3, p. 159.

Half of the participants assigned to either one of the three treatment groups were fully compliant with the treatment (N = 375, 49.9%). Table 5.14 presents the results of the 2SLS regression, where Column 1 provides the first-stage regression and Column 2 and 3 the two primary outcomes of interest.

Table 5.14: Instrumental variable estimates of CACE estimates

Variables	*Compliance: intended recipients received full schedule of 35 texts*		
	(1) First stage	(2) Attendance rate	(3) Achievement rate
Supporter only - assigned	0.556+		
	(0.023)		
Student only - assigned	0.561+		
	(0.023)		
Supporter + Student - assigned	0.432		
	(0.023)		
Supporter only – alerted		0.049+	-0.044
		(0.029)	(0.062)
Student only - alerted		-0.002	0.050
		(0.030)	(0.064)
Supporter + Student - alerted		0.041	0.104
		(0.038)	(0.083)
Student-level covariates		Yes	Yes
Observations		923	970
Control mean		0.721	0.203

Notes: Treatment effects on primary pre-registered outcome variables are estimated using 2SLS regressions with the instrumented alerted variable, an indicator for students and supporters who received the full schedule of 35 texts. All regressions include a set of student-level demographic covariates, of gender, age, subject, and resit status. Baseline achievement was not available for 30% of our sample and therefore not added into the model in order to preserve sample size. Robust standard errors (in parentheses). + = p < 0.10, * = p<0.05, ** = p<0.01.

The effect of receiving the full schedule of text messages is not statistically significant at the benchmark of $p < 0.05$, neither for attendance nor achievement. I first turn to the CACE estimates on attendance rates. Column (2) shows that receiving the full schedule of texts increases attendance by 4.9% points for students in the supporter only group ($p = 0.09$). This finding does not hold for students who only received texts themselves ('student only group') or where both parties receive the texts 'supporter +student arm'. Yet, the findings suggest that compliers fared better than those who were offered the treatments but who did not or only partially receive the text messages.

The CACE estimates of receiving texts on achievement follow similar patterns, where two of the three treatment groups produce positive effects on achievement rates but not statistically significantly so. Receiving the full dose of SMS in the 'supporter + student' arm results in a 10.4 % point increase in achievement ($p = 0.213$), but the evidence remains inconclusive. The precision of the treatment estimates is low, and the instrument appears to be relatively weak. Compliers in the 'student only' group fared better than those in the 'supporter only' group, a difference of almost ten percentage points. Minimal and optimal thresholds of compliance are explored as bounds of the treatment effects in Appendix 34 (p.279).

5.6.3 Assessing contamination due to within-classroom nominations

As discussed in more detail in section 5.3.4, a total of 96 students appeared in the data both as a student and nominated study supporter of another student in the data set. Of these 96 students, 55 received a different treatment than originally allocated (e.g. they were allocated to control but received texts as a supporter for another student). In order to assess the influence of such contamination on the treatment estimates, these students were re-allocated to treatment groups based on their observed treatment. Those who should have been in the control group but received texts as supporters are re-allocated to the 'student only' arm (N = 22) since they ended up receiving texts about their learning. These text messages referred to the classmate who nominated them, but since texts are tailored to the subject rather than the individual, the informational value was identical.

Second, in instances where only the students' supporters should have received texts but said students also received texts in the role of supporter were allocated to the 'student + supporter' group (N = 33). Finally, forty-one students who had nominated themselves as supporter (albeit sometimes under a different name, the phone number was identical) were re-allocated. Within this group, students who were originally assigned to 'supporter only' or 'supporter + student texts', are re-allocated to the 'student only' arm, as no one other than themselves received the intended communications. In total, 80 students are re-allocated to treatment groups. See Appendix 35 for a breakdown of re-allocated treatment groups (p. 280). Sensitivity checks were performed by (1) reassigning them to the observed treatment and (2) removing these non-complying students from the regression. Appendix 36 to Appendix 39 display these results (p 281 - 284). None of the inferences made from the primary regressions are altered by the re-allocation or removal of contaminated students, which provides evidence for the structural validity of treatment estimates.

5.6.4 Modelling dropout dynamically

Thus far, the quantitative analyses have focused on attendance rates in maths and English. Class attendance is an important predictor of academic achievement (Credé et al., 2010), and is an outcome measure in many parent-school interventions that inspired this thesis (e.g. Berlinski et al., 2016; Rogers et al., 2017). The variable is typically of interest because it can be easily obtained through administrative datasets and because it is a behavioural proxy of student (dis)engagement (Cabus & De Witte, 2012). Every class students fail to attend is marked on the attendance register, which produces a detailed and longitudinal dataset. Rather than focusing on the average attendance rate only, the dynamic nature of this rich dataset can be utilised to explore variation in the time to drop out across the four treatment arms.

Early school withdrawal is generally understood to be a dynamic, multidimensional and cumulative process; driven by individual and institutional factors (De Witte, Cabus, Thyssen, Groot, & Van Den Brink, 2013; Rumberger et al., 2017; Rumberger & Lim, 2008). Theoretical models consider the lack of self-esteem (Finn, 1989) and lack of social and academic integration with the institution (Tinto, 1975) as primary reasons for school leaving. Although truancy is predictive of long-term disengagement with college (Cabus & De Witte, 2012) the vast majority of students in our sample are absent from class intermittently (the average attendance rate is 79%). A much smaller proportion of the sample drop out altogether.

The intervention discussed in this chapter could reverse the cumulative process of disengagement through positive, actionable and timely communication between the college, student, and their social network. Similar text messages geared towards increasing students' sense of social belonging with the college have proven effective at improving course completion rates (Chande et al., 2017).

In order to capture the more severe disengagement, a duration modelling approach is applied to the rich attendance dataset. It is used to explore whether the intervention had an effect on the rate at which students drop out from the course. Duration models are only occasionally used in education research yet they are particularly adept at handling chronological data and exploring the timing of events (for example, see Alcott, 2013; Anders, 2017;

Bradley & Lenton, 2007; Cabus & De Witte, 2012). Section 1 summarises the relevance of dropout as an outcome measure, section 2 sets out the data structure and modelling approach, and section 3 presents the findings.

5.6.4.1 Why focus on dropout rates?

College dropout is a relevant outcome measure in its own right. First, dropout rates are of policy concern. The retention of students is taken into account in the calculation of funding allocated to each post-16 institution in England. Naturally, total student numbers are taken into account, but whether enrolled students (1) stay at least the first 42 days, the qualifying period, and (2) stay until the end of the anticipated qualification end date ultimately determines the total funding colleges receive for their cohort of students on GCSE qualifications (Education and Skills Funding Agency, 2018). If students fail to stay enrolled during the first six week, the college receives no funding. If the student stays beyond the qualifying period but drops out before the end of the academic year, the college receives only 50% of the full funding rate. It is no surprise, therefore, that college staff dedicate a considerable amount of resources to ensure course completion.

Second, dropout is predictive of poor life outcomes for young people. School dropouts have higher rates of unemployment (Belfield & Levin, 2007) and are more likely to commit juvenile crimes (Belfield & Levin, 2009) than their peers who stayed in education. In 2016/17, 85.5% of FE college students who reached the end of their 16-18 studies were retained until the end of their vocational programme of study (Department for Education, 2018). A recent analysis of year-on-year dropout showed that 25% of GCSE catch-up students who fail to catch up in year 12 (i.e. at further education or sixth form colleges) do not return to retake their maths or English (Impetus PEF, 2017). These students are left without a Level 2 qualification in English and maths, which may limit their employment opportunities (Machin et al., 2018).

Finally, dropout from schooling can be predicted from students' patterns of truancy over time. Using a duration modelling approach, Cabus and De Witte (2012) show that the risk of dropout for students with a truancy record is 37.4 % higher than their regular attending peers in vocational education or training. Therefore, interventions that are effective at reducing school absenteeism may in the long term reduce eventual dropout.

5.6.4.2 *Preparing the data*

The data was structured as follows. Week-by-week attendance data was collected via college administrative data for the 922 students we have complete attendance data for. Absences are noted for every possible session, for the full academic year. I take the first week of November 2016 as the starting point for the analysis of dropout behaviour, when randomisation took place. Since the participating colleges did not use specific markers to signal student drop out, a student is recorded as a dropout if she fails to attend the treated maths or English class four weeks in a row after the point of randomisation. It can be argued that missing four weeks of teaching results is a reasonable marker of disengagement with the course. Re-entry is not examined in the duration model; once students fail to attend four weeks in a row, they are coded as having dropped out for the rest of the academic year. In total 116 unique students dropped out, which corresponds to 12.6% of the sample. It should be noted that students only signed up in week 5 and 6 of the academic year, at which point the most disengaged students may have already dropped out of the course.

The descriptive statistics displayed in Table 5.15 show that baseline differences between continuing and dropout students are not as obvious as might have been expected. No differences are found in the gender and age composition between regular attenders and dropouts, nor on baseline achievement on the bksb diagnostic test administered at the start of the academic year. Not surprisingly, those who eventually drop out have worse attendance in the weeks leading up to randomisation than students who complete the course. Dropouts are more likely to indicate they are resitting at an FE college for the first time, but this difference between the groups fails to meet the threshold for statistical significance ($p > 0.05$). The proportion of white students is considerably higher in the dropout group, which is compatible with studies showing that particularly White British working class students are at-risk of disengagement (House of Commons Education Committee, 2014; Strand, 2014).

Table 5.15: Baseline characteristics by dropout status

Characteristic	Continuing	Dropout	*p*-value
Gender: male (proportion)	0.515	0.509	0.900
	(0.018)	(0.047)	
Age (in years)	18.928	18.975	0.933
	(0.193)	(0.573)	
Ethnicity: white (proportion)	0.328	0.448	0.010
	(0.017)	(0.046)	
Baseline score on bksb test (proportion)	0.632	0.644	0.627
	(0.009)	(0.022)	
Baseline attendance (proportion)	0.835	0.605	0.000
	(0.006)	(0.030)	
Resit status: first time retake at FE college (proportion)	0.422	0.509	0.078
	(0.017)	(0.047)	
N	806	116	

Notes: Standard errors in parentheses. Aside from age and baseline attendance which are continuous variables, all characteristics report proportions. The final column reports *p*-values from a t-test of the null hypothesis of no difference in characteristics between dropout and continuing students.

To assess the sequence of attendance, non-attendance and ultimate dropout, a hazard-based duration model is constructed. Duration models are able to examine the timing of 'failures' (i.e. drop out) dynamically and flexibly. Additionally, duration models take into account censored data (Box-Steffensmeier & Jones, 2004, Ch. 2, p. 8). For example, one college provided only 31 weeks of data and two colleges provided data for 33 weeks. The fourth college provided additional data for the Summer term, whereas the other colleges did not (i.e. this final college provided data for in total 40 weeks). Students could drop out from their course between week 31 and the end of the academic year, thus the end point of a spell (i.e. from attendance to dropout) is not observed.[45] The data from the three colleges are right-censored, as students in three colleges are not observed after week 31 and 33, respectively.

Attendance is observed on a weekly basis for the duration of the academic year. This data is handled as continuous-time rather than discrete-time data.[46] The observations enter the analysis when the randomisation was performed and do not re-enter after dropout; the maximum number of spells

[45] The results are not affected by censoring all colleges to 31 weeks, so we proceed with the full observed data (i.e. we keep the observations beyond week 31).

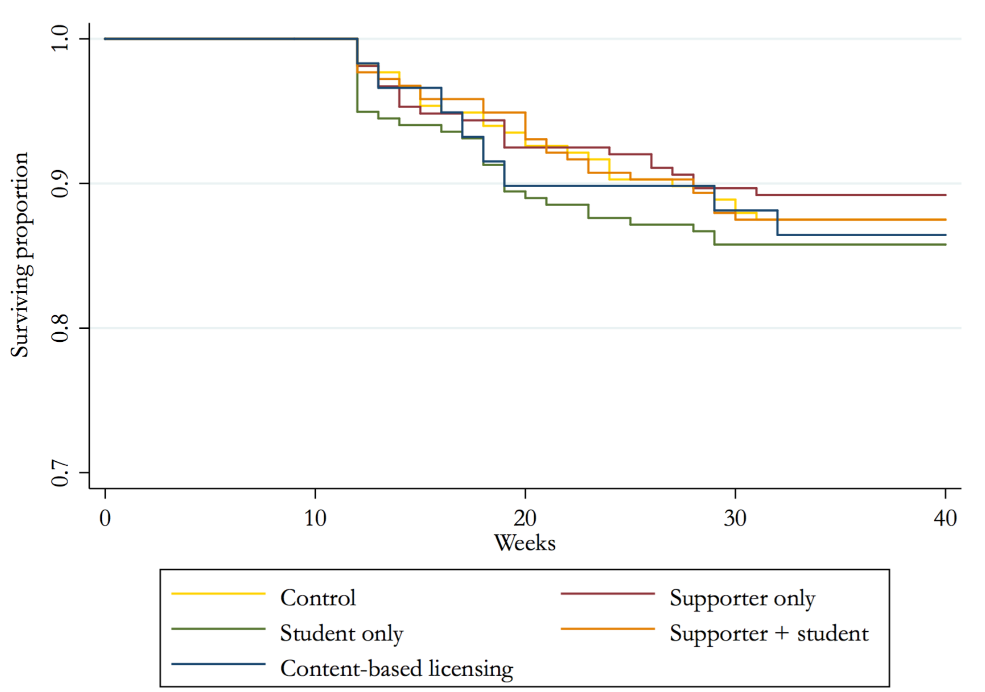
[46] As a robustness check, treating the data as discrete by using a complementary log-log regression did not affect the results.

per person is 1. The outcome of interest is the hazard rate, or the "probability that the observation will fail at a certain moment in time (t)" (Cabus & De Witte, 2012, p. 6), conditional on it not having occurred before time t. The eventual event of interest is dropout, at four consecutive weeks of non-attendance. The clock only starts ticking in week 9, hence the first possible failure occurs in week 12. In line with Bradley and Lenton (2007), all sequential observations (i.e. weeks) are coded as 0 until a dropout occurs, which is coded as 1.

5.6.4.3 Results

First, the survival probabilities of students across treatment arms are plotted using the nonparametric Kaplan-Meier estimator (1958). Since attendance is reported in a discrete week-by-week fashion, the graph visualises failure by a step-wise function (Cabus & De Witte, 2012). Figure 12 displays the differences in relative risk of cumulative incidence of dropout between the three treatment groups and control group.

Figure 12: Kaplan-Meier curve, cumulative hazard of dropout



Notes: Students are treated as 'at risk' from week 9 when randomisation occurred. The first possible failure occurs in week 12, four weeks after the intervention launched. The plot time starts at the start of the academic year, but failures occurring before randomisation are not taken into account.

Figure 12 shows that the proportion of students who are classified as having dropped out does not differ strongly between treatment arms. This is perhaps not surprising, since the primary analysis found no discernible effects of the intervention on average attendance rates (see Table 5.10, p. 175). The non-parametric log-rank test also finds no difference between survival curves of the treatment arms (*Chi-square* = 1.27, p = 0.87), therefore the null hypothesis of no difference between the probability of dropout between treatment arms cannot be rejected.

Next, inference testing is performed using a semi-parametric Cox proportional hazard model. This makes the assumption of proportional hazards, or the "effect of any covariate having a proportional and constant effect that is invariant to when in the process the value of the covariate changes" (Box-Steffensmeier & Jones, 2004, Ch. 8, p. 131). This model controls for the same pre-treatment covariates controlled for in the primary analyses; age, gender, resit status, subject, and college-specific dummy variables. The proportional hazards assumption holds for students in our sample, suggesting that the baseline hazard of dropout is constant over time (*Chi-square* = 368; *Prob* > *Chi-square* = 0.451).

Table 5.16 presents the estimation results. Model 1 displays the basic Cox model (with no covariates other than the treatment arm binary variables), Model 2 controls for the pre-treatment covariates discussed above. The estimated hazards are slightly lowered in the 'supporter only' and 'supporter + student' conditions; students assigned to these groups are slightly less likely to drop out in comparison to the control group, but this is not statistically significant ($p > 0.05$). Additionally, the practical significance of the effects is negligible, especially for the 'supporter + student' group where students have a 1.4% lower likelihood of drop-out than their control peers. Students assigned to the 'supporter only' treatment have a 12.8% lower likelihood of dropout compared to students assigned in control (*n.s.*). The remaining two treatment arms appear to increase the hazard of dropout, but not statistically significantly so. The 'student only' treatment appears to have been least beneficial to students. Controlling for pre-treatment covariates, students who receive the weekly texts directly and without the mention of a study supporter have a 22.8% higher hazard of dropout than students in the control group (*n.s.*). Holding the other factors constant, students on maths courses are significantly more likely to drop out from the course (p = 0.021) than participating students on English courses.

Table 5.16: Estimated hazard ratios of dropout by treatment group

	(1) Cox: basic	(2) Cox: inc. covariates
Supporter only	0.864	0.872
	(0.245)	(0.249)
Student only	1.166	1.228
	(0.307)	(0.324)
Supporter + Student	0.999	0.986
	(0.272)	(0.270)
Content-based licensing	1.095	1.088
	(0.441)	(0.440)
Gender: male		0.992
		(0.187)
Age		1.007
		(0.017)
Subject: maths		1.570*
		(0.306)
Resit status: first time		1.424+
		(0.273)
College-specific dummies	No	Yes
Person-time observations	26,142	26,142
Observations	922	922
No. Failures	116	116

Notes: Standard errors in parentheses. Hazard ratios (i.e. exponentiated coefficients from the underlying Cox regression model) are displayed. Stars indicate statistical significance as follows: ** p<0.01, * p<0.05, + p<0.1

In summary, students across the treatment arms have a probability of dropping out from their courses over the analysis time that is not distinguishable from their control group peers. Adding control variables does not significantly change the estimated hazard ratios, and neither does tightening (5 weeks; 96 failures) or relaxing (3 weeks; 185 failures) the rule for defining dropout. A joint significance test of the treatment dummy variables shows that the null-hypothesis of no difference in survival rates between the treatment arms is not rejected (p = 0.80). These results suggest that the supportive communication intervention did not reduce the rate of dropout.

5.7 Qualitative methods and results

5.7.1 Rationale

This thesis combines the rigour of randomised controlled trials with the in-depth and nuanced exploration of the lived experiences of participants using interpretative qualitative approaches (see Chapter 3, section 3.6 for a review of the literature on mixed methods, p. 66 - 69). The process and implementation evaluation of the Study Supporter intervention, reported in Chapter 4, suggested that the intervention benefits students who nominated study supporters whom (1) they felt close to and communicated with regularly, (2) provided emotional support rather than instrumental support, and (3) actively asked questions and were interested in the student's experience at college.

The qualitative component also suggested that the supportive text messages may not lead to improved communication when students nominate weak relationships (such as classmates they met at the start of the academic year), and when these relationships were not emotionally supportive to start with. Therefore, the present qualitative analysis focuses on understanding the quality of the relationship between student and Supporter.

5.7.2 Research questions

The primary purpose of the previous chapter's qualitative component was to help inform and design the recruitment and consent procedures for the trial introduced in this chapter. As foreshadowed, this second wave of qualitative inquiry focuses particularly on exploring the barriers to and facilitators of engagement with the intervention content. Second, the qualitative data is used to understand whether students experienced benefits beyond those captured through college administrative datasets. Students as well as nominated supporters were interviewed to gain a better understanding of the process of seeking and providing social support from both the student and supporter perspective.

Table 5.17: Qualitative research objectives, Project Success trial

Qualitative objective	*Qualitative research questions*
To document the experience of resitting GCSE maths or English.	How do students describe their learning experience in maths and English?
To explore barriers to, and facilitators of engagement with the intervention.	What factors influence students' openness to the intervention?
To assess the feasibility and acceptability of the intervention in practice.	Do student welcome supportive communication about their college experience?

5.7.3 Sampling approach and data collection

The qualitative data collection was carried out alongside the delivery of the intervention in order to explore its implementation in depth and assess changes in student perceptions over time. The qualitative data was collected in two phases. First, students were interviewed one month after the launch of the text messaging programme (December 2016; Phase 1). Second, study supporters were invited for a phone interview post-intervention, in June 2017 (Phase 2).

Since the qualitative component of this second field experiment focused explicitly on the relationship between students and their nominated study supporter, students in the 'supporter only' and 'student + supporter' groups were interviewed. The quality and quantity of their interactions about the weekly text messages was the focus point of the topic guide. Thus, interviewing students who only received the texts themselves (i.e. 'student only' group) or did not receive any texts at all (i.e. control group) would not help address the primary question of interest. In summary, the goal of this qualitative study is not to explore how students across different conditions experience college and support, but rather, to zoom in on the absence or provision of supportive communication between student and supporter.

5.7.3.1 Phase 1 sample selection

A purposive sampling strategy was employed. Only four colleges participated in the present experiment, and therefore all colleges were selected for qualitative fieldwork. As it proved more difficult than expected to recruit a sufficient number of student in Chapter 4 due to timetabling issues and dropout or non-attendance, students were again oversampled by 50%. In total, 30 students were selected with the aim to conduct fifteen interviews. Since College 3 and 4 recruited a larger proportion of students to take part in the intervention (73% of the sample combined), 20 students were selected across these two colleges, and the remaining 10 students from College 1 and 2.

Students were purposively selected using the following criteria:

1. Treatment condition; even split between 'supporter only' and 'supporter + student';
2. Type of study supporter nominated; spread of family members, peers and partners (i.e. the most common categories);
3. Gender of student; even split between males and females; and,
4. Subject in the trial; even split between maths and English.

After drawing a purposive sample, college administrators were asked to contact the students on the list. The administrators then scheduled individual interviews with students who expressed an interest in taking part. The interviews were conducted at the college and lasted approximately forty-five minutes. Students were authorised to leave the classroom to take part in the interview. In total fifteen in-depth interviews were carried out across the four colleges. Students signed a consent form after reading an information sheet, and all interviews were recorded and transcribed verbatim.

Demographic characteristics of the qualitative sample are displayed in Table 5.18. The final sample of interviewees was somewhat unbalanced on subject; 60% of interviewees received the intervention in their English class. Seven student interviewees were women and eight were men. Nominated study supporters were family (N = 7), peers (N = 4), partner (N = 2) or other (N = 2). All but one student were resitting their GCSE qualification after failing to obtain a passing grade the previous year. Finally, the majority of students were aged between 16 and 19 (N =17). The high proportion of 16-18 year olds is a reflection of the distribution of the student population at the participating colleges. Balance on treatment conditions was not achieved. More students assigned to 'student + supporter' texts accepted the interview invitation (N = 9) than those in the 'supporter only' group (N = 4). Due to unforeseen challenges with students' timetables, two students (allocated to 'student only' and control) were convenience sampled at the college when the original interviewees were not able to attend.

Table 5.18: Participant characteristics of phase 1 interviewees

						Text delivery statistics		**Student demographics**		
Interview ID	*College*	*Pseudonym*	*Treatment group*	*Subject in trial*	*Relationship with supporter*	*% successful, student*	*% successful, supporter*	*Gender*	*Age*	*Ethnicity*
01LE01	1	Jack	Supporter + student	English	Brother, but turned out as self-nomination	100%	-	M	19	White British
01LE02	1	Max	Supporter + student	English	Mother	49%	100%	M	17	White British
01LE03	1	Hana	Supporter only	Maths	(Ex) Girlfriend	-	100%	M	16	Black British
02LE01	2	Wojtek	Supporter only	English	Girlfriend	-	100%	M	17	Other white
02LE02	2	Molly	Supporter + student	English	Grandmother	97%	100%	F	16	White British
02LE03	2	David	Supporter + student	Maths	Friend outside college	87%	25%	M	18	Indian
03LE01	3	Alex	Supporter + student	English	Friend outside college	100%	100%	F	16	Mixed
03LE02	3	Ivan	Supporter + student	English	Father	94%	91%	M	17	Other white
03LE03	3	Plus	Supporter only	Maths	Cousin	-	100%	M	19	Black African
03LE04	3	Isabelle	Supporter + student	English	Sister	100%	100%	F	16	White British
04LE01	4	April	Supporter + student	Maths	Mother	90%	100%	F	17	White British
04LE02	4	Rohan	Supporter + student	English	Friend inside college	0%	77%	M	17	Indian
04LE03	4	Priya	Supporter only	Maths	Mother of boyfriend	-	100%	F	18	Pakistani
04LE04	4	Zoe	Control	Maths	Mother	-	-	F	17	White British
04LE05	4	Lucy	Student only	English	Friend outside college	-	-	F	17	Black British

5.7.3.1 *Phase 2 sample selection*

The post-intervention interviewees were recruited using a different method. First, students interviewed in Phase 1 were re-contacted using a recruitment text message. Second, the supporters of the previously interviewed students were approached for a phone interview. These interviews would allow a deeper exploration of the relationship between the student and supporter and triangulate the student's responses. The response rate to these target recruitment activities was low. One previously interviewed student took part in the second interview, and two study supporters of previously interviewed students consented. Therefore, all texted study supporters received an invitation for a 30-minute phone interview. The sample was therefore constructed using convenience sampling, rather than the intended snowball sampling method. Four study supporters were recruited using the general invitation. The second phase of interviews was conducted over the phone.

The majority of the text messages focused on maths (N = 5) rather than English (N = 2). This balances the overall sample. Limited demographic information about the study supporters was collected; age, ethnicity, gender and occupation of study supporters are displayed in Table 5.19. Finally, five interviewees were assigned to the 'supporter + student' group and two to the 'supporter only' condition.

Table 5.19: Participant characteristics of phase 2 interviewees

								SMS delivery statistics		**Supporter demographics**		
ID	*Coll-ege*	*Type*	*Student also interviewed?*	*Pseudonym*	*Treatment group*	*Subject in trial*	*Relation-ship*	*% success, student*	*% success, supporter*	*Gender*	*Age*	*Occupation*
01SS01	1	Supporter	No	Bethany	Supporter + student	Maths	Classmate and friend	66%	98%	F	17	FE college student
02SS01	2	Supporter	No	Preethi	Supporter + student	Maths	Mother	100%	100%	F	49	Child practitioner
03LE02B	3	Student	Yes – 2nd interview	Ivan	Supporter + student	English	Father	94%	91%	M	17	FE college student
03SS01	3	Supporter	Yes, 03LE02	Thomas	Supporter + student	English	Father	94%	91%	M	46	Broadcast engineer
03SS02	3	Supporter	No	Louis	Supporter + student	Maths	Friend	100%	100%	M	18	FE college student, BTEC
04SS01	4	Supporter	No	Shannon	Supporter only	Maths	Mother	-	100%	F	-	Business owner
04SS02	4	Supporter	Yes, 04LE03	Latha	Supporter only	Maths	Boyfriend's mother	-	100%	F	39	Nurse

All interviews started were completed with no mention of discomfort over the interview questions, and none of the questions were skipped by participants. All interviews were audio recorded, labelled with a unique identifier, and sent to a transcription agency to be transcribed and de-identified. A total of 22 interviews were conducted at the four colleges ($N_{students}$ = 16, $N_{study\ supporters}$ = 6). All names are pseudonyms.

5.7.4 Topic guide

Two topic guides (one for each phase) were prepared before data collection to ensure consistency between the interviews. The topic guide was semi-structured and included primary questions which were addressed in every interview, and potential follow-up questions which served to clarify and follow-up responses. The guide was developed from an understanding of the literature on social support and the interviews conducted in the previous trial. The guide was piloted with two further education college students. The open-ended questions focused on specific themes, for example: how students seek extra maths and English support, what they expected of the programme at sign-up, how often they speak to the study supporter about their learning, and what made them feel this person cares about doing well at college. See Appendix 40 (p. 285) for the information sheet and consent form, and Appendix 41 (p. 286) for the interview topic list.

The student interview guide (first phase) also included an exercise which was designed to help visualise students' social network. An example is displayed in Appendix 42, (p. 287). The exercise clarified the perceived strength of the relationship between the student and their supporter, relative to others in their social network.

5.7.5 Analysis and interpretation

The qualitative component of the Project Success study follows the principles set out in Chapter 4 (section 4.8.5, p. 121). A thematic analysis approach was used, following the guidelines provided by Braun and Clarke (2006). An a-priori coding schedule was developed based on the research questions and themes identified in the literature and supplemented with inductive codes gathered throughout the coding process. The first four interviews were coded in this fashion, after which the coding framework was finalised. The remaining interviews followed this coding framework. Throughout the coding process, memos were used to track newly emerging themes and to maintain researcher reflectivity. Due to the scope and timing of the qualitative inquiry, it was unfeasible to employ multiple researchers to code the data in order to establish coding reliability. The coding framework was reviewed one month after initial coding to assess whether codes were comparable and reproducible.

The Consolidated Framework for Implementation (CIFR), which informed both the coding and analysis of the tutor interviews in Chapter 4, was not used in this follow-up phase. First, tutors were not interviewed for this study. Second, the research questions of chapter 4 focused primarily on programme implementation, whereas the current study focuses on individuals' narratives about seeking, receiving, and providing social support.

5.7.6 Results

Several themes emerged from the analysis. Those that serve to elucidate the qualitative research questions set out in Table 5.17 are addressed here, and the remaining themes are further addressed in Chapter 6.

Chapter 6 explores in more detail the typologies of social support provided, and whether students perceived the intervention as a potential social commitment device.

5.7.6.1 Struggling to engage with maths and English

The most common barrier to effective communication about the text messages in specific, or maths and English in general, was participants' reluctance towards studying these subjects. Out of the ten students who discussed their previous learning experience in maths or English, none recounted positive experiences in secondary school. The majority of participants (75%) also did not engage with the subject beyond the two hours of class time each week. The students argued that they were not required to complete coursework outside of classroom hours, and therefore did not feel the necessity to discuss the subject with their classmates. Those who still live at home with their parents or guardians (88% of full trial sample) described how the subject was not one they liked to touch upon. Isabelle, an English student said: "They ask me about it but I don't really go into conversations about it. I just feel like I've been there all day, so I just don't really want to talk about English and math much" *[03LE04]*. This reluctance towards the GCSE qualification appears to stem primarily from a perceived lack of autonomy. Students did not choose to study maths or English at post-16; they are required to take these subjects. One student described this as a general sense among the student population. When asked how many students are willing to learn, he said: “Thirty percent. The rest just turn up because they have to. And there [are] a few people in the classroom that have been doing English for five years, so it just gets really repetitive”*[01LE02]*.

Students spoke about disengaging with the lessons for a variety of reasons: the lack of structure in the classroom, a negative relationship with the teacher, boredom, and fear of maths. Such negative feelings and attitudes appeared to result in avoidance behaviour. David continued to attend class, but admitted he caused trouble during the lesson because he was not “in that maths zone”. He then added: “And by the time the exam comes you either do it, or shy away from it. You are the coward innit” *[02LE03]*. Other students avoided confrontation with their low sense of self-efficacy in maths or English by avoiding class altogether.

The interviews did not suggest that the students were disengaged with learning in general. This observation is particularly clear when contrasting student's perception of their vocational course to their maths or English course. The majority of students spoke about future career goals, met up with fellow classmates to work on assignments, completed work outside of the college, and shared news about grades or upcoming deadlines with their close relationships. They used words like “pride”, “enjoy”, or other positive words to describe their vocational courses; sentiments that were absent from their description of basic skills courses. Participants described the value of maths and English in general, slogan-like terms, such as “it just opens up many more doors” *[01LE02],* or “I know that there is lots of government standards, like, everyone needs English and maths*” [01LE01]* but only one student then proceeded to discuss the importance of the subject to her own future career and skillset.

Study supporters’ attitudes about maths and English were distinctly more positive than those of the students they were supporting. Five out of six interviewed study supporters felt a sense of responsibility to ensure the student would complete the course successfully. Louis, an 18-year old male, took his study supporter role seriously:

> “I try my best to make sure that he gets that push in the back and the support that he needs. [...] He is from Italy and his parents are from

his country, and he is only living with his nanny - aunty, and he doesn't really get much motivation and support from the family let's say. I try to make sure that he gets that." (Louis, study supporter, 03SS02)

Louis then compared his interaction with Colin (the recipient) before and during the intervention, and said he now has the habit to check in with Colin every time he receives a text message whereas before they did not speak as often. Only one study supporter, Bethany, who was nominated by her classmate in maths, expressed a relative lack of concern about her friend's progress in maths: "I wasn't too bothered. I was more focused on her getting her health and social done, and then the GCSEs". She added: "Sometimes she was like, "I can't be bothered with maths", and I would be like "yeah, I can't be bothered either" *[01SS01]*. This excerpt illustrates why classmates may not be suitable study supporters. Students were discouraged from nominating their direct classmates in the current trial, but in total 218 students nominated a classmate.[47] It could be argued that there is little chance they are able to provide the necessary encouragement if they feel dejected about the course themselves. On the other hand, students perceived several benefits to nominating a fellow classmate, including their insider perspective, day-to-day contact and a shared understanding of learning techniques. In fact, several students who had nominated an older person (i.e. parents, grandparents) felt unable to ask questions because they were used to different instruction methods, causing confusion instead of clarification: "if I go to them with a question, they'll help me, but I can tell that like they're struggling themselves with the question" *[04LE05]*.

The above quote illustrates a potential barrier to the effectiveness of the intervention. If students do not feel confident their nominated supporter can provide suitable assistance, they may be less likely to address the topic. The text messages were written to be accessible to people with limited background knowledge of maths and English. For example, jargon is avoided and where jargon cannot be avoided a URL is provided with further information about the topic. Additionally, the schedule of texts alternated between subject-specific texts (e.g. punctuation or how to use a calculator) and general motivational texts (e.g. ask what he/she is most proud of

[47] Approximately half of these nominated classmates also signed up to take part in the intervention themselves (N = 98); the threat of contamination is discussed in-depth in sections 5.3.5 and 5.6.3.

achieving this term) to ensure supporters with low confidence in maths and English would still feel able to fulfil the role of study supporter.

5.7.6.2 A potential mediator of the intervention: future learning goals

Although the majority of students had either an ambivalent or negative attitude towards their maths or English class, some recognised the value of the qualification for future study. Six students (40%) who planned to continue studying and some hoped to eventually make the transition to a university course, explicitly discussed achieving their GCSE qualification as a gateway to success. This subset of students seemed to engage more consistently with the intervention than those who did not explicitly link basic skills to their own opportunities in life. For example, students who recognised the importance of GCSEs for further study initiated conversations about the text messages, instead of waiting for the supporter to do so. Those who saw the course merely as a 'tick-box exercise' did not appear to speak to their study supporters frequently, and in some cases failed to read the texts they themselves received from Project Success.

This theme underlines the importance of sufficient baseline motivation to engage in learning and touches upon an important limitation of the intervention. In the first place, the most disengaged students may not sign up to take part, and if they do, they may disregard the text message or resist engaging with their study supporter meaningfully. Hana, a maths student, explained why he and his friends took no notice of the weekly texts:

> "If something like a text came by, they would read it. [But] if it was something like maths or English they wouldn't bother. If it was like a test they would bother, but a simple text or coursework being sent out for revision, nah." *[Hana, GCSE maths, assigned to supporter only group, 01LE03].*

5.7.6.3 Students' global perceptions of the intervention

In contrast to the previous intervention (where most interviewees were unsure whether their supporters received any text messages) all students interviewed as part of the follow-up study were aware that their nominated supporter was involved in the programme. Students' recall of the intervention was more detailed than the previous study described in Chapter

4, although this may be due to the interviews occurring earlier in the year. The immediate response elicited from students was a set of generalised positive comments, such as “I think the project is good and obviously it gives someone else to talk to and support on maths” *[04LE01]*. Additionally, students were better able to recall the details of the text messages and resulting conversations in the current trial. Three students did not engage in conversations about the text messages, but the overwhelming majority recalled precise details, as discussed in the sections below.

Only two students were overtly negative about their participation in the trial. The first student felt he was not in need of external encouragement. The second critical interviewee commented that “I tend not to message people from college and keep things separate to my life. [...] I’d rather they just give me the work instead of messaging her” *[01LE03]*. Ultimately, it transpired that this student had nominated his girlfriend at the time, whom he had broken up with since. Sociological research finds that adolescent romantic relationships are often more superficial than same-sex peer friendships and based on idealised expectations rather than real intimacy (Giordano, 2003). Since the sample population primarily consisted of 16 to 18- year-olds, close friendships may provide more consistent support in “a more settled and comfortable social arena” (Giordano, 2003, p. 270).

5.7.6.4 Bridging the gap when college support is lacking

There was a general consensus among students and supporters that the primary benefit of the intervention was to fill in the gaps when existing college information was lacking or insufficient. For some students it primarily appeared to be an issue of low engagement with the course materials, but others attended every class and still felt uncertain about their learning and progress. This is illustrated by these quotes from two student participants:

> “It’s just, "did you get the message?", "yeah I did", when we got messages about what we should be studying. And "we haven’t been studying that" and we just go on and look at what we should be studying. [...] Because where we haven’t been getting help at the moment at the college, we have just been taking that for ourselves through the messages in what we should be looking at.” *[Molly, GCSE English, allocated to supporter + student texts, 02LE02].*

> "If I've missed the English class and I get a text message, let's say about a mock the other day and I wasn't in for that. And I got a text message about that and I knew that's what we were doing in class. So I messaged my friends, "oh what did you do for that, was there any homework?" *[Alex, GCSE English, allocated to supporter + student texts, 03LE01].*

These excerpts illustrate that the intervention may fill the gaps where existing college communication is lacking. In total, 30% of the interviewed students felt that existing college support was insufficient. In these situations, the programme prompted self-study. Especially the text messages with links to revision websites were appreciated within this context.

Second, when students failed to attend lessons, the text messages informed them what topics they had missed. Interestingly, the use of texts as a tool for self-reflection occurred both when students received the texts directly, and when their supporter communicated the content. However, students who received the texts themselves more explicitly recognised their reliance on the programme to stay on track.

5.7.6.5 Finding the balance between privacy and support

The topic of privacy is an important consideration in the design of information interventions. The text messages did not include personalised information such as student grades or in-class behaviour, yet a small subset of students appeared reluctant to discuss their learning at all. Thomas, the father and study supporter of Ivan (also interviewed), illustrated this tension:

> "Whenever I get updates from you, that he's preparing for the next exam, [will] sometimes be the first time when I am hearing about them. So, even though I try to not to be very inquisitive, or you know, looking into his own private matters. I respect his privacy as much as possible. *[Thomas, father, nominated supporter, 03SS01].*

Ivan, his son, previously explained that "sometimes I want to keep it a secret from them" *[03LE02],* wanting to gain independence and not wanting to bother his parents whom he perceived to be busy with work and family life. A small subset of the interviewed students disengaged with the intervention because they felt it interfered with their sense of personal responsibility and privacy (N = 3). Two students whose supporters received weekly texts would have preferred to receive the texts themselves, and one only signed up for the lottery (a £25 voucher).

The majority of students did not share the abovementioned concerns about their potential loss of privacy. As a matter of fact, most appreciated the proactive communication by their nominated study supporters. Students recounted specific situations where their study supporter contacted them to discuss the text message. The ensuing conversation was typically short but focused, and perhaps most importantly, positive:

> "Then Lauren will call me and we will just like talk about it. Because yeah, just reflecting on it. It's like when you talk about something to someone else, they can then like lengthen it out a little bit" *[Isabelle, assigned to supporter + student arm].*

The positive wording of the text messages appears to be key: students used the words "confidence", "feeling positive", "reassurance" and "capable" to describe the personal change they experienced as a result of the text-messaging programme.

5.7.6.6 The importance of making the right study supporter choice

The most critical element of the intervention remains study supporter choice. In the previous study, Chapter 4, student experiences could be neatly categorised into two groups (see Section 4.8.7.2, p. 129). Overall, those who nominated individuals out of convenience did not experience strong benefits of the intervention, whereas those who nominated an individual they had an existing strong relationship strongly identified with the intervention. In this study, an identical pattern emerges. Table 5.20 displays engagement with the intervention by strength of relationship with supporter.

Table 5.20: Categories of closeness

	Weak relationship	*Strong relationship*
Positive change in perceived closeness with study supporter	Not recorded.	Last year, she wasn't close like this. But this year she is very close to me because of these text messages, she is sharing everything, she is telling me what is going on, what is done, what she didn't do and why she needs help, and then I give her support (Latha, supporter, 04SS02).
No change in perceived closeness	She tells me about [the texts] but I don't pay attention. So every time she might message me or something, [...] I tend not to speak to her. I'll read it and not look at it. Or she'll come up to me and tell me she's got a message and I'll just say okay, cool. (Hana, student, 01LE03)	We're still the same, like we've always been really close, and got on and spoke all the time, so it's just the same, except when we are talking about English sometimes as well (Molly, student, 02LE02).

None of the interviewees reported that they communicated with the person they had nominated less frequently as a result of the intervention. No students could be identified who initially nominated someone they self-reported they were not close to, but subsequently became much closer to the individual. This suggests that when the existing relationship is weak, the intervention is not capable of forging stronger relationships. When the existing relationship is strong, both students and supporters recounted relatively frequent and targeted conversations about the messages and their learning. For some, this resulted in greater perceived closeness, and others did not perceive a change in the nature of their relationship:

1. Greater closeness: A small subset of students (N = 3) felt the more frequent sharing of information resulted in a stronger bond. In each of these cases, the student was previously reluctant to discuss maths and English outside of the classroom. The texts facilitated the seeking and providing of support about these specific topics; support in other domains was already strong.

2. Neutral: The intervention did not affect the strength of the relationship (N = 6). These students typically already received targeted maths or English support from the nominated individual before the trial. The intervention simplified communication but did not change the nature of the conversations or the support provided.

5.7.6.7 What makes a good study supporter?

All interviewees were asked "what makes a good study supporter". The most common response centred around the emotional availability of the study supporter. Students used the words "someone who will not slack off", "someone who cares", "someone you can rely on", or "someone you can trust" to verbalise why they chose their study supporter. Especially important, in their view, is whether this person is emotionally available and who will "stick around". Implicit in these responses is the need for continuity and comfort. Interviewees recognised that people they had long-term relationships with would likely be more reliable study supporters. These responses also show students internalised the sign-up guidance well: they were advised to nominate an individual they trusted and felt comfortable talking to.

The maturity of potential study supporters played an important role in the decision process as well. Classmates were generally deemed unsuitable as they would "just laugh and play around" *[04LE03]* or "take it in a joke way" *[03SS02]*. In a similar vein, choosing someone who sees the value of maths and English, who "isn't going to put a negative effect on it" *[04LE01]* was listed by students as an important decision criterion.

Finally, most interviewees spoke to their supporters every day or every other day. Although students were wary of constantly bringing up the topic of maths or English: "keep on talking about the same thing again and again and I would be getting bored with it" *[04LE02]*. Yet, none of the interviewees complained of too frequent texts or conversations. Most interviewees felt the frequency of text messages was just right, or felt they would benefit from messages twice a week. Future studies could systematically vary the frequency of the communications to test optimal levels of communication.

5.7.6.8 *Lessons learnt*

The qualitative research illuminated three ways in which the implementation of the study would be improved. All three issues would not have become apparent without the qualitative research. They are easily resolved by small tweaks to the existing procedures, as described below.

Although only one out of fifteen interviewed students experienced a breakdown in relationship with their nominated supporter, future pragmatic iterations of the intervention could facilitate changing supporters throughout the year.

The content of the text messages did not always match the college curriculum. The texts were written at the start of the academic year, using the scheme of work (SOW) as a guideline. College leads would communicate changes to the schedule so that the texts would remain relevant and timely, but this appears to have slipped in several cases. When asked if the texts were relevant to their courses, four students recounted feeling confused because the texts did not correspond to what was being taught in class. In all cases, this mismatch was due to the teacher being behind on schedule. Future iterations of similar information interventions should consider ways to make the updating of information frictionless. A clear feedback process is essential. If tutors were to receive a weekly prompt to check the text message and make small amends or defer it until a future date, the texts continue to be relevant throughout the academic year. In this study, tutors signed off the full schedule of 35 texts at the start of the academic year. Tutors were then prompted to provide additional information at the start of the spring semester, but more frequent check-ins are likely beneficial to the relevance of the text messages.

Finally, student accounts of their interactions with supporters made clear that the degree to which participants engage with the treatment is not well-captured by SMS delivery statistics. The three disengaged interviewees received the texts but proceeded to ignore them. Similarly, both students and supporters admitted that they sometimes forgot to discuss the text, citing clashing schedules as the main culprit. Future studies could incorporate mid-year and post-intervention surveys to assess responsiveness (i.e. the degree to which participants engage with the intervention) and reach (i.e. the rate and score of participation; Humphrey et al., 2016). A compound measure of delivery statistics and self-reported engagement with the text messages will likely constitute a more accurate account of compliance for the CACE analysis.

5.8 Discussion

The chapter set out to present and discuss the findings of a second field experiment testing the potential of a supportive text messaging intervention. It has made several contributions to the thesis: (1) it has provided robust empirical evidence to answer the research questions; (2) it raises further questions about the intervention mechanisms; and (3) it offers new insights on the manner in which students seek, perceive, and receive social support from people in their social networks. The discussion below summarises these contributions, and then highlights the remaining lines of enquiry that frame the next and final empirical chapter.

5.8.1 A closer look at the research questions

In response to research question 1, which asks if a supportive communication intervention can boost attendance and attainment at further education colleges, this chapter provides mixed evidence of the intervention's effectiveness. None of the treatment variations in which either students, supporters, or both received weekly messages, successfully improved overall attendance rates. Overall, average attendance levels were relatively high in this study at 72%, in comparison to 58% in Chapter 4. The multiply imputed results are similar to the main results, which suggests that missing outcome data do not drive the results.

The results are more promising for attainment than attendance, although not all variations on the treatment resulted in improvement. The trial demonstrates that the best performing treatment, the 'supporter + student' group, boosted attainment by 6 to 9 % points, but the treatment effect was only significant at the 10% level (simple model, Table 5.11, p. 176). The estimated effects of this treatment are similar in size to those reported in Chapter 4 and translate to 18% of the control group standard deviation, or Hedges g = 0.17. The failure to reach conventional levels of statistical significance (p = 0.05) may be primarily a power issue. The Study Supporter trial in Chapter 4 was a two-arm RCT in which over 750 students were assigned to each arm. This study, on the other hand, has four trial arms and 250 students assigned to each arm.

Furthermore, direct texts to students were found to be more effective (4 - 8 % points improvement in achievement rates depending on model specification) than texts sent to study supporters only (-1 to +3 % points). This finding leads us to research question 2, which asks whether communication with a designated study supporter is as effective or more effective than direct college-student communication. The treatment effect is greater for students assigned to both receiving the texts themselves and having a study supporter who received texts. However, the difference in treatment estimates between the 'student only' and 'student + supporter texts' is not statistically significant, thus research question 2 is not convincingly answered.

A surprising finding emerged from this study, as the original configuration of the intervention (see Chapter 4 for the intervention design, 4.3.6, p.96) did not replicate successfully. The version of the intervention that resulted in statistically significant increases in attendance and attainment in the previous trial, the 'study supporter only' texts, did not prove effective in the current trial. In Chapter 4, the treatment resulted in a 3.1% point improvement in attendance and a 6% point improvement in achievement. In the present chapter, the treatment coefficients are close to zero.

It is unclear why these findings diverge. The interviewees assigned to the 'supporter only' group were as positive about the programme as those in the 'supporter + student' group. Similarly, the difference cannot be explained by higher opt-out rates from texts or a lower text delivery rate for the supporter only group. The analysis finds effects of the 'student only' and 'supporter + student' treatments, but not of the 'supporter only' treatment. A replication with a larger sample size may be required to provide more robust evidence on the disparities in effectiveness between the configurations of the supportive communication intervention. I am currently running a follow-up study with 3800 students randomly assigned to one of the four treatment arms, funded by the Education Endowment Foundation and independently evaluated by NatCen Social Research. This trial will hopefully be able to provide a more conclusive picture of the relative effectiveness of direct college-student communication versus involving a close relationship in the conversation. The scale-up is further discussed in Chapter 7.

5.8.2 Evidence of heterogeneous treatment effects

The analysis provides evidence of heterogeneous treatment effects. In this study, male learners benefited more from the interventions (a 13.8 % points increase in attainment in 'student only' and 13.7 % points for 'supporter + learner' groups) than females. Female learners achieved better grades on average: 31.9% achieved a good pass in the control group, whereas only 12.8% of male learners achieved a passing grade in the absence of the intervention. This data is suggestive that the supportive communication intervention benefits lower-attaining students most. Since the subgroup analysis was exploratory in nature, these patterns are subject to replication in future studies. This finding also contradicts that of Chapter 4, where treatment effects were found for male students' attendance rates, and female students' attainment. It should be noted that the smaller subgroup samples used for the heterogeneity analysis limit the conclusions that can be drawn from these findings.

Exploratory analyses on the effectiveness of the intervention in maths versus English courses raise the prospect that some subjects are more conducive to a social support intervention than others. The treatment effect is driven by improvements in English attainment rates, rather than maths. It would be interesting to further explore whether students and supporters find it easier to weave conversations about English into their daily interactions, and struggle more with the maths content. It is also possible that maths anxiety is more common than anxiety about English, leading students and their social networks to avoid situations where they have to use mathematics or discuss related topics. The qualitative component provided limited evidence of students' and supporters reluctance to discuss maths-related texts, however. The next section summarises findings from the in-depth interviews with students and study supporters.

5.8.3 Qualitative follow-up

The qualitative data show that implementation matters and that it is essential that students nominate trustworthy and caring study supporters. The intervention is undermined when students are ill-informed about its purpose and choose someone they do not feel comfortable speaking about their learning with. On the other hand, the potential of the supportive texts is evident when supporters take the encouragement to heart and engage in a positive dialogue with the student.

The recruitment procedure and materials were modified based on qualitative insights gleaned from the Study Supporter intervention, which had a positive effect on the acceptability of the intervention. Additionally, good adherence to the intervention was documented in the current study. Both students and supporters recounted specific conversations about the text messages, and all interviewees would recommend the programme to future students. Finally, the intervention did not generate controversy and both students and supporters reported that they had found the experience rewarding. Collectively, the qualitative findings suggest that this study, Project College Success, leverages existing relationships to encourage students to stay engaged with their learning, and that students generally appreciate such proactive involvement from their social networks.

5.9 Conclusions

This chapter set out to present the findings of a second and more sophisticated field experiment on the potential effectiveness of a supportive communication intervention. This experiment estimated the effects of the following three treatments on student academic outcomes: (1) texts sent to nominated supporters; (2) direct texts to students; and (3) a combination of the two. A fourth, smaller, arm added a curiosity-inducing element to the 'supporter + student' texts, by texting a question to the student and simultaneously sharing the clue with the supporter. The thirty-five weekly texts touched on a variety of subject-specific topics and were always positively worded.

The chapter presents robust empirical evidence. The experiment, which built on the Study Supporter intervention discussed in the previous chapter, found no effect of the 'student only' and 'supporter only' text messages across the general student population. The third and most intensive treatment, 'supporter + student texts', was most successful at improving achievement rates ($p < 0.10$). Considering that the text messages only cost £3.50 per student in the single-recipient arms and £7.00 per student in the double-recipient arm this approach warrants further exploration and replication with larger sample sizes.

The study addressed both the primary and secondary research questions posed in this thesis. Research question 1 asks whether the supportive communication intervention can improve attendance and attainment in maths and English courses at further education institutions. The overall effectiveness of the supportive communication intervention received mixed support in this chapter, as the best performing treatment fell short of conventional statistical significance levels. Nevertheless, the size of the effect (Hedges' g = 0.17) is promising.

Further, the research design of the study allowed exploration of the secondary research question, which asks if direct communication with students is more or less effective than communication via their nominated study supporters. Communication with students and their supporters simultaneously resulted in greater treatment effects than communication with students or supporters only. The treatment estimates of these two variations are not sufficiently different to conclude that 'student + supporter' text messages are superior. Yet, if practitioners were interested in implementing this approach, the combined evidence of Chapter 4 and 5 suggests that involving study supporters can have a significant impact on student success.

6 SEEKING SUPPORT VERSUS SEEKING COMMITMENT

6.1 Introduction

Qualitative methods, embedded within two randomised controlled trials, were used to support three primary aims. The first was to develop effective recruiting, consent, and implementation practices. The in-depth interviews with tutors and student during the Study Supporter trial (Chapter 4) focused on barriers and facilitators of successful implementation and interpretations of study experiences. The qualitative data helped gain insight into the processes of supportive communication between the college, study supporter, and student.

The second trial, Project Success (Chapter 5), was further developed using the insights gleaned from the process evaluation of the Study Supporter trial. Its qualitative component focused on identifying potential facilitators and barriers of intervention effectiveness. For example, the strength of the existing relationship between student and their nominated supporter appeared critical to engagement with the intervention, as well as students' internalisation of learning goals. As such, the qualitative data provided a more nuanced understanding of the social support processes set in motion by the intervention.

The final aim of the qualitative approach used in this thesis was to gain a deeper understanding of the potential mechanisms of the two interventions set out in this thesis. The literature review identified two broad potential mechanisms of the supportive communication intervention. The intervention could facilitate social support by breaking down boundaries to seeking and providing support, which in turn improves student engagement and academic outcomes. Alternatively, the intervention could help improve academic outcomes if students use the knowledge that someone can monitor their learning as a social commitment device. Exploring students' motives for participation in the intervention and their subsequent experience of support versus monitoring is the focus of this chapter. The qualitative data collected in both phases is combined to provide a full account of student and study supporter experiences.

6.2 Theoretical mechanisms of change

Before setting out the qualitative approach to the final empirical question, I briefly review two potential mechanisms of the intervention. It is unclear why communication with study supporters works: is it mainly because interaction with the supporter leads to greater self-esteem and resilience, or is it mainly because the study supporter holds the student accountable? The first pathway fits neatly within the literature on social support and conceptualises the study supporter as a mentor. Social support interventions typically activate individuals' close relationships to become more involved with and better informed of the needs of the recipient (Taylor, 2011). The second potential pathway can be placed within the literature on commitment devices and commitment contracts, on the assumption that students opt to take part because they are motivated to achieve their coursework but know that they might struggle to follow through on their good intentions.

Theoretical models of social support posit that support from close relationships enables individuals to cope more effectively with stressors (Cohen & Wills, 1985; Thoits, 1986, 2011). Students who report having access to adequate social support are at lower risk of drop-out, spend more time studying, pay better attention in class and achieve better grades than students who perceive lower availability of support (Rosenfeld, Richman, & Bowen, 1998; Rosenfeld et al., 2000). The proposed mechanisms through which social support improves educational outcomes include uncertainty reduction, improved self-esteem, cognitive reappraisals and positive affect (Bodie & Burleson, 2008; Burleson, 2009; Burleson & Goldsmith, 1996). The first purpose of this chapter is to assess what type of (supportive) interaction students expected from their nominated study supporters at sign-up, and to what degree supportive communication helped buffer students from stressful life events and daily stressors.

On the other hand, the intervention may have appealed to students because it presents a helpful tool to ensure they keep themselves in check. The students in our sample may have the sophistication to know that the intervention is beneficial to them, even if a resulting increase in monitoring is not particularly enjoyable. Sophisticated individuals recognise their dynamically inconsistent preferences and choose to constrain their future choices (Laibson, 1997) in this case by electing to be monitored by a third party. In other words, students might perceive the interaction with their study supporter as a social commitment device[48] to help overcome internal conflict between "should" and "want" urges (Milkman, Rogers, & Bazerman, 2008). A commitment device is defined as an arrangement which limits the individual's future choices by lifting self-imposed restrictions only when the goal is accomplished or imposing penalties for failing to accomplish said goal (Rogers, Milkman & Volpp, 2014). The cost of failure could be conceptualised as a psychological cost in the form of reputational damage, shame, or loss of self-esteem (Bryan, Karlan, & Nelson, 2010).

It should be noted that this intervention may be at the boundary of what can be categorised as a commitment device. It relies on a third party, the study supporter, to engage with the intervention consistently. The study supporter could fail to engage with the text message in the way initially intended by the student, which arguably reduces the cost of failure. After all, a disengaged study supporter probably would not be disappointed if the student fails her exam. The student cannot know the extent of the study supporter's engagement at sign-up, so for the purpose of this chapter I assume that students incur a psychological cost if they fail their course.

Participation in the two field experiments would not be a commitment device if sign-up primarily stems from the motivation to influence the actions of others. It should arise from the motivation to influence the actions of one's future 'self' (Bryan et al., 2010). This defining characteristic of commitment devices clarifies how student motives for participation can be distinguished; students may either wish to seek greater involvement and encouragement from a close relationship (i.e. motivation to influence actions of study supporter) or use the knowledge that this person receives frequent communications to commit to study effort (i.e. motivation to influence future self). This chapter explores if wanting to constrain one's future behaviour

[48] Although the commitment device relies on interaction between two individuals, the student and the supporter, I do not conceptualise the intervention as a commitment contract. Commitment contracts are usually conceived as formal contracts between two parties (Bryan et al., 2010). Study supporters receive the weekly communications without having consented into the study. Further, they do not confirm that they accept the role of supporter.

was a primary reason for signing up or whether students primarily hoped to activate greater support from their social networks.

6.3 How do participants conceptualise participation?

6.3.1 A qualitative exploration

In order to synthesise the qualitative data collected during the two field experiments all transcripts are combined. The combined qualitative dataset is comprised of 21 in-depth student interviews conducted across 5 further education colleges, and 5 phone interviews with study supporters. The majority of students were assigned to the treatment group (90%) since the interviews focused on students' experiences of interactions with nominated study supporters. The data source of the supporting quotes are indicated as experiment 1 (Chapter 4; Study Supporter) and experiment 2 (Chapter 5; Project College Success).

The transcripts were previously coded in line with the qualitative research objectives in Chapter 4 and 5, respectively. For the purpose of this chapter, the data was re-coded using a simple coding scheme to identify the different ways in which students expected and experienced interactions with their nominated study supporters. The coding scheme focuses specifically on student motives for signing up and classifies students into two categories: (1) signing up in the expectation of social support, versus (2) the expectation of monitoring.

However, examination of the transcripts revealed that a number of students' expectations of the programme did not neatly fall within these two categories. A number of students indicated that they expected direct reminders and did not anticipate involvement of a study supporter. All tutors were asked to frame the interventions as the opportunity for greater involvement of key relationships in students' learning. It appears that some tutors explicitly introduced the project as a 'reminder service' instead, deviating from the instructions. Table 6.1 displays the resulting coding framework.

Table 6.1: Coding framework

Category	Coding criteria	Supporting quotes
Social support	Expected: • supportive interaction about their learning • sustained involvement of nominated supporter	*What convinced me was the idea that someone would help me to sort of plan my – I don't know how to explain it fully, but someone would guide me along*. - (03LE02, experiment 2)
Commitment device	Expected: • needing someone to help overcome self-control problems • psychological costs of failure, such as shame or embarrassment.	*I [nominated] my boyfriend's mum because I know how strict she can be at times, so that puts me on track. So at any point I go off track then she puts me on track so it helps along*. - (04LE03, experiment 2)
Factual support: reminders	Expected: • factual information • direct college-student communication Did *not* expect third party to help motivate them.	*I thought [the text messages] would be a good way to keep up with my learning to actually pass my exam*. - (03LE01, experiment 2)

Seven students (35%) were identified as expecting support from a close relationship, two (10%) expected texts as reminders, and only one student (5%) could be categorised as signing up as a commitment strategy. Three interviewees (15%) did not remember their motivation for signing up and two students (10%) expected both social support and reminders. Surprisingly, five interviewees (25%) were not aware that someone might receive weekly text messages if they signed up.[49] This simple coding exercise resulted in limited evidence in support of the commitment device hypothesis, and some support for the hypothesis that students hoped to activate social support.

This qualitative dataset is limited in a number of ways. First, the small sample size (N = 20) and purposeful nature of the qualitative sample

[49] These individuals thought they were simply completing a questionnaire. Since all students completed the sign-up procedure in class, it can only be assumed that tutors did not dedicate enough time to explaining the project in sufficient detail. Future intervention studies should pre-test the sign-up materials to ensure they are engaging and informative.

selection precludes generalisation to the larger student population. Second, students were asked to reflect on their expectations at a later point in time, a delay of 2 months for the Project Success interviews (Chapter 5), and of 7 months in case of the Study Supporter interviews (Chapter 4). Students' expectations may have been coloured by subsequent experiences of the intervention. The next section therefore explores student motives for participation in the moment of signing up.

6.3.2 Exploring motives for participation at sign-up

To explore motives for participation in a larger sample, students signing up to the follow-up trial funded by the Education Endowment Foundation, which results are not included in this thesis,[50] were asked to complete a short questionnaire. The scale-up of Project Success was implemented across 31 further education colleges. In total 4,226 students nominated a study supporter and completed the survey.

First, students were asked to indicate which statement corresponded best to their primary motive for participation, from the following three statements:

I am concerned that I will...

1. Lose motivation and focus, so I want my study supporter to hold me accountable to stick with [subject];
2. Lose my schedule and forget when things are due, so I want my study supporter to remind me;
3. Need specific content-related help, so I want my study supporter as my class-content coach.

These three options were based on the most common responses in our qualitative data. Half of the students indicated they wanted someone to hold them accountable (N = 2,057, 49%). Fewer hoped for content support (N = 1,238, 29%) or reminders (N = 936, 22%). It appears that a sizeable portion of students recognise they may need someone to remind them of their long-term goals in the face of distractions and short-term desires.

Further, students rated whether they were motivated to avoid embarrassment and disappointment if they were to fail the course (binary; yes/no). These responses allow for a proxy measure of the psychological cost

[50] The trial ran from September 2017 until June 2018. Outcome data collection commences in October 2018 and results will be released in Spring 2019.

of failure. The vast majority of students responded that they were highly motivated to avoid embarrassing (82.5%) or disappointing (88.7%) their nominated study supporter. Students also indicated how embarrassed and disappointed their nominated supporter would be if they failed their course. This item was included in order to assess whether students had nominated individuals who would take their potential failure to achieve the course seriously. The results displayed in Table 6.2 suggest that students did indeed nominate people who would care about their educational success.

Table 6.2: Expectations of supporter response to course failure at sign-up

Rating	*How [...] would [study supporter] be if you failed your course?*	
	Embarrassed N (%)	Disappointed N (%)
Not at all	907 (21.4)	356 (8.4)
A little	1,240 (29.3)	1,109 (26.2)
Quite	1,089 (25.7)	1,262 (29.8)
Very	995 (23.5)	1,504 (25.6)

These findings support the hypothesis that students incur a psychological cost in the event they fail the course. It is important to note that these questions are hypothetical in nature and invite respondents' predictions about their emotional reactions to future events.

Recent experimental studies find that people often exaggerate the emotional impact of future life events, such as the breakup of a romantic relationship or not getting that dream job (Eastwick, Finkel, Krishnamurti, & Loewenstein, 2008; Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998). Failing to achieve one's GCSE in maths or English is likely an emotional event, since it results in being required to retake the subject again the following year and being unable to qualify for higher-level study. [51] The survey responders may have overestimated how much they will think about disappointing their study supporter in the event of a poor GCSE grade, neglecting the many other topics that will be competing for their attention (Wilson & Gilbert, 2005). Yet, the perceived emotional impact at sign-up is

[51] If the student is aged 16 – 18. Resits are no longer required for continued funding once the student turns 18 (before the start of the academic year).

of interest because it is the perceived cost of failure that makes a soft commitment device a powerful motivator.

6.4 What types of supportive interactions take place?

Having assessed students' expectations of the programme, there is some indication that students hoped to activate more support from their immediate social networks. The survey data shows that a study supporter could be a useful commitment device as students indicate that the psychological cost of failure would motivate them to stay on track. This section explores student's actual interactions with their study supporters.

Transcripts of both qualitative phases are combined (N = 26) and analysed for evidence of supporting versus monitoring interactions. Interviewed students and supporters described their experiences of interaction or communication with their supporter, and vice versa. As such, the interviews focused on enacted support rather than hypothetical support (i.e. support students were initially expecting to receive). The meaning that the supportive behaviours had to recipients was often implicit in these accounts. In line with the most common typologies of social support (e.g. Gottlieb & Bergen, 2010; Thoits, 2011) interview excerpts are coded as being supportive in an informational, instrumental, or emotional way. The coding framework was informed by the types of social support set out by Thoits (2011, p. 146). The final criterion examines if students perceive study supporters as a commitment device.

Table 6.3: Coding framework, types of social support

Category	Coding criteria	Supporting quotes
Informational support	Provision of: • Facts or advice • Feedback/guidance • Help (solving a problem, accomplishing a task)	"*Whenever I thought I was having a hard time I went straight to my Nan, and told her everything that was going on and she gave me the best advice on what I could do*" - (02LE02, experiment 2)
Instrumental (tangible) support	Offering of: • Assistance with practical tasks or problems • Material resources	*"Paying tuition fees for him every week for English and maths."* - (03SS02, experiment 2)
Emotional support	Communication of: • Caring, love • Encouragement • Esteem and value	*"They encourage me, they always speak to me, if you do read a lot or concentrate in your work, you can learn and you can do better."* - (04LE04, experiment 1)
Commitment device/ monitoring	Awareness of: • Inconsistent preferences • Strategic choice to nominate supporter • Supporter engages in monitoring behaviour	"*She'll be saying, "have you got any work to do" and I'll say "yeah", or she'll say to me, "get up and do your work". She does this thing where she'll constantly pull me."* - (03LE01, experiment 2)

The coding exercise discovered that interactions between students and their study supporters fulfil several functions at once. Table 6.4 displays the frequency with which students and study supporters recount specific types of supportive interaction. Seventy per cent of the students recounted specific episodes of receiving support from their nominated study supporter. Students refer to informational (40%) and emotional support (40%) most often, and rarely mention receiving instrumental support (10%).

Table 6.4: Frequency of types of support episodes

Type of interaction	*Source*	
	Student N (%)	Study supporter N (%)
Informational support	8 (40%)	5 (83%)
Instrumental support	2 (10%)	2 (33%)
Emotional support	8 (40%)	4 (60%)
Constrain impulses by nominating supporter	0 (0%)	N/A
Monitoring	5 (25%)	3 (50%)
Total unique transcripts	*14 (70%)*	*6 (100%)*

Examples of informational support included content-related help (e.g. marking practice exam questions), help navigating broader college issues (e.g. calling the tutor to request extension of a deadline), and general guidance (e.g. discussing study plan, reminding about upcoming exam). Instrumental support was rare, and exclusively consisted of examples where the study supporter organised transport to college. Examples of emotional support were more general in nature, and include general encouragement, receiving praise, and feeling cared for.

None of the interviewed students displayed evidence of an inner struggle between competing goals or desires, although this is likely a reflection of the interview setting and focus of the interview questions rather than proof of the absence of internal dialogue of time-inconsistent preferences. Students were asked to describe their study behaviours and long-term goals, but not in sufficient detail to tease out internal tussles between their 'planner-self' and 'action-self'. Several instances of monitoring behaviour were recounted by interviewed students (25%), such as handing over the mobile phone until homework is completed. The six students who did not recount any specific supportive interactions with their study supporters (a) forgot they had signed up, or (b) resisted communication, as discussed in more detail in the previous chapters.

These descriptive counts of type of supportive behaviours initially mask an interesting observation. Eight out of the ten students who described their experiences of informational and instrumental support also implicitly attached emotional meaning to them. For example, students recounted experiences in which their supporter provided advice or drove them to an open day at college and assigned emotional rather than transactional meaning to them. Implicitly, most descriptive passages conveyed that the interaction with their study supporter helped students cope with their emotions. One student recounted the numerous instances in which her study supporter helped with her assignments:

> "She is like my teacher, so the same way you would behave around your teacher and expect your teacher to help you. She does it without any like reliance, so she will just come and take it upon herself to help me and stuff like that." *[02LE02, female, age 16, experiment 2, nominated her grandmother].*

Students' affective evaluation was also elicited during the interview. This student was then asked what made her grandmother's behaviour helpful. She responded:

> "I do love it, because no matter what she will support me, and she helps me like her best, so it takes a lot of weight off me. Knowing that I have somebody there to support me." *[02LE02, female, age 16, experiment 2, nominated her grandmother].*

The interactions with study supporters comforted students, helped them to see things in a more positive light, and assisted students in coping more effectively with challenges. Even monitoring behaviours, although not always experienced as particularly enjoyable, were interpreted in a positive light:

> "She keeps mentioning it, she keeps nagging, "Have you got any homework, have you got any coursework?", "Have you got this, have you got that?". I'm like "Yeah, okay I'll get on with it". But it's good in a sense because it supports you and reminds you." *[05LE03, female, age 17, experiment 1, nominated her mother].*

Although social desirability bias may have affected interview responses, only one student recounted interactions with his study supporter which he found unhelpful.

The intervention explicitly encourages study supporters to provide informational support (e.g. “Ask [student] if they have already bought a calculator”, “Ask [him/her] to tell you the difference between a radius and diameter”) but the qualitative data shows that students often experience these interactions as providing emotional support. These findings are in line with the hypothesis that social support primarily serves to satisfy a need to belong, a need to feel relatedness to close others (Baumeister & Leary, 1995; Furrer & Skinner, 2003; Ryan & Deci, 2000). Additionally, emotional support is positively associated with successful coping with stressful events and emotional well-being (Taylor, 2011; Thoits, 2011) and academic outcomes (Burleson & MacGeorge, 2002). The qualitative data suggests that one of the pathways through which our supportive communication intervention improves educational achievement is that of greater perceived emotional support. I return to this idea in the discussion chapter.

Finally, this thesis takes a brief look at the types of support described by study supporters to assess how they interpreted the text message prompts. Study supporters use similar language to describe supportive interactions with students. They primarily provided informational (83%) and emotional (60%) support, each interviewee recounting multiple occasions of each. The data also shows that supporters did not stick to providing just one type of support. Supporters appeared to be responsive to situation-specific needs; they helped prepare for exams when a deadline approached and consoled students when they needed emotional support. Perhaps the most interesting theme that emerged from the data was the close emotional connection study supporters felt with the students. One study supporter, who was nominated by her son’s girlfriend (also interviewed; 04LE03), tells:

> “She is missing her parents most of the time… that’s why sometimes she doesn’t want to go out, she doesn’t want to study more. […] I explain her “listen, you are staying here sometime, you come to eat with me, I am your mum, don’t think like your parents are not here.” I just give her love like she is my daughter. Then she forgets everything.” *[04SS02, female, age 39, experiment 2, supporter of son’s girlfriend].*

The qualitative findings that emerged from the thematic analyses performed in Chapters 4 and 5 underlined that the existing strength of the supporter-student relationship is an important facilitator of the effectiveness of the intervention. Students who nominated casual acquaintances typically received limited supportive communication. Those who nominated someone they trusted, felt close to, and already communicated with regularly showed considerably more evidence of benefiting from the intervention, at least psychologically.[52] Future iterations of the intervention should take into account that the choice of study supporter appears essential to the effectiveness of the intervention, and guide students towards a thoughtful choice.

6.5 Discussion

This chapter set out to explore descriptive evidence of two potential pathways through which the intervention helps improve academic outcomes. It should be noted that especially the sample of study supporters was small (N = 6) with likely selection bias due to convenience sampling. All interviewed study supporters were positive about the programme and took their role as study supporter seriously, which is unlikely to be representative of the wider sample. The sample of interviewed students (N = 20) was purposively sampled and more diverse. Future studies would benefit from drawing a random sample from the sample of trial participants for the embedded qualitative component. Acknowledging these limitations, the chapter provides a rich insight into students' and supporters' experiences of supportive communication.

The first section considered students' expectations of their participation in the programme to assess whether they hoped to use their study supporter as a coping resource or whether they nominated someone as a commitment strategy. Two data sources were used to answer this question: interviews conducted during and after the trials, and questionnaires collected during the sign-up procedure (before trial launch). The interview data suggested that students primarily expected receiving social support, but their reflections may have been coloured by subsequent experiences of the intervention itself. The survey data, on the other hand, provided evidence that students were implicitly aware of their self-control problems and that the psychological tax

[52] Due to the small sample size of interviewed students (N = 20) spread over two field experiments it was not deemed feasible to assess whether greater expressed closeness resulted in better attendance and achievement outcomes. Future studies could use random sampling to construct a qualitative sample to more robustly explore associations between perceived closeness and effectiveness of the intervention.

of disappointing or embarrassing their study supporter motivated them to do well in the course. This data was limited by the hypothetical nature of the questions. People are prone to overestimation of future emotions (Wilson & Gilbert, 2005), so what might feel motivating at the start of the college year may fade into the background once many other events compete for their attention. Yet, this data provides the first glimpse into participants' motives for signing up to a communication intervention. The analysis found evidence of both potential pathways. Seeking social support from close relationships is by definition inter-personal and seeking a commitment device to lock-in future study effort is mostly intra-personal. Future RCTs could elicit students' motivations for participation at sign-up to assess whether the intervention benefits particularly those who perceive it as a strategy to keep themselves on track or those who hope to receive more support.

The interview data also provides fresh insight into students' and study supporters' supportive communication with one another. This data is the first of its kind within the literature on communication interventions delivered via text message. Although some studies charted text recipients' behavioural responses to the treatment (see for example Bergman & Chan, 2017; Chande, 2016), these were limited to multiple-choice survey reports. This chapter presents rich qualitative data not only on the type of behaviours study supporters engaged in, but also what meaning students attributed to these interactions.

The academic literature has long distinguished between different types of support (e.g. informational, instrumental and emotional) but has to date failed to elucidate their different meanings. Most studies on social support and education simply describe how students perceive different types of support from different sources (i.e. parents, peers, teachers; Malecki & Demaray, 2003). This analysis has shown that the type of support received may not be the essential ingredient, but the meaning it carries. Study supporters engaged in various types of supportive and monitoring behaviours, but the common denominator was students' interpretation of these interactions: caring, understanding and comfort. The analysis contributes to the scholarly debate on communication interventions by offering a new perspective: subject-specific support such as help with homework or reminding the student of an upcoming exam may be subordinate to the implicit communication of social belonging and esteem.

This chapter also provides novel qualitative evidence that support from parents and friends helps students overcome challenges through appraisal processes. Students often mentioned feeling more positive and hopeful through interactions with their study supporters. This data therefore offers new support for the popular hypothesis in the published literature (e.g. (Feeney & Collins, 2015) that improved coping with emotions is a primary mechanism of the beneficial effects of social support on well-being and academic outcomes.

7 SUMMARY AND CONCLUSIONS

This thesis has contributed both to the scholarly debate on social support interventions and to the growing body of work showing that inexpensive information interventions can help improve student outcomes. This chapter summarises the empirical evidence gathered in this thesis and then addresses limitations of the research design, directions for further study, and offers concluding remarks.

7.1 Research findings

The two research questions read as follows:

RQ1: Can supportive text messages improve students' attendance and attainment?

RQ2: Are the effects greater if students also receive the text messages?

The first research question is addressed in both Chapters 4 and 5, and the second question is addressed in Chapter 5. I now turn to a brief summary the findings and support for each research question in turn.

7.1.1 RQ 1: Can supportive texts improve student success?

The first experiment, the Study Supporter trial, demonstrated that weekly text messages to study supporters leads to an 3.1 % point (or 7 percent, Hedges g = 0.11) improvement in class attendance and 6 % points (27 percent, Hedges g = 0.16) improvement in students' probability of passing the qualification.

The second experiment, Project Success, provided a weak replication of the effectiveness of the supportive communication intervention. The intervention did not result in better class attendance (*n.s.*). I find promising effects of the intervention on achievement, but only for the most intensive treatment arm where both parties received weekly communications ('supporter + student' arm, $p < 0.10$, Hedges g = 0.17).

The considerably smaller sample size per treatment arm in the second experiment may, at least partly, explain the diverging results. The Study Supporter trial consists of only two treatment arms and involved over 1600 students. In contrast, Project Success is a four-arm trial and managed to recruit just short of 1000 students.

In sum, this thesis demonstrates that potentially cost-effective way to improve post-16 success is to inform students' family, friends and wider social network about their learning. The simple, positive and actionable weekly text messages sparked conversations, reminded students of upcoming exams, and ultimately led to better class attendance and higher pass rates. The second intervention shows effect sizes that are similar to the first (Hedges' g = 0.16 in Chapter 4 and Hedges' g = 0.17 for the 'student + supporter arm' in Chapter 5) but the experiment is underpowered, so larger-scale replications are needed to determine whether the treatment effects found in Chapter 4 can be replicated.

7.1.2 RQ 2: are 'direct' or 'social' text messages most effective?

The second experiment incorporated four trial arms to explore whether it was primarily the informational content of the texts that produce a treatment effect, or whether it is driven by activation of the social network. Differences between the treatment arms are not statistically significant. The proportion of students who achieved their qualification is larger in the 'supporter + student' arm but not distinguishable in size from the 'student only' arm. Further, the 'student only' texts resulted in better student outcomes than the 'supporter only' arm. These differences between the three treatment arms are not estimated with precision so they are interpreted as suggestive patterns.

The experiment also produced a surprising finding: the treatment that so successfully improved attendance and achievement in Chapter 4 did not produce positive effects in the follow-up field experiment, Chapter 5. In the follow-up experiment, the treatment in which only study supporters received the weekly texts resulted in a fairly clear null-result. Ex post, it is unclear why texting supporters only did not benefit students in the second year. I find no indication that students nominated 'inferior' supporters or that the treatment was not successfully delivered. Students assigned to the 'supporter only' group were no more or less positive about the programme than their peers assigned to the 'supporter + student' arm.

In sum, convincing answers to the second research question are wanting. The thesis delivers convincing evidence that text messages can help improve student success, but it is not yet clear whether these are best delivered directly to the student, or via their friends and family. Even so, the evidence points to the value of informing a close third party. After all, the most intensive treatment where both students and their supporters received the supportive text messages showed a larger treatment effect than informing students only. In Chapter 2, I hypothesised that social support from close others such as family and friends can help students cope better with stressors and improve their sense of self-confidence. The empirical findings support the hypothesis that greater involvement of third parties can boost student success. Further, preliminary qualitative evidence indicates that students' felt emotionally supported by their nominated study supporters which is in turn strongly predictive of better educational outcomes in adolescence (Malecki & Demaray, 2003; Wentzel et al., 2016). Nevertheless, direct communication with students appears a promising avenue for further research as well.

7.1.3 Effect sizes

The effect sizes of the study in Chapter 4 on attendance (3.1 - 4.8 % point increase in attendance rates, Hedges' g = 0.13) are similar to or larger than those found in similar interventions. For example, Chande et al. (2017) who sent weekly text messages to further education college learners, find a 6.6 % point increase in attendance from a control mean of 42.1%, Hedges' g = 0.06. Rogers et al. (2017) sent personalised postcards about the importance of attendance to over 50,000 students' homes and found a 2.4 per cent reduction in student absences, or Hedges' g = 0.02. Bergman and Chan (2017) find a 17% increase in attendance, Hedges' g cannot be calculated due to incomplete reporting of standard errors and sample sizes. As summarised above, the second follow-up field experiment did not yield statistically significant impacts on attendance.

The effect sizes of the intervention on achievement were more pronounced than for attendance and are larger than or comparable to those reported in similar intervention studies. This thesis found a 6 % point increase in pass rates (Chapter 4, Hedges' g = 0.16) [53] and 7.4 % point increase for the best-performing treatment in Chapter 5 (Hedges' g = 0.17)[54]. These treatment effects are only somewhat smaller than found in the study by Chande et al. (2017), where receiving weekly text messages about maths or English improved pass rates at post-16 institutions by 8.7 % points (Hedges' g cannot be calculated due to missing standard deviations). Berlinski et al. (2016) find a positive impact of 2.8% points on the proportion of students achieving the cut-off for passing the subject, and Bergman and Rogers (2017) find a 10 per cent reduction in the number of courses failed for treated students, or Hedges' g = 0.06. In sum, the supportive communication intervention may be a promising tool in the educator's toolkit.

7.1 Contributions to the academic literature

This thesis focused on enhancing involvement of students' existing social ties. The literatures on social support and information interventions were interrogated in Chapter 2. This thesis makes contributions to the scholarly debate within both fields by offering fresh evidence that existing dyadic relationships can be leveraged to help build behaviour change and improve educational outcomes. Furthermore, the embedded mixed methods design is an innovation within the rapidly growing literature on information interventions in education. The qualitative data enabled a greater understanding of the social context in which the studies were situated. Finally, this thesis offers contributions for policy by presenting robust evidence that greater student support can be delivered at a low cost. These contributions are expanded below.

7.1.1 Reflections on the social support literature

The literature on social support offers several directions for intervention design. First, the literature offers descriptive evidence that perceived support may be more important for individual wellbeing than enacted or received support (Gottlieb & Bergen, 2010). The qualitative findings suggest that general perceptions of the availability of support were enhanced by the interventions. Students were often unable to recount specific episodes of social support and instead felt encouraged that they could rely on their study

[53] Using the estimates of the simple model where the outcome variable is regressed on the treatment indicator.
[54] Ibid.

supporter should it be necessary. This opens up the possibility that the simple act of nominating a study supporter confers benefit to students. A future replication study could include a control condition in which students cannot nominate a third party, alongside a control group where they nominate an individual who is then not contacted (i.e. current design of the control group).

Further, researchers have attempted to explain the contradictory evidence on the effectiveness of social support provision by introducing the concept of invisible support (Bolger et al., 2000; Howland & Simpson, 2010; Maisel & Gable, 2009). Maisel and Gable (2009) argue that responsiveness of the support provider to the needs of the recipient is paramount. If the support provider is able to provide gentle support without making the recipient aware they are doing so, recipients may feel guilt or indebtedness to a lesser degree. The two field experiments did not offer guidance to study supporters on how to provide skilful support. However, the text messages always offered general and positive conversation topics.

The balance between the provision of social support and at the same time making sure it is not too obvious has received growing attention. A recent experimental study finds that it can be ineffective to clearly communicate supportive intentions (Hooker, Campos, & Pressman, 2018). Participants who were about to undergo a stressful task received text messages from their romantic partners. Unbeknownst to them, the text messages were scripted by the experimenter. When the text messages were mundane (i.e. "It is cold in here", Hooker et al., 2018, p. 488) participants' showed lower blood pressure responses to stressors. When the texts were clearly supportive (i.e. "don't worry, it's just a psych study. You'll be fine :)", Hooker et al., 2018, p. 488) participants' cardiovascular responses were greater. The authors argue that supportive text messages can inadvertently create evaluative threat, by subtly suggesting that the recipient of the texts *needed* the support (Hooker et al., 2018). Testing different configurations of the text messages with more mundane versus supportive prompts appears to be a relevant avenue for further research. Further, future studies may benefit from surveying students to assess whether they perceive the interactions with study supporters to be supportive to their needs or convey (stress-inducing) social evaluation. Both types of responses have been found in the experimental literature (Bolger & Amarel, 2007; Maisel & Gable, 2009).

The social support literature also clearly stipulates that relationship closeness is a key determinant of effective support provision. Qualitative results in Chapters 4, 5 and 6 suggests that students who nominated close relationships engaged in supportive communication with their study supporters. Those who nominated mere acquaintances did not appear to have built positive interaction habits. These indicative patterns underline the importance of a well-designed on-boarding process where students are guided towards choosing a trusted and skilled study supporter. Yet, it is encouraging to see statistically significant treatment effects in this thesis, where students' were not consistently guided towards choosing skilled supporters. If anything, this shows that social support interventions are a promising avenue even when the pool of potential support providers is diverse in age, qualification levels, and closeness to students.

7.1.2 A new approach to information interventions

This thesis illustrates that information interventions applied to education settings can be informed by social support theory. The recent and growing body of information interventions have thus far exclusively focused on the parent-child relationship and rely on personalised information about students' performance in school. Most information interventions reviewed in Chapter 2 are defined by their focus on providing parents with detailed information about their child's in-class behaviour, homework completion, and exam performance. For example, Bergman and Rogers (2017) alert parents if their child had a missing assignment, a class absence or a low average course grade. These automated alerts help overcome parents' (upward) biased beliefs about their child's performance in school, lower the cost of monitoring, and increase the salience of monitoring benefits.

In contrast, the field experiments in this thesis only share general and often prospective information about the maths and English courses. The text messages set out in this thesis never disclose students' in-class behaviour or achievement. Instead, the messages generally attempt to convey the value of engagement with learning. The text messages are written in a way to encourage third parties such as family and friends to initiate supportive conversations with students and build habits of engagement. This subtle distinction is also found in the details. The majority of published parent-texting interventions approach parents to request participation (e.g. Berlinski et al., 2016; Smythe-Leistico & Page, 2018); the student has no say in the matter. In contrast, the two field trials presented in this thesis give students full autonomy over the choice of recipient of the weekly communications.

Finally, the information interventions offered in this thesis are also distinctly easy to implement because they do not require teachers to supply frequent information about students' behaviour. Most information interventions (e.g. Bergman & Rogers, 2017; Cunha et al., 2017; Kraft & Monti-Nussbaum, 2017; Kraft & Rogers, 2015) alert parents about poor attendance or missed assignments, essentially to enable parents to better monitor their child's progress in school. In order to deliver such interventions, research teams collect weekly teacher ratings of test results (Berlinski et al., 2016), on-task behaviour (Kraft & Dougherty, 2013) or missed assignments (Bergman & Chan, 2017). In these studies, every text message contains personalised information about the child's progress. The Study Supporter intervention (Chapter 4) and its successor, Project College Success (Chapter 5), are not personalised to each individual student. The texts are written using the Scheme of Work (i.e. curriculum plan) and are fully scheduled weeks in advance, requiring little to no teacher involvement. This feature makes the intervention scalable and low-cost, especially in comparison to other education interventions aimed at improving student retention.

7.1.3 Policy relevance and practical applications

The cost of each experiments was considered in the respective chapter discussions, but here I consider the cost in comparison to other programmes. Text-messaging interventions such as the ones described in this thesis deliver supportive communication at a fraction of the cost of more intensive academic support programmes that involves professional support staff such as mentors or coaches. For example, Van der Steeg, van Elk, and Webbink (2015) evaluated an intensive coaching programme where per class a full-time coach contacted students when they were absent from class, visited their home, and taught them how to plan their learning. This programme, unsurprisingly, cost €3000 per student per year (Van der Steeg et al., 2015). Yet its estimated effect was similar to the treatment effects reported in this thesis, as dropout was reduced by approximately 7.3 % points (Van der Steeg et al., 2015).

Another experiment compared the effectiveness of personalised coaching versus a programme of generally informative texts about college (Oreopoulos & Petronijevic, 2017). The general text messages were sent by a team inside the college and provided students with general academic and study preparation advice. As such, the intervention is similar to the Study Supporter programme of texts. The researchers contrast this general text-message support with a treatment in which trained senior-year students act as coaches. The coaches pro-actively contacted students, asked if they needed help with anything, and built trust over time. In other words, the coaches fulfilled a support role. The general text-messages did not produce any positive treatment effects, but the coaching treatment improved course grades significantly. This coaching programme was several orders of magnitude more expensive than the general programme of texts: it cost $13,000 to service seventeen students whereas the general texts cost only $1,200 for 1500 participants (Oreopoulos & Petronijevic, 2017).

This thesis presents a twist on the general texts and in-person coaching treatment. Rather than training (and paying) students to pro-actively initiate contact, friends and family were encouraged to take up this role. The benefits of this approach are multifold: the student and supporter already know (and trust) each other and the texts can still be crafted by college staff. It was estimated that the year-long programme cost approximately £45 per student, including administrator and research costs.

7.2 Limitations

The research design could have been improved in three key areas: recruitment difficulties and resulting sample size, limited outcome data and contamination.

Sample sizes were smaller than expected, especially in the second experiment. Since the intervention relies on active opt-in, the final sample size was known only after all intervention materials were prepared. Ex post, one or more treatment arms should have been dropped from the Project College Success trial. The sample size of 250 students per arm was too ambitious, partly due to over-optimistic assumptions that fed into the power calculation. I had assumed that baseline attendance and achievement would explain approximately 30% of the variation in outcomes and student-level covariates would explain a further 20%. In fact, baseline achievement and student-level covariates combined explained a mere 12.6% of variation. The optimistic power calculations resulted in a minimum detectable effect size (MDES) of 6.6% points for a four-arm trial and in total 1000 students. In other words, the proportion of students achieving a passing grade would have to be 6.6% points higher in the treatment group in comparison to the control. Since the Study Supporter trial resulted in a 6 % point improvement in achievement, this MDES was deemed realistic. Using the corrected assumptions results in a MDES of 10.8% points. An improvement of almost 11 percentage points is quite unfeasible, given the average achievement rate of students on maths and English retakes: 23.6% (Department for Education, 2017). In Project College Success, 20.2% of students in the control group achieved a C or above. In this context, achieving a 10.8% point improvement would have been very remarkable indeed. The 7.6% point improvement in attainment for students in the 'supporter + student' group was not statistically significant ($p = 0.066$) whereas the 6% point improvement in the Study Supporter trial did reach the benchmark for statistical significance ($p = 0.008$). It stands to reason that the smaller sample size and resulting insufficient power at least partly explain this difference.

Second, outcome data was exclusively collected via college administrative data. Further education colleges collect attendance and achievement regularly and faithfully because their year-on-year funding allocation depends on accurate reporting to the Education Skills Funding Agency (ESFA). Unfortunately, they do not collect additional outcomes of potential interest, such as student in-class behaviour. Similar text-messaging interventions, carried out in primary- and secondary school settings, have examined the effect of the weekly teacher-family communications on homework completion, on-task behaviour, and classroom participation (Kraft & Dougherty, 2013). Then again, class attendance and course achievement are arguably the most important behavioural variables of student engagement. If students fail to pass the final exam, students have to re-take the same qualification year-on-year. Yet, it would have been interesting and relevant to collect more data about the interactions between students and their study supporters. This point is further addressed in the section on directions for future research.

Finally, it only became apparent that some students had nominated classmates who also took part in the trial after randomisation. Chapter 5 presents two analyses to explore to what degree this issue caused biased treatment effects (see p. 184). These analyses (1) re-assigned students to their observed treatment or (2) dropped these contaminated students from analysis altogether. The causal inferences did not change as a result of this exercise. The trial procedure would ideally have guarded against this problem in the first place, by checking duplicate phone numbers prior to randomisation. Further, spillover was difficult to assess in the current trials because class-level data was not available. Future experiments should explore spillover by exogenously varying the proportion of students allocated to treatment within classes (for example, see Berlinski et al., 2016).

7.3 External validity

Both experiments likely suffered from self-selection of field collaborators. The colleges are unlikely to be representative of the 200+ further education colleges in England, as senior staff at each of these colleges pro-actively reached out to take part in intervention research. The senior staff at most colleges was relatively engaged with the research efforts, so it is ambiguous what the uptake of the intervention might be in the absence of senior-management championing. Such selection into the experiment at the college level impacts the external validity of the results (Belot & James, 2016). As a simple fix, future studies should pro-actively contact colleges instead of relying on switched-on college principals to initiate contact. More generally, the nature of selection into the experiments should be more clearly documented. Belot and James (2016) suggest that field experimental researchers should present summary statistics on colleges that responded and contrast these with approached colleges who did not do so (similarly, colleges who ended up participating versus colleges who dropped out).

Further, both experiments required participants to opt in. The student opt in rate was 35% and 65% in the first and second trial, respectively. We found evidence of positive selection bias in the Study Supporter trial (Sanders & Groot, *forthcoming*, see Appendix 44, p. 289 - 303*)*. The nature of the intervention makes it more challenging to design an opt-out procedure. After all, students have to provide contact details of an individual of their own choice. Future experiments could improve recruitment materials and tutor induction which could hopefully motivate the least engaged students to sign up.

7.4 Implications for future research

There are a number of open questions. First, greater personalisation and interactivity of text message content could lead to greater effectiveness of the intervention. Due to resourcing limitations as well as the intention to keep the implementation light-touch for college staff, the text messages were not personally tailored to students. Students and their supporters were also not encouraged to text back and initiate a conversation. Such two-way communication would require an administrator's assistance for the duration of the academic year. Questions were only occasionally asked, and I responded by asking the individual to get in touch with the tutor directly. Due to this lack of interactivity, our text messages may not have been able to establish rapport and trust with recipients. Finally, the messages were signed off with '#SUCCESS' or 'the Project Success team'. Two alternative approaches could be tested: concluding the text with the college name or signing off with the tutor name.

Another question arising from this thesis is whether and how the intervention may interact with other channels of communication. Schools and colleges now routinely use text-messaging platforms to contact students and their guardian with frequent and personalised information about absences or exam dates. Colleges participating in the two field experiments were asked to limit their communication with students and parents to procedural topics including classroom changes and revision reminders. However, in the absence of intervention protocols, the intervention may at times duplicate ongoing communication channels. It is not known whether receiving multiple texts from multiple sources attenuates the treatment effect.

Similarly, a fruitful avenue for further research may be to test how varying frequencies, timings and content of text messages determine intervention effectiveness. Both studies reported in this thesis sent out one text message per week, at 7:00 PM on a Thursday evening. Other text-messaging studies have allowed a greater degree of tailoring of content and timing. For example, a weight loss intervention delivered daily via text message allowed participants to alter number and timing of the texts (Patrick et al., 2009). Text-messaging interventions delivered within education settings have not explored this level of tailoring to date. Allowing participants to alter content and timing potentially reduces scalability of the intervention. However, a low-cost approach to greater personalisation may be to connect the text-messaging platform to schools' management information systems. Bergman and Chan (2017) delivered high-frequency information via this integrated system, informing parents via automated alerts at extremely low cost.

Finally, the longer-run effects of the supportive communication intervention are unknown. Students in our sample were followed for the duration of a full-academic year, but it is relevant to further explore whether treated students are more likely to continue to study higher-level qualifications or gain employment. This question is particularly relevant because students are paired up with study supporters who ceased to receive weekly reminders at the end of the academic year. If the supporter terminates the supportive relationship abruptly after the experiment finishes, the intervention may backfire. After all, a key ingredient of effective social support is the stability of the supportive relationship (Baumeister & Leary, 1995).

The potential disruptive effect of discontinuation of a social support intervention is illustrated by a famous intervention in which elderly retirement home residents were paired up with student volunteers who visited them regularly (Schulz, 1976). During a two-month period, residents either received visits on a pre-defined schedule, a random schedule, or no visits from college students. Elderly residents who could predict visits used fewer medications and were rated as healthier than their peers who received random or no visits. Two to four years post-intervention, the residents assigned to the predictable visits, who initially benefited most from the intervention, accelerated in physical and emotional decline in comparison to elderly residents across the other groups (Schulz & Hanusa, 1978). This backfiring of the intervention was hypothesised to be the result of learned helplessness: the intervention gave residents a sense of personal control, and then abruptly removed this control after study termination (Schulz & Hanusa, 1978). Fortunately, students in our sample nominated individuals they were already well-acquainted with, so the risk of termination of the relationship is somewhat less plausible. Nevertheless, future studies could assess whether study supporters effectively built a positive habit to have conversations about learning or if they need the weekly reminders to engage. In other words, do study supporters who received the weekly communications continue to engage more in conversations than those who never received the treatment?

The literature on behavioural science interventions has extensively engaged in discussions about longer-run effects of the types of light-touch interventions (Allcott & Rogers, 2014; Frey & Rogers, 2014; Rogers & Frey, 2015). In this thesis, the processes set in motion by the treatments likely diverge depending on the recipient of the intervention. Study supporters are only an indirect target; they pass the text message content on to the student. On the one hand, they may simply be the conduits of messages, adding little value beyond communicating directly with students. On the other hand, study supporters may provide skilful social support, discussing the content of the text message but also providing emotional support to the student. In the latter case, a strong supporter-student bond may have formed over time. This scenario is a better recipe for persisting treatment effects than the former where supporters act as passive messengers.

The second potential recipient of the intervention, introduced in Chapter 5, is the student. Students who receive the direct communications may build habits of greater engagement over time because the text messages increase the salience of the importance of learning. Alternatively, the texted students may benefit through a simple reminder effect. In the latter case, the treatment effect is less likely to persist. Once the reminders cease, the student has access to fewer channels of personalised college information. The former scenario, where the student builds a habit of regular study time, may result in higher persistence of treatment effects even after the frequent communications have ceased.

Future work could also investigate more robustly the role of different sources of support. The studies introduced in this thesis did not introduce exogenous variation in the type of student-supporter relationship. It may be relevant to understand whether peer support or parental support manifests itself in different ways within the context of this intervention. Future intervention development work may benefit from a greater focus on these design choices. The social support literature suggests that both the source of support and type of support provided influences its effectiveness (Taylor, 2011). Future research may also explore the optimal closeness of the student-supporter relationship, whether co-habitation status determines treatment effectiveness, and whether weak or strong ties are most beneficial.

Finally, replication of the two studies introduced in this thesis has already commenced. The Education Endowment Foundation, an education funding body dedicated to generating robust evidence on the effectiveness of education innovations, is funding the follow-up trial to the two field experiments discussed here. The four-arm trial, a direct replication of the study design in Chapter 5, reached over 3800 students across 31 general further education colleges in the 2017-18 academic year. I was fortunate enough to lead this scale-up and further develop consent procedures and text-message content using the qualitative insights gained from this thesis. The trial is independently evaluated by NatCen Social Research and results will be published in Spring 2019. This study will contribute to a greater understanding of the effectiveness of this supportive communication intervention and may also elucidate whether the effects are greater when students receive the texts directly or when their study supporters are involved. Chapter 5 provided mixed evidence on the differential effects of the treatment arms, but the replication study will have greater statistical power to conduct these comparisons. Additionally, the trial is powered to detect moderate treatment effects for disadvantaged students (as measured by Free School Meals status). Chapter 2 briefly touched on the socio-economic gradient of parenting behaviours and access to supportive adults (Kalil, 2015). An intervention that seeks to redress imbalances in access to supportive close others may therefore have a disproportionate effect on disadvantaged students.

7.5 Conclusion

This thesis has robustly demonstrated that supportive text messaging programmes have the potential to help improve student success. The text messages were particularly effective at improving qualification achievement rates, by 6 % points to 7 % points across the best performing treatment arms in the two trials. The second experiment showed that most intensive treatment, a combination of student and supporter texts, was most effective at improving achievement.

The qualitative findings offer new insights into the dynamics of supportive interactions between students and their study supporter. Chapter 4 showed that students with a strong existing relationship with their nominated study supporter recounted supportive episodes whereas those who nominated mere acquaintances did not. The next experiment, Project Success, thus set out to better inform students about the importance of nominating a trusted relationship; someone they felt comfortable talking to about their maths or

English. The qualitative component of Chapter 5 finds that students and their study supporters appreciate the weekly prompts. Further, the qualitative data showed that study supporters felt encouraged to engage in more frequent and focused conversations with students about their learning. I conclude this thesis with a quote from a mother who was nominated by her son. Her reflection on the changing nature of their conversations captures the core aim of the thesis well:

> *It was more of a focused conversation. For example, when he [the student] was saying that there was a certain topic that he was studying or there was a maths exam that would prompt me to keep that conversation open or to start it. [Before] It would probably just be, have you done your work?*
> *[04ss01, Experiment 2]*

8 APPENDICES

8.1 Chapter 4 appendices

Appendix 1 *Descriptive statistics for sample colleges*

College	Students (N)		Achieved A*-C (%)	
	Maths	English	Maths	English
A	260	370	26.3	29.9
B	210	140	28.2	8.7
C	240	300	7.9	23.9
D	300	330	24.7	28.5
E	500	810	10.4	16.6
F	150	260	14.4	22.3
G	930	1130	16.7	16.4
H	960	1090	20.0	17.4
I	550	710	19.4	28.1

Appendix 2 *Student characteristics by survey completion mode*

	Online survey	Paper-based survey	Normalised difference	Online survey	Paper-based survey
	M (SE)	M (SE)	$\hat{\Delta}_{ct}$	N	N
Female (proportion)	0.486 (0.022)	0.539 (0.016)	0.074	516	935
Age (years)	17.124 (0.115)	19.514 (0.200)	0.339	516	935
White (proportion)	0.620 (0.021)	0.390 (0.016)	-0.316	516	935
GCSE (proportion)	0.744 (0.019)	0.501 (0.016)	-0.344	516	935
Maths (proportion)	0.841 (0.016)	0.418 (0.016)	-0.567	516	935
High self-reported belonging	0.693 (0.021)	0.593 (0.016)	-0.146	472	914
Qualification achievement rate	0.250 (0.019)	0.263 (0.014)	0.021	516	935
Study Supporter 1 age (years)	28.360 (0.642)	27.979 (0.445)	-0.02	500	866
Study Supporter 2: Age (years)	27.705 (0.697)	26.700 (0.496)	-0.051	420	733
On how many of the past 7 days have you spoken with SS1?	5.756 (0.085)	5.137 (0.078)	-0.199	499	935
On how many of the past 7 days have you spoken with SS2?	5.049 (0.105)	3.960 (0.090)	-0.291	469	935
Assignment to treatment group	0.494 (0.022)	0.506 (0.016)	0.016	516	935

Notes: Mean values reported with robust standard errors in parentheses for all continuous and binary variables. Age and communication frequency with study supporters are continuous variables, and all other variables are binary. Self-reported belonging was assessed by asking respondents "How often do you wonder: maybe I don't belong here?" measured on a scale from "never" to "always". Data are recoded so that high belonging signifies "never" and "rarely". The survey data was merged with the attainment dataset; therefore sample sizes correspond to the attainment data.

Appendix 3 Information sheet about Study Supporter intervention

The college wants to be able to give students the best possible learning experience. One way Harrow College is doing this is by being involved with research projects. These projects help the college to find out new ways of supporting its learners, such as using helpful text messages that could support your learning.

Participating in the Study Supporter Program means that:

- You will participate in your courses just as you normally would.
- We may send occasional text messages and/or emails to your "Study Supporter" encouraging them to talk to you about your studies at your College. The messages may be about upcoming deadlines, online resources that are available to you, and questions about what you are learning. You may also receive a copy of these messages.
- Your Study Supporter may be asked for feedback about their experience at the end of the project.
- You authorize your College to share your data with researchers of the Behavioural Insights Team. If you like, you can be provided with a record of any data that is shared.
- You authorize your College to contact your Study Supporters.

How will my and my Supporters' privacy be protected?

- The names of participant(s) who win one of the six £250 Amazon vouchers will be made publicly available.
- Only your unique learner number will be shared; your full name will not be shared with anyone outside the college.
- Researchers will have temporary access to your phone number for sending you and/or your Supporter text messages only.
- Your Supporter can unsubscribe from texts or e-mails at any time by responding "STOP."
- You and your Supporter will not be identifiable in any resulting research.

What are the possible risks and benefits of participation in this project?

- You and your Supporter will not be paid for participating. You will be entered to win a £250 Amazon voucher and your chances of winning depend on the total number of participants but they will be approximately be one in thousand.
- We see no likely risks or discomforts for you or your Supporter.
- You may benefit from having a Study Supporter to encourage you during this term. Also, regardless of whether you are randomly chosen to have your Study Supporter contacted, this study will help you identify a friend or family member who could be an academic supporter for you.

If I have any questions or concerns about this project, who can I talk to?

If you have questions or concerns, you can speak to either your tutor or e-mail the Behavioural Insights Team at ask@behaviouralinsights.co.uk.

This sounds good. How do I participate?

- Please decide on two Study Supporters and ask them for permission to share their contact details with us so we can send them messages about your learning.

Appendix 4 Student sign-up survey and information sheet

You could win a £250 Amazon voucher by just filling out this form! {{College}} would like you to give us the names and contact details of two 'Study Supporters'. A Study Supporter is someone you think would be good at supporting you in your learning. It could be a friend, family member, colleague, or anyone else. If possible, <u>at least one of the Study Supporters should be someone you don't live with.</u>

We may send your Study Supporters messages about what you're learning, study tips or important dates (but never your grades). Research has found that Study Supporters can really help learners succeed! 6 people who complete this form will win a **£250 Amazon voucher**. The winners will be announced on Friday 23rd October. You will be automatically entered if you complete this form.
You can find more information on the rules of the lottery from your tutor, who has a copy of the prize drawing rules available.

Study Supporter 1	**Study Supporter 2**
First Name:	First Name:
Last Name:	Last Name:
Age:	Age:
Male/Female:	Male/Female:
Mobile phone:	Mobile phone:
E-Mail:	E-Mail:
Relationship to you:	Relationship to you:

Later, make sure you check with your Study Supporters that they are happy to receive messages about your learning. They will be able to stop the messages at any time.

We would like to know more about your Study Supporters. Please answer a couple of questions for us. First, please read the three questions below and circle the Study Supporter that fits best.

1. Which Study Supporter are you closer to?

Study Supporter 1 **or** **Study Supporter 2**

2. Which Study Supporter do you think would do the best job of supporting you at college?

Study Supporter 1 **or** **Study Supporter 2**

3. Which Study Supporter do you talk with more?

Study Supporter 1 **or** **Study Supporter 2**

*Now, for **<u>each</u>** of your Study Supporters, please tick **<u>one</u>** of the boxes for each Study Supporter:*

4. Do you live with your Study Supporters?

Study Supporter 1	**Study Supporter 2**
☐ Yes ☐ No	☐ Yes ☐ No

5. Compared to all the other people you know, how close are you to your Study Supporters?

Study Supporter 1	**Study Supporter 2**
☐ Not close at all ☐ A little close ☐ Close ☐ Very close	☐ Not close at all ☐ A little close ☐ Close ☐ Very close

6. On how many of the past 7 days have you spoken with your Study Supporters?

Study Supporter 1	**Study Supporter 2**
☐ 0 days ☐ 1 to 3 days ☐ 4 to 6 days ☐ All 7 days	☐ 0 days ☐ 1 to 3 days ☐ 4 to 6 days ☐ All 7 days

7. Do your Study Supporters work or study at your College?

Study Supporter 1	**Study Supporter 2**
☐ Yes ☐ No ☐ I don't know	☐ Yes ☐ No ☐ I don't know

8. Did your Study Supporters get 5 GCSEs or more (or similar exams, like O Levels)?

Study Supporter 1	**Study Supporter 1**
☐ Yes ☐ No ☐ They are taking GCSEs now ☐ I don't know	☐ Yes ☐ No ☐ They are taking GCSEs now ☐ I don't know

9. Do your Study Supporters have a university degree?

Study Supporter 1	**Study Supporter 2**
☐ Yes ☐ No ☐ They are currently in university ☐ I don't know	☐ Yes ☐ No ☐ They are currently in university ☐ I don't know

Appendix 5 Example Study Supporter Programme text messages

Type	Text message
Intro text	Hi [SS forename], [learner forename] [surname] would like your help. [He/She']s asked us at [college] to send you weekly texts so you can support them in their [English/Maths/Numeracy/Literacy] class. It's half term now, please ask [him/her] how the class is going so far. Reply STOP if you don't want to receive these messages.
General	Hi [SS forename], [learner forename] returned to their [English/Maths] class this week. Please ask if [he/she] has made a plan for when, where, and how [he/she] plans to study going forward. Thanks, [College]
FS MAT	Hi [SS forename], please ask [learner forename] to think of something that was challenging this week and what she can discuss about it in [his/her] next Maths class. Thanks, [College]
FS MAT	Hi [SS forename], please ask [learner forename] what [he/she] is most proud of accomplishing in [his/her] maths class this term. Thanks, [College]
FS MAT	Hi [SS forename], please encourage [learner forename] to think about how this maths class can help [him/her] do well in [his/her] vocational course. Thanks, [College]
FS ENG	Hi [SS forename], it's Scottish and Australian Week in English class next week, for [learner name]. Ask [him/her] to tell you more about the Scottish culture and language. Thanks, [College]
FS ENG	Hi [SS forename], encourage [learner forename] to take plenty of time to proofread after finishing a writing task. This will really help [him/her] notice spelling and grammar mistakes. Thanks, [College]
FS ENG	Hi [SS forename], [learner forename] is preparing for [his/her] speaking and listening assessments, happening in English class next week. Ask [him/her] how [he/she] is preparing for them. Thanks, Great Yarmouth College
GCSE MAT	Hi [SS forename], [learner forename] has recently learnt about percentages. Ask [him/her] to calculate the final price of a £250 TV after adding 20% VAT (tax on things you buy) and show you how [he/she] worked it out. Thanks, [College]
GCSE MAT	Hi [SS forename], [learner forename] is learning about Pythagoras' theorem in maths class. Ask if [he/she] has checked out the BBC Bitesize website, which has a great explanation on this topic. [He/she] can find it at http://www.bbc.co.uk/schools/gcsebitesize/maths/geometry Thanks, [College]
GCSE MAT	Hi [SS forename], [learner forename] is learning about statistics. Please ask [him/her] to explain the difference between the mean and the median and how to calculate them. Thanks, [College]

GCSE MAT	Hi [SS forename], [leaner forename] has a milestone assessment (test) in maths class next week. Ask how [he/she] is revising for the assessment - every bit counts! Thanks, College
GCSE ENG	Hello [SS forename], next week, [learner forename] will start on creative writing in English class. Ask [him/her] what writers can do to grab the reader's attention. Thanks, [College]
GCSE ENG	Hi [SS forename], ask [learner forename] when reading & writing has come up in [his/her] daily life this week. For example, has [he/she] read anything interesting in the news? Thanks for helping [him/her] succeed in college! Thanks, [College]
GCSE ENG	Hi [SS forename], next week is the last controlled assessment (exam) in English class for [learner forename]. It is a creative writing task about 'Shutter Island'. Please ask [him/her] to describe the story to you. Thanks, [College]
GCSE ENG	Hi [SS forename], we wanted to thank you for your continued support of [learner forename]! Please remember that just having a chat with [him/her] can make a big difference, especially now the GCSE English exam is happening soon, on June 7th. Thanks, [College]
GCSE ENG	Hi [SS forename]. The BBC skillswise website has lots of useful practice materials. Encourage [learner forename] to set aside some time to practice spelling on /www.bbc.co.uk/skillswise Thanks, [College]
General	Hello [SS forename], please ask [learner forename] whether [he/she] is involved in extracurricular activities at Lakes College (for example, sports or drama) or whether [he/she] would like to be? Getting involved in college life is a great way to meet new people and get extra skills. Thanks, [College].
General	Hello [SS forename] it's half term for [learner forename] next week. Please ask [him/her] what [he/she] plans to revise over the break. Ask if [he/she] has already started preparing for reading and writing assessments. Thanks, [College]

Appendix 6 Effect size formulae

All formulae are taken from (Enzmann, 2015, p. 2 - 3)

Cohen's d:

$$d = \frac{\bar{x}_t - \overline{x_c}}{s^*}$$

where:

$$s^* = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}}$$

and:

$\bar{x}$	= sample means;
t and c	= denotes treatment and control group;
s^*	= pooled within sample estimate of the population standard deviation;
s_t and s_c	= standard deviation of treatment and control groups, respectively; and
n_t & n_c	= sample sizes.

Cohen's *d* where standard deviations are not known but F statistics are available:

$$d = \sqrt{F\left(\frac{n_t + n_c}{n_t * n_c}\right)\left(\frac{n_t + n_c}{n_t + n_c - 2}\right)}$$

F	= F-test statistic

Hedges g:

$$g = d * c(m)$$

Where:

$$m = n_t + n_c - 2$$

And:

$$c(m) \approx 1 - \frac{3}{4m - 1}$$

The approximation of c(m) above is used for the calculations. The exact formula is:

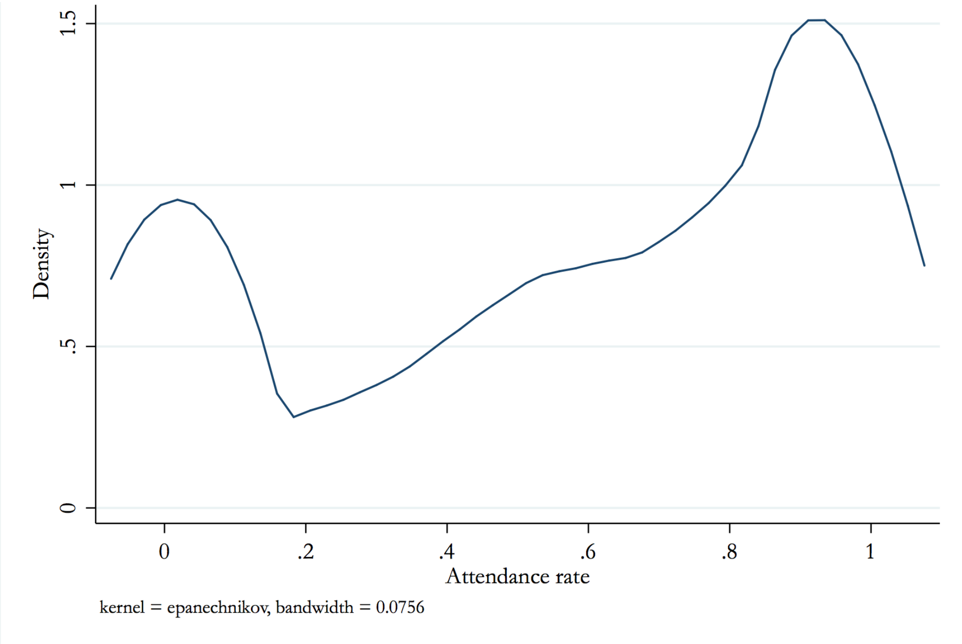
$$c(m) \approx \frac{\Gamma(\frac{m}{2})}{\sqrt{\frac{m}{2}} * \Gamma(\frac{m-1}{2})}$$

Finally, Glass' Δ is calculated using the following formula:

$$\Delta = \frac{\bar{x}_t - \bar{x}_c}{s_c}$$

Where $\boldsymbol{s_c}$ is the within sample estimate of the population standard deviation in the control group.

Appendix 7 *Kernel-density plot of attendance rate*



Appendix 8 *Outcome data by college and treatment group*

	Attendance			**Attainment**		
College	Control	Treat	*Total*	Control	Treat	*Total*
College A	73	58	*131*	63	52	*115*
College B	136	145	*281*	101	122	*223*
College C	27	25	*52*	-	-	-
College D	172	165	*337*	142	153	*295*
College E	127	134	*261*	128	121	*249*
College F	47	47	*94*	40	42	*82*
College G	107	98	*205*	91	77	*168*
College H	113	132	*245*	109	125	*234*
College I	18	14	*32*	49	36	*85*
Total	820	818	*1,638*	723	728	*1451*

Appendix 9 *Balance between treatment and control groups, attainment dataset*

Covariates	Control M (SE)	Treat M (SE)	Normalised difference ($\widehat{\Delta}_{ct}$)
Male (proportion)	0.489	0.471	-0.024
	(0.019)	(0.019)	
Age (years)	18.520	18.806	0.038
	(0.184)	(0.206)	
White (proportion)	0.477	0.467	-0.014
	(0.019)	(0.019)	
GCSE (proportion)	0.558	0.617	0.085
	(0.018)	(0.018)	
Maths (proportion)	0.570	0.566	-0.008
	(0.018)	(0.018)	
Observations	723	728	

Notes: Mean values reported with robust standard errors in parentheses for all continuous and binary variables. Age and belonging are continuous variables, and all other variables are binary. Normalised differences are calculated using. Data on student demographics was merged with the achievement outcome data rather than the attendance dataset, therefore the sample size used to assess balance corresponds to analyses reported for attainment.

Students on GCSE courses are overrepresented in the treatment group (61.7%) in comparison to the control group (55.8%). To ensure this imbalance in qualification type does not affect the results, the regressions control for qualification type, although the results are robust to omitting this covariate.

Appendix 10 *Distribution of nominated study supporters*

Relationship type	Study Supporter 1		Study Supporter 2	
	N	%	N	%
Parent	431	29.7	270	18.6
Sibling	158	10.9	135	9.3
Extended family	76	5.2	100	6.9
Significant other	57	3.9	29	2
Peer	600	41.4	571	39.4
Child	8	0.6	7	0.5
Professional support	32	2.2	28	1.9
Colleague	86	5.9	7	0.5
Missing	3	0.2	304	20.9
Total	1451	100	1451	100

Notes: Parents include stepmother/father as well as guardian. Extended family includes grandparents, aunts, uncles, nieces and nephews. Significant other indicates spouse, boyfriend or girlfriend. Peer includes friends and classmates, although very few students categorised their relationship as 'classmate' (N = 22). Professional support comprises teachers, tutors, social support workers, or other support staff at the college.

Appendix 11 *Descriptive statistics of study supporter demographics*

Question	Category	Study Supporter 1 N (%)	Study Supporter 2 N (%)
Do you live together?	Yes	639 (44.0)	358 (24.7)
	No	762 (52.5)	883 (60.9)
	Missing	50 (3.5)	210 (14.5)
Do they have 5 GCSEs?	Yes	553 (36.7)	474 (32.7)
	Studying towards	147 (10.1)	135 (9.3)
	No	278 (26.0)	452 (43.0)
	I don't know	375 (25.8)	350 (24.1)
	Missing	20 (1.4)	40 (2.8)
Do/did they attend university?	Yes	166 (11.4)	120 (8.3)
	Studying towards	56 (3.9)	49 (3.4)
	No	978 (67.4)	866 (59.7)
	I don't know	189 (13)	197 (13.6)
	Missing	52 (4.3)	219 (15.0)
How close are you?	Very close	815 (56.2)	588 (40.5)
	Close	408 (28.1)	411 (28.3)
	A little close	135 (9.3)	181 (12.5)
	Not close at all	40 (2.8)	64 (4.4)
	Missing	53 (3.7)	207 (14.3)
Question	Category	M (SD)	M (SD)
What is their age?	Parent	43.7 (7.1)	44.2 (7.1)
	Sibling	21.7 (6.2)	21.3 (5.1)
	Extended family	33.0 (18.7)	36.3 (20.7)
	Significant other	25.1 (9.6)	21.1 (6.3)
	Peer	19.5 (7.2)	19.3 (6.0)
	Child	18.4 (8.9)	14.3 (3.5)
	Professional support	39.3 (12.1)	40.4 (11.2)
	Colleague	17 (0)	20.2 (5.0)
	Missing (relationship type)	22.1 (9.7)	28.6 (15.5)

Appendix 12 *Comparisons of closeness between both supporters*

Question		%
Who are you closer to?	Study Supporter 1	56.4
	Study Supporter 2	22.3
	Equally close	20.8
	Missing	0.5
Who supports you best?	Study Supporter 1	58.8
	Study Supporter 2	33.0
	Equally well	1.2
	Missing	7.0
Who do you speak to more often?	Study Supporter 1	60.7
	Study Supporter 2	28.3
	Equally often	1.5
	Missing	9.5

Appendix 13 *Effect sizes, primary analyses*

	(1)	(2)	(3)	(4)
	Attendance rate		Achievement rate	
	Basic model	*Inc. covariates*	*Basic model*	*Inc. covariates*
Treatment	0.048**	0.031*	0.071**	0.060**
	(0.018)	(0.015)	(0.023)	(0.022)
Constant	0.555**	0.619**	0.223**	0.159**
	(0.013)	(0.039)	(0.015)	(0.056)
Student-level covariates	*No*	*Yes*	*No*	*Yes*
College fixed effects	*No*	*Yes*	*No*	*Yes*
Cohen's d	0.129	0.105	0.163	0.136
Hedges g	0.129	0.105	0.163	0.136
Glass' delta	0.126	0.082	0.171	0.143

Appendix 14 *Effect sizes, heterogeneous treatment effects on attendance*

	(1) GCSE	(2) FS	(3) 16- 18	(4) 19+	(5) Female	(6) Male
Treatment	0.030	0.029	0.042*	0.001	0.004	0.053**
	(0.020)	(0.022)	(0.017)	(0.030)	(0.022)	(0.020)
Constant	0.671**	0.609**	0.644**	0.542**	0.662**	0.600**
	(0.052)	(0.055)	(0.035)	(0.114)	(0.072)	(0.048)
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cohen's d	0.095	0.070	0.116	0.001	0.001	0.135
Glass' delta	0.091	0.068	0.112	0.001	0.001	0.133

Notes: All analyses are OLS regressions, including fixed effects at the college level. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) and missingness dummies as pre-specified. Huber white standards errors, clustered at the student-level, in parentheses. Effect size calculations use unconditional standard deviations, and covariate adjusted means are used for all calculations. + = p < 0.10, * = p<0.05, ** = p<0.01.

Appendix 15 *Effect sizes, heterogeneous treatment effects on achievement*

	(1) GCSE	(2) FS	(3) 16- 18	(4) 19+	(5) Female	(6) Male
Treatment	0.051+	0.064+	0.073**	0.021	0.065*	0.058+
	(0.030)	(0.033)	(0.026)	(0.043)	(0.032)	(0.032)
Constant	0.207**	0.232*	0.163**	0.106	0.036	0.221**
	(0.068)	(0.105)	(0.057)	(0.129)	(0.080)	(0.074)
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cohen's d	0.115	0.152	0.171	0.044	0.144	0.138
Glass' delta	0.119	0.161	0.182	0.045	0.151	0.145

Notes: All analyses are OLS regressions, including fixed effects at the college level. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) and missingness dummies as pre-specified. Huber white standards errors in parentheses.
+ = p < 0.10, * = p<0.05, ** = p<0.01.

Appendix 16 *Logistic regression: ATE of the intervention on achievement*

	(1)	(2)
	Basic model	*Inc. covariates*
Treatment	1.453**	1.398**
	(0.176)	(0.175)
Student-level covariates	No	Yes
College fixed effects	No	Yes
Marginal Effect	0.071	0.060
N Observations	1451	1451
Pseudo R^2	0.006	0.058

Notes: Odds ratios and marginal effects reported in table. Logistic binary regression, includes as covariates student-level covariates and college fixed effects. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) as pre-specified. Robust standard errors in parentheses. McFadden's R^2 reported in table. += p < 0.1 *, = p < 0.05, ** = p <0.01

Appendix 17 *Logistic regression: treatment effect on achievement rate by subgroups*

	(1) GCSE	(2) FS	(3) 16-18	(4) 19+	(5) Female	(6) Male
Treatment	1.315^{+}	1.528^{*}	1.501^{**}	1.171	1.439^{*}	1.403^{+}
	(0.210)	(0.329)	(0.214)	(0.356)	(0.254)	(0.260)
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes
College fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Marginal effect	0.051	0.064	0.073	0.022	0.065	0.058
N	852	592	1114	337	755	696
Pseudo R^2	0.060	0.147	0.019	0.309	0.099	0.030

Notes: Odds ratios and marginal effects reported in table. Logistic binary regression, includes as covariates student-level covariates and college fixed effects. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) as pre-specified. Robust standard errors in parentheses. McFadden's R^2 reported in table. += p < 0.1 *, = p < 0.05, ** = p <0.01

Appendix 18 *Characteristics of selected colleges for qualitative follow-up*

College	Region	Subject implemented	Ofsted rating	Implementation difficulties	N students in trial
College A	South East - rural	Maths	Good	low tutor engagement	131
College C	East of England - rural	English	Required improvement	IT and classroom disruption	52
College H	Greater London - urban	Maths	Outstanding	No	245

Appendix 19 Interview information sheet and consent form

What is the purpose of this project?
The Behavioural Insights Team, a research company, is working with your College to better understand what you think about studying at a Further Education College and what motivates you to learn. We are also interested how you've found the exercise(s) you did this year, and what you've learnt and how you found the session(s). We are going around other Colleges around the UK too, and trying to find out what participants most enjoy and find useful so we can help the College support their students better in the future. We would like you to participate in this study, but you do not have to.

What happens if I take part in this project?

- You will be interviewed by a researcher from the Behavioural Insights Team for approximately 60 minutes.
- The interviewer will take notes on your answers, and also make an audio recording of the interview.
- You authorise to share the interview data with other researchers from the Behavioural Insights Team. If you like, you can be provided with a record of any data that is shared.
- If at any point of the interview you may wish to withdraw, you may do so.
- You will be paid £10 in Amazon vouchers for participating.

If I participate, how will my privacy be protected? What happens to our data?

- All information collected during the interview will be kept strictly confidential and only used for this project. We will use a random number to label and store data instead of your name.
- We may provide your school with a general summary of insights. This summary will not identify you.

If I have any questions or concerns about this project, whom can I talk to?

- If you have questions or concerns, you can speak to Bibi Groot at the Behavioural Insights Team. Email: bibi.groot@behaviouralinsights.co.uk. Mobile: +44 7938871985.

This sounds good. How do I participate?

- Please sign the consent form attached.

What if I change my mind?

- If you change your mind, you can withdraw at any time by e-mailing Bibi Groot (bibi.groot@behaviouralinsights.co.uk). We will destroy all records of your interview if you make this request.

CONSENT FORM
If you are happy to take part in the interview, please sign this form.
Your name:
Your Student ID:
Date:
Signature:

Appendix 20 *Student interview schedule*

Main objective	**Purpose of section**	**Guide timings**
1. Introductions and background	Explains the purpose and ground rules for the interview.	5 mins
2. Social interaction mapping exercise	Participatory exercise to visualise the interviewee's social network.	15 mins
3. Social Support and Social belonging	Focused on subjective experiences of social support from peers, tutors, parents and other adults.	25 mins
4. Implementation evaluation	Focused on the design and implementation of the Study Supporter Programme.	8 - 10 mins
5. Close	Thank you and close	2 mins

Appendix 21 *Tutor interview guide*

Main objective	**Purpose of section**	**Guide timings**
1. Introductions and background	Explains the purpose and ground rules for the interview.	5 mins
2. Implementation evaluation	Series of more structured questions on the implementation of the interventions.	25 mins
3. college belonging	Focused on tutors' perspectives on supporting learners and college environment.	25 mins
4. Close	Thank you and close	2 mins

Appendix 22 *Tutor self-rated assessment of implementation readiness*

[rating between 0 (completely disagree) – 10 (completely agree)]

Self-rated implementation readiness (1 - 10)		
I would implement these exercises again next year with a good deal of enthusiasm	10	1
The exercise logistics/procedures easily fit in with my current practices	8	7
I understand the procedures of these exercises	7	10
A positive home–college relationship is needed to get the most out of the exercises	10	10
The total time required to implement the exercises was manageable	10	8
Material resources needed for the study are reasonable	10	3
The logistics of administering the exercises are too complex to carry out accurately *(reverse-scored)*	8	9
Use of these exercises is consistent with the mission of my College	10	10
My work environment is conducive to implementation of exercises like these	10	10
I would need additional resources to carry out this study next year *(reverse-scored)*	9	9
Overall rating	9.2	7.7

8.2 Chapter 5 appendices

Appendix 23 *Student sign-up survey including guidance on choosing a supporter*

Welcome! Your College is participating in a project to help learners succeed in their GCSE maths and English. Many learners like you have already signed up!

- Where do you study?
- Which courses are you taking?

Please click "Next" to watch a short video about the project.
Video: https://youtu.be/WjY-yS9c2Kk

We will never text your supporters about your grades or attendance.
You'll need to provide a name and a mobile phone number for each person you nominate. If you can only think of one Study Supporter, that's fine too. As an added bonus, if you sign up, you can win one of six £25 Amazon vouchers. Winners will be contacted on the 1st of November.

If you click "yes" below, you agree for the College to share your attendance and attainment data with the project team.

Your data will always be locked away safely and is never shared with anyone outside the project team.
For more information about the prize drawing rules and the Study Supporter programme, see www.projectcollegesuccess.co.uk

Yes, I want to participate
No, I do not want to take part

- What letter grade do you expect to receive for GCSE maths this year?
    - A* - U
- What letter grade do you expect to receive for GCSE English this year?
    - A* - U
- Are you resitting the course?
    - Yes, I took maths/English GCSE before at this college
    - Yes, I maths/English GCSE before, but at a different school/college
    - No, this is the first time I am taking maths/English GCSE

- What is your...
- First name
- First letter of last name
- Student ID or reference number
- What is your mobile phone number? Please type without any spaces.
- What is your date of birth? (dd/mm/yyyy)
- Are you..?
    - Male
    - Female
- Do you still live at home with your parents or guardians?
    - Yes
    - No, I live by myself
    - No, but I do live with family

The guidance on nominating a suitable study supporter was presented here, see Appendix 2.

- Study Supporter 1:
- First name
- Last name
- What is their mobile phone number?
- What is their gender?
    - Male
    - Female
- What best describes your relationship? He/she is your: (select one)
    - Father
    - Mother
    - Sister
    - Brother
    - Grandmother
    - Grandfather
    - Cousin
    - Aunt
    - Uncle
    - Friend outside college
    - Friend inside college
    - Coworker (someone you work with)
    - Tutor/teacher
    - Guardian
    - Carer
    - Partner/spouse
    - Boyfriend
    - Girlfriend
    - Family friends
    - Boss
    - Daughter
    - Son
    - Mentor
    - Coach
    - Roommate
    - Other

- Pick the top reason why you chose [SS name] to be your Study Supporter.
    - He/she is there when times are tough
    - He/she makes me feel safe
    - He/she believes in me
    - He/she inspires me
    - He/she goes out of his/her way to help me
    - He/she can make me laugh whenever
    - He/she gives good advice
    - He/she tells the truth
    - He/she helps keep me on track towards my goals
    - He/she brings out the best in me

- Do you live with ${q://QID8/ChoiceTextEntryValue/1}?
    - Yes
    - No

- Compared to all the other people you know, how close are you to [SS name]
    - Very close
    - Close
    - A little close
    - Not close at all

- In the last 7 days, on how many days have you spoken with [SS name]?
    - 0 = less than once a week
    - 1 = once a week
    - 7 = every day

- How do you usually talk to each other? You can tick multiple boxes.
    - Mostly face-to-face
    - Mostly via text, whatsapp, Facebook, messenger, snapchat, etc.

    - Mostly by phone or skype

- What is [SS name]'s first language?
    - English
    - Polish
    - Punjabi
    - Urdu
    - Arabic
    -
    - Spanish
    - Bengali
    - Tamil
    - Turkish
    - Other

- Does [SS name] understand written English, for example in a text message?
    - Yes
    - No

- Is [SS name] your classmate or your tutor?
    - Yes, [SS name] is my classmate in this subject
    - Yes, [SS name] is my classmate in my vocational course but not in this subject
    - Yes, [SS name] is my tutor in this subject
    - Yes, [SS name] is my tutor in my vocational course but not in this subject
    - No, [SS name] is not my classmate or tutor

- Did [SS name] get 5 GCSEs or more (or similar exams, like O Levels)?
    - Yes
    - No
    - [SS name] is taking GCSEs now
    - I don't know

- Does [SS name] have a university degree?
    - Yes
    - No
    - [SS name] is currently at university
    - I don't know

Click "Next" to enter the details of Study Supporter 2.

<u>The above questions are repeated for the second nomination.</u>

<u>For students who nominated two study supporters:</u>
- Which Study Supporter are you closer to?
    - Study Supporter 1
    - Study Supporter 2

- Who do you think will be better at helping you succeed in your course(s)?
    - Study Supporter 1
    - Study Supporter 2

- Which Study Supporter do you talk with more?
- Study Supporter 1
- Study Supporter 2

- Do [SS name] and [SS name] know each other?
    - Yes
    - No

Later, make sure you check with your Study Supporters that they are happy to receive messages about your learning.
They will be able to stop the messages at any time.
Please click 'Next' to answer a few questions about yourself.

- Use the slider (you can drag it from left to right) to indicate how you feel right now.
    - ______ How much do you feel like you belong at your college?

Answer If Do you still live at home with your parents or guardians? Yes Is Selected

- Do your parents/guardians know the parents of your friends in College?
    - Yes
    - No

Answer If Do you still live at home with your parents or guardians? Yes Is Selected

- What is the highest level of education your father / mother has completed?
    - University degree
    - College/vocational degree
    - Upper secondary school (A levels)
    - Lower secondary school (GCSE, O grade)
    - No schooling completed
    - I don't know

You are almost done. Only a few questions left!
Tick the box that best reflects how much these statements are true for you.

- If I put in enough effort, I can succeed in maths
- Whether or not I do well in maths is completely up to me
- Family demands or other problems prevent me from putting a lot of time into my maths homework
- If I had different teachers, I would try harder in maths
- If I wanted to, I could do well in maths

Scoring: strongly disagree (1) to strongly agree (5).

- I work hard at college
- I concentrate on my homework
- I am a responsible student
- I complete my homework regularly

Scoring: strongly disagree (1) to strongly agree (5).
<u>Repeat the above two scales if students study both subjects.</u>

Appendix 24 Guidance, choosing a suitable supporter

<u>This guidance was presented after the consent procedure, and before nominating a study supporter.</u>

Your Study Supporter(s) can be anyone who is 16 or over that you talk to regularly and cares about your success in college. Students in colleges around England have chosen parents, siblings, cousins, colleagues, friends, football coaches, aunts, grandparents, relationship partners, teachers, godfathers, and others.

Let's think about who you might like to be your Study Supporters.
We found it can really help if you nominate two people. One should be someone you live with and the other person should be someone you don't live with.
First, think about:

- someone who cares about you and how you do in college
- someone who is around when you're in need
- someone who brings out the best in you

Or maybe you prefer to think about a person you would go to if...

- you were feeling lonely or nervous?
- you were having a tough time completing your homework?
- you needed some advice?

You don't need to think of a person who fits all of the questions.
A Study Supporter can be anyone, really! As long as you talk to each other regularly.

Can you think of someone? Tip: you can ask your tutor for help if you're not sure.

- Yes, I can think of 2 people!
- Yes, I can think of 1 person
- No

If you're struggling to come up with a person who could be your Study Supporter, you might want to ask your tutor for help. Or maybe you can get some inspiration from the people other learners have chosen:

- ... parents, carers, brothers, sisters, cousins,
- ... friends, coaches, aunties, grandparents, family friends
- ... partners, teachers, godfathers, social workers

If you still cannot think of anyone who might be a good Study Supporter, ask you tutor. If you can only think of one, that's fine!

Appendix 25 *Ex ante power calculations, unadjusted*

	Student opt-in scenarios			
	500	750	1000	1250
Trial arm scenarios	MDES δ (N per arm)	MDES δ (N per arm)	MDES δ (N per arm)	MDES δ (N per arm)
2 Arms	0.119 (250)	0.096 (375)	0.083 (500)	0.074 (625)
3 Arms	0.147 (166)	0.119 (250)	0.102 (333)	0.091 (416)
4 Arms	0.170 (125)	0.138 (188)	0.119 (250)	0.106 (313)

Notes: Power was set at 0.80 and alpha at 0.05. The allocation proportion was set equal group sizes. Unadjusted power calculations.

Appendix 26 Example text messages, by content type

Topic	Recipient	Text message content
Course content	Supporter	Hi {supporter first name}, ask {student first name} to explain the 4 conditions that can be used to determine whether two shapes are congruent.
	Student	Hi {student first name}, your study supporter [SS name] received a message about the 4 conditions that can be used to determine whether two shapes are congruent. Please explain it to [him/her] when you see them!
Upcoming exam deadlines	Supporter	Hi {supporter first name}, it is just a week until {student first name} takes their first GCSE maths exam! It would be a good time to ask them how their revision is going. #SUCCESS
	Student	Hi {student first name}, it is just a week until your first GCSE maths exam! Why not chat with {supporter first name} about how your revision is going? #SUCCESS
Extra tutorial sessions	Supporter	Hello {supporter first name}, please ask {student first name} if they have been to any extra tutorial sessions. The Learning Resource Centre offers sessions that can be booked directly to suit their schedule.
	Student	Hello {student first name}, don't forget to speak to {supporter first name} about the extra tutorial sessions [College] offers. The Learning Resource Centre offers sessions that can be booked directly!
Academic resources	Supporter	Hello {supporter first name}, {student first name} is learning about descriptive writing skills in {{custom2}} class. Why not suggest they take a look at the BBC Bitesize website which has lots of extra resources? {{url=http://www.bbc.co.uk/education/guides/zx499j6/revision}} #SUCCESS
	Student	Hi {student first name}, please ask your Study Supporter {supporter first name} about the website link we sent over to them. This website has a lot of extra resources to help you learn about descriptive writing skills. #SUCCESS
General motivation	Supporter	Hi {supporter first name}. Sometimes, when things are tough, remind yourself: "Life isn't about waiting for the storm to pass. It's about learning to dance in the rain". Why not text this quote to {student first name}, and tell them you're there when they need it? #SUCCESS
	Student	Hi {student first name}. Sometimes, when things are tough, remind yourself: "Life isn't about waiting for the storm to pass. It's about learning to dance in the rain". Don't forget: {supporter first name} can support you when you need it! #SUCCESS

Appendix 27 Example text messages, by treatment group

Treatment group	Subject	Recipient	Text message content
Supporter only	English	Supporter	Hi {supporter first name}, term starts next Tuesday (18th) for {student first name}. They will have a mock exam to prepare for the real GCSE exams happening this summer. Please encourage {student first name} to prepare by practicing past papers. That way, {student first name} will learn how the questions are asked and what good answers look like. #SUCCESS
Student only	English	Student	Hi {student first name}, term starts next Tuesday (18th). You will have a mock exam to prepare for the real GCSE exams happening this summer. Tip: prepare for the mock exam by practicing past papers. That way, you will learn how the questions are asked and what good answers look like. #SUCCESS
Supporter + Student	English	Supporter	*Identical to supporter text in 'supporter only' group*
	English	Student	Hi {student first name}, term starts next Tuesday (18th). You will have a mock exam to prepare for the real GCSE exams happening this summer. Tip: prepare for the mock exam by practicing past papers and let {supporter first name} know how it's going. That way, you will learn how the questions are asked and what good answers look like. #SUCCESS
Content-based licensing	English	Supporter	Hi {supporter first name}, we just texted {student first name} with an English riddle. One possible answer is SWIMS. Why not talk to {student first name} to give them the answer & find out what the question was? If they text us back with the answer they can win a £5 voucher! #SUCCESS
	English	Student	Hi {student first name}, here's an English riddle: what is a word which looks the same upside down? Not sure? We have sent {supporter first name} the answer. Text us back with the answer and we'll send a £5 M&S voucher to the first three correct replies! #SUCCESS
Supporter only	Maths	Supporter	Hi {supporter first name}, {student first name} has been learning about 'vectors' and 'scalars' in GCSE Maths this year. Ask {student first name} to explain what vectors and scalars are and how they are different from each other. If they aren't sure, they can ask their tutor. #SUCCESS

Student only	Maths	Student	Hi {student first name}, you have been learning about 'vectors' and 'scalars' in GCSE Maths this year. Make sure you understand what vectors and scalars are and how they are different from each other. If you are not sure, you can ask your tutor. #SUCCESS
Supporter + Student	Maths	Supporter	*Identical to supporter text in 'supporter only' group*
	Maths	Student	Hi {student first name}, you have been learning about 'vectors' and 'scalars' in GCSE Maths this year. Explain to {supporter first name} what vectors and scalars are and how they are different from each other. If you are not sure how, you can ask your tutor. #SUCCESS
Content-based licensing	Maths	Supporter	Hi {supporter first name}, we just texted {student first name} with a maths riddle. The answer is "they are the same". Why not talk to {student first name} to give them the answer & find out what the question was? If they text us back with the answer they can win a £5 voucher! #SUCCESS
	Maths	Student	Hi {student first name}, here's a maths riddle: which is heavier, a kilogram or feathers or a kilogram of bricks? We have sent {supporter first name} the answer. Text us back with the answer and we'll send a £5 M&S voucher to the first three correct replies! #SUCCESS

Appendix 28 Student responses to opt-out questionnaire

Statement	Frequency	% agree
I don't want to participate in research	189	50%
I cannot think of anyone	119	31.5%
I don't want to get messages	101	26.7%
I don't know the mobile number(s) of my study supporters	56	14.8%
I don't have a mobile phone	28	7.4%
I am afraid to ask people to be my study supporter	28	7.4%

Notes: non-consenting students were asked to tick all statements they agreed with. They could tick multiple boxes, or tick none. Total N = 378.

Appendix 29 *Covariate balance between groups with and without missing data*

	Complete case		Missing attendance data		Normalised difference
	M (SE)	N	M (SE)	N	$\hat{\Delta}_{ct}$
Gender: male (proportion)	0.514	923	0.426	47	-0.123
	(0.016)		(0.073)		
Age (years)	18.918	923	21.595	47	0.255
	(0.181)		(1.245)		
Ethnicity: white (proportion)	0.340	923	0.298	47	-0.064
	(0.016)		(0.067)		
Baseline grade (0 – 1)	0.634	656	0.536	16	-0.356
	(0.008)		(0.035)		
Baseline attendance	80.126	923	67.500	2	-0.434
	(0.746)		(2.500)		
First time resit at FE college	0.430	920	0.468	47	-0.053
	(0.016)		(0.074)		

Appendix 30 *Descriptive statistics of study supporter demographics*

Question	Category	Treated Study Supporter	
		N	%
Do you cohabit with [Study Supporter]?	Yes	478	49.03
	No	483	49.54
	Missing	14	1.44
Does [Study Supporter] have 5 GCSEs?	Yes	381	39.08
	Studying towards	115	11.79
	No	187	19.18
	I don't know	269	27.59
	Missing	23	2.36
Did [Study Supporter] achieve a university degree?	Yes	148	15.18
	Studying towards	48	4.92
	No	608	62.36
	I don't know	148	15.18
	Missing	23	2.36
How close are you to [Study Supporter]?	Very close	628	64.41
	Close	238	24.41
	A little close	78	8.00
	Not close at all	17	1.74
	Missing	14	1.44
Does [Study Supporter] understand written English?	Yes	935	95.79
	No	23	2.36
	Missing	18	1.85
Of the past 7 days, on how many days did you speak with [Study Supporter]?	None	5	0.51
	1 – 3 days	86	8.82
	4 – 6 days	218	22.36
	Every day	645	66.14
	Missing	21	2.15

Appendix 31 *Effect size calculations for primary analyses*

Outcome	Treatment group	Type	Effect size
Attendance rate	Supporter only	Cohen's d	0.022
		Hedges' g	0.022
		Glass' Δ	0.023
	Student only	Cohen's d	-0.076
		Hedges' g	-0.076
		Glass' Δ	-0.079
	Supporter + Student	Cohen's d	0.045
		Hedges' g	0.045
		Glass' Δ	0.047
Achievement rate	Supporter only	Cohen's d	-0.029
		Hedges' g	-0.029
		Glass' Δ	-0.029
	Student only	Cohen's d	0.100
		Hedges' g	0.100
		Glass' Δ	0.104
	Supporter + Student	Cohen's d	0.173
		Hedges' g	0.173
		Glass' Δ	0.183

Notes: effect size calculations are presented for the simple model reported in column 1 of Table 5.10 and Table 5.11, where the treatment indicator is regressed on the outcome of interest. In the interest of parsimony, effect sizes for the pilot group ('content-based licensing') are not reported; this comparison is underpowered due to its small sample size.

Appendix 32 *Average treatment effects of intervention on achievement, logistic regression*

	(1) Simple model	(2) Including covariates
Supporter only	0.930	0.883
	(0.220)	(0.219)
Student only	1.274	1.246
	(0.288)	(0.291)
Supporter + student	1.502^{+}	1.311
	(0.333)	(0.305)
Content-based licensing pilot	1.520	1.391
	(0.502)	(0.497)
Student-level covariates	No	Yes
College fixed effects	No	Yes
N	970	970
Pseudo R^2	0.006	0.092

Notes: Odds ratios and marginal effects reported in table. Logistic binary regression, includes as covariates student-level covariates and college fixed effects. Student-level covariates include age, gender, subject (maths/English), and resit status as pre-specified. Robust standard errors in parentheses. McFadden's R^2 reported in table. += p < 0.1 *, = p < 0.05, ** = p <0.01

Appendix 33 *Heterogeneous treatment effects of intervention on achievement, logistic regression*

	(1) English	(2) Maths	(3) Female	(4) Male
Supporter only	1.387	0.509^{+}	0.654	1.269
	(0.488)	(0.192)	(0.217)	(0.515)
Student only	1.275	1.208	0.631	2.643**
	(0.443)	(0.378)	(0.202)	(0.976)
Supporter + student	1.709	0.965	0.732	2.592*
	(0.578)	(0.320)	(0.227)	(0.965)
Content-based licensing	2.159	0.892	0.708	3.158*
	(1.142)	(0.431)	(0.360)	(1.698)
N	472	498	476	494
Pseudo R^2	0.118	0.084	0.102	0.105

Notes: Odds ratios and marginal effects reported in table. Logistic binary regression, includes as covariates student-level covariates and college fixed effects. Student-level covariates include age, gender, subject (maths/English), qualification type (GCSE/FS) as pre-specified. Robust standard errors in parentheses. McFadden's R^2 reported in table. += p < 0.1 *, = p < 0.05, ** = p <0.01

Appendix 34 *Instrumental variable estimates of CACE effect, upper and lower bound*

	≤ 25% of texts received			≤ 50% of texts received			≤75% of texts received		
Variables	(1) First stage	(2) Attendance rate	(3) Achieve-ment rate	(1) First stage	(2) Attendance rate	(3) Achieve-ment rate	(1) First stage	(2) Attendance rate	(3) Achieve-ment rate
Supporter only - assigned	0.843** (0.027)			0.810** (0.030)			0.759** (0.033)		
Student only - assigned	0.812** (0.027)			0.775** (0.030)			0.735** (0.033)		
Supporter + student - assigned	0.940** (0.027)			0.926** (0.0300			0.895** (0.033)		
Supporter only – alerted		0.032 (0.020)	-0.029 (0.043)		0.033 (0.021)	-0.030 (0.044)		0.036 (0.022)	-0.032 (0.047)
Student only - alerted		-0.001 (0.02)	0.034 (0.044)		-0.001 (0.021)	0.036 (0.046)		-0.002 (0.023)	0.038 (0.049)
Supporter + Student - alerted		0.019 (0.018)	0.048 (0.038)		0.019 (0.018)	0.049 (0.039)		0.020 (0.019)	0.051 (0.040)
Student-level covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations		923	970		923	970		923	970

Notes: Treatment effects on primary pre-registered outcome variables are estimated using 2SLS regressions with the instrumented alerted variable, an indicator for students and supporters who received the full schedule of 35 texts. All regressions include a set of student-level demographic covariates, of gender, age, subject, and resit status. Baseline achievement was not available for 30% of our sample and therefore not added into the model in order to preserve sample size. Standard errors in parentheses.
+ = $p < 0.10$, * = $p<0.05$, ** = $p<0.01$.

Appendix 35 *Reallocated treatment groups based on observed noncompliance*

		Original allocation (at randomisation)					
		Control	Supporter only	Student only	Supporter + student	Content-based licensing	**Total**
Re-allocated, observed	Control	**205**	-	-	-	-	**205**
	Supporter only	-	**194**	-	-	-	**195**
	Student only	22	16	**230**	21	4	**293**
	Supporter + student	-	17	-	**207**	-	**224**
	Content-based licensing	-	-	-	-	**59**	**59**
	Total	**227**	**227**	**230**	**228**	**63**	**975**

Notes: Students were re-allocated to the treatment they received due to having been nominated as a supporter. Students who were originally allocated to control but who received texts about their classmate's learning were re-assigned to the 'student only' group. Second, students who were originally allocated to 'supporter only texts' but also received texts as their classmate's supporter were reassigned to the 'student + supporter group'. Finally, students who nominated themselves as supporter were re-assigned to 'student only' texts if they had originally been assigned to any supporter-texts arm.

Appendix 36 *Average treatment effect on attendance, re-allocating contaminated students*

	(1) Simple CCA	(2) Simple, imputation	(3) Inc. covariates CCA	(4) Inc. covariates, imputed
SS only	-0.007	-0.004	0.016	0.015
	(0.022)	(0.022)	(0.018)	(0.018)
Student only	-0.026	-0.024	-0.013	-0.013
	(0.020)	(0.020)	(0.016)	(0.016)
SS + Student	-0.002	-0.001	0.003	0.003
	(0.022)	(0.022)	(0.018)	(0.017)
Content-based licensing	-0.018	-0.019	-0.026	-0.024
	(0.036)	(0.035)	(0.025)	(0.025)
Constant	0.731**	0.734**	0.220**	0.219**
	(0.015)	(0.015)	(0.053)	(0.053)
Student-level covariates	No	No	Yes	Yes
College fixed effects	No	No	Yes	Yes
Observations	923	975	923	975
R-squared	0.002	0.002	0.344	0.350

Notes: As described in the main text, in total 80 students (8% of full sample) received other treatments than allocated due to having nominated classmates who were also part of the intervention, or due to having nominated oneself. The same student-level pre-treatment covariates are added to the models as in the primary regressions. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1

Appendix 37 *Average treatment effects on attendance, removing contaminated students*

	(1) Simple CCA	(2) Simple, imputation	(3) Inc. covariates CCA	(4) Inc. covariates, imputed
SS only	-0.007	-0.004	0.015	0.014
	(0.022)	(0.022)	(0.018)	(0.018)
Student only	-0.027	-0.024	-0.014	-0.013
	(0.022)	(0.022)	(0.018)	(0.017)
SS + Student	0.001	0.003	0.004	0.004
	(0.022)	(0.022)	(0.018)	(0.018)
Content-based licensing	-0.018	-0.019	-0.027	-0.025
	(0.036)	(0.035)	(0.025)	(0.025)
Constant	0.731**	0.734**	0.234**	0.233**
	(0.015)	(0.015)	(0.055)	(0.055)
Student-level covariates	No	No	Yes	Yes
College fixed effects	No	No	Yes	Yes
Observations	847	895	847	895
R-squared	0.003	0.002	0.344	0.351

Notes: Students who were nominated by a classmate, and therefore contaminated by their classmate's participation in the experiment (N = 80), were removed from the models reported in this table. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1

Appendix 38 *Average treatment effect on achievement, re-allocating contaminated students*

	(1) Simple CCA	(2) Simple, imputation	(3) Inc. covariates CCA	(4) Inc. covariates, imputed
SS only	-0.017	-0.019	0.036	-0.018
	(0.040)	(0.040)	(0.046)	(0.039)
Student only	0.013	0.012	0.054	0.014
	(0.038)	(0.037)	(0.042)	(0.036)
SS + Student	0.071+	0.071+	0.088+	0.056
	(0.041)	(0.041)	(0.045)	(0.040)
Content-based licensing	0.088	0.078	0.074	0.072
	(0.067)	(0.066)	(0.067)	(0.062)
Constant	0.210**	0.210**	-0.033	-0.092
	(0.029)	(0.029)	(0.106)	(0.138)
Student-level covariates	No	No	Yes	Yes
College fixed effects	No	No	Yes	Yes
Observations	970	975	672	975
R-squared	0.007	0.007	0.113	0.096

Notes: As described in the main text, in total 80 students (8% of full sample) received other treatments than allocated due to having nominated classmates who were also part of the intervention, or due to having nominated oneself. The same student-level pre-treatment covariates are added to the models as in the primary regressions. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1

Appendix 39 *Average treatment effects on achievement, removing contaminated students*

	(1) Simple CCA	(2) Simple, imputation	(3) Inc. covariates CCA	(4) Inc. covariates, imputed
SS only	-0.017	-0.019	0.035	-0.018
	(0.040)	(0.040)	(0.046)	(0.039)
Student only	0.035	0.034	0.077+	0.035
	(0.040)	(0.040)	(0.045)	(0.039)
SS + Student	0.070+	0.070+	0.083+	0.053
	(0.042)	(0.042)	(0.046)	(0.041)
Content-based licensing	0.088	0.078	0.074	0.072
	(0.067)	(0.066)	(0.067)	(0.062)
Constant	0.210**	0.210**	-0.022	-0.094
	(0.029)	(0.029)	(0.109)	(0.142)
Student-level covariates	No	No	Yes	Yes
College fixed effects	No	No	Yes	Yes
Observations	890	895	624	895
R-squared	0.007	0.007	0.116	0.093

Notes: Students who were nominated by a classmate, and therefore contaminated by their classmate's participation in the experiment (N = 80), were removed from the models reported in this table. Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1

Appendix 40 Interview information sheet and consent form

We'd like to invite you to take part in a study about studying maths and English GCSEs at College, and how you are supported in your learning. This document tells you a bit more about the study and asks for your consent to take part. Please read the information carefully and if you have any questions just ask myself.

Who is doing this study?

The study is being done by the Behavioural Insights Team. We are an organisation which conducts research to try and improve how people experience the world around them. You might have seen text messages coming from 'Project Success'. We are working with the College to try to improve student motivation to achieve their maths and English GCSEs, and the text messages and these interviews are part of this project.

Why are we doing this?

We are trying to learn more about how you study for your English and/or maths GCSEs, and who supports you in your learning. You have been chosen to participate in the study because we are interested in your experience of nominating a Study Supporter earlier this year, who receives weekly messages about what you are learning at College. Your opinions are very important to us.

What will I be asked to do?

You will be asked to answer a few questions we have prepared. The interview should take around 45 minutes. We will record this interview so we can type out the interview. This makes it easier for us to look at your answers and write up a story about our findings. If you don't want to be recorded, please let the interviewer know.

What do you do with my data?

The recording will be stored securely on a password-protected hard drive. Only the interviewer and the transcriber will have access to the recordings, and the recording will not have your name attached to it. We might like to include things that you have told us in a report but we will not mention your name or any of your friend's names. Anything you tell us will be kept confidential unless we think you might be at risk of harm in which case we would tell someone who could help you.

What happens if I no longer want to take part?

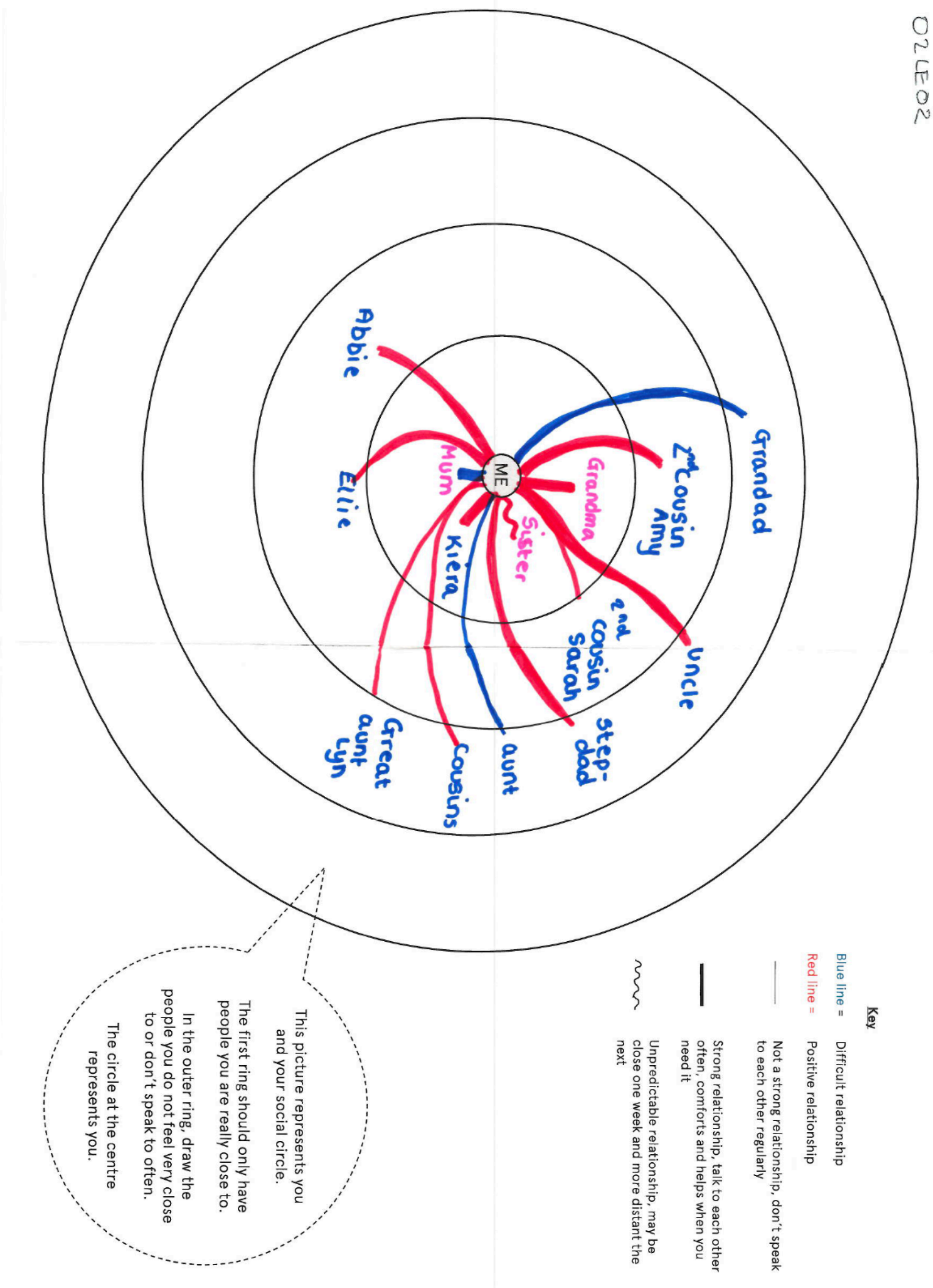
You are free to decide whether you'd like to take part. To give consent, please sign the form below and give it back to the interviewer. You can change your mind at any time about participating. You don't need to give a reason to withdraw.

If you have any questions, you can email us at bibi.groot@bi.team.

Appendix 41 *Student interview schedule*

Main objective	Purpose of section	Guide timings
1. Introductions	Explains the purpose and ground rules for the interview.	5 mins
2. Study behaviours	Understanding students' study behaviour and how they seek and receive support.	10 mins
3. Social support map	Exploring students' relationships with key individuals.	10 mins
4. Feeling supported	Understanding students' personal experiences of having a study supporter and what college support looks like.	10 mins
5. Implementation of trial	Get feedback on project implementation and content and frequency of texts.	7 mins
6. Close	Thank you and close.	2 mins

Appendix 42 Example output of the social mapping exercise



Appendix 43 Treatment assignment by interview participants

Treatment group	Experiment 1: Study Supporter	Experiment 2: Project Success
Control	1	1
Supporter only	4	4
Student only	-	1
Supporter + student	-	9
Total	5	15

Appendix 44 *Sanders & Groot (forthcoming). How important is selection bias in educational research? Evidence from a field experiment. Widening Participation and Lifelong Learning.*

Sanders, Michael, Behavioural Insights Team and University of Oxford Blavatnik School of Government

Groot, Bibi, Behavioural Insights Team and University College London School of Public Policy

Email: Michael.sanders@bi.team

Abstract Education research has undergone a revolution as the use of randomised controlled trials has become more prevalent. There remains however some debate as to the relative virtues of RCTs. It is widely recognised that randomised trials can, if well designed, produce unbiased estimates of treatment effects. However, what is unclear is to what extent second-best forms of evaluation, such as propensity score matching, are comparable to RCTs in terms of their accuracy. In this paper we exploit a dataset containing both random assignment and the opportunity to evaluate quasi-experimentally the effect of an intervention designed to increase student attendance at further education colleges. We find not only that there is substantial evidence of selection bias, but also that this bias is only marginally reduced by the inclusion of covariates, and by using propensity score matching.

Key words methodology; randomised controlled trials; bias; further education

Introduction and Context

The field of education research has been transformed in the last ten years by the dramatic rise in the number of randomised controlled trials (RCTs) being conducted. The Higher Education Funding Council for England (HEFCE), for example, has championed the experimental approach to assess the impact of the National Collaborative Outreach Programme (NCOP), in order to build a more systematic evidence base. Robust evidence on approaches to increase the proportion of young people from identified areas is currently lacking; a gap the NCOP research programme aims to address. Similarly, the Education

Endowment Foundation (EEF) has commissioned 145 RCTs to examine 'what works' in closing the attainment gap since its foundation in 2011, involving over 970,000 children and young people (The Education Endowment Foundation, 2018). The Department for Education's decision to fund the Behavioural Insights Research Centre for Adult Skills in Knowledge (ASK) in 2014 showed a desire to see more randomised experiments carried out in lifelong and adult learning, including among those returning to education.

The case for conducting such trials has been made in considerable detail elsewhere (see Fryer, 2017; Sadoff, 2014; Whitehurst, 2012), but it remains important to highlight that experimental methodologies can be especially valuable. If well designed, RCTs can isolate causal relationships from contextual factors (Cook, Campbell, & Shadish; 2002). Non-experimental studies cannot rule out third variables influencing the relationship between two variables such as attending university (cause) and job prospects (effect). For example, intelligence, parental education or family socio-economic status could cause both success in education and a higher income, complicating the inferences that can be made from such studies. Randomised controlled trials, through random assignment, are uniquely capable of removing such selection bias in comparisons between treated and non-treated individuals (Athey & Imbens, 2017).

At the same time, however, there has been considerable criticism of the use of randomised controlled trials, both by practitioners (Smith, 2013) and some academics (Christ, 2014; Schanzenbach, 2012). We will not debate here the ethical or logistical arguments against randomised trials, not shall we consider the (in our view incorrect) argument that RCTs stifle innovation. Instead, this paper seeks to contribute to the debate on the use of RCTs in education by considering how it fares compared with other methodologies in terms of handling selection bias.

Selection bias occurs when individuals or organisations are able to self-select into the treatment (Belot & James, 2016). In the absence of randomisation, universities with strong operational efficacy or managerial ability might be able to run more effective programs, leading to an overestimation of its impact (Alcott, 2015). Similarly, students with higher intrinsic motivation to achieve their qualification might be more likely to volunteer to take part in an educational intervention. Both the university's strong existing operational efficacy and the student's intrinsic motivation are likely correlated with outcomes. Indeed, they might have achieved positive outcomes even in the absence of the intervention. The concern here, is that those who are most likely to benefit from the treatment self-select into the programme (Gertler et al., 2016; List & Rasul, 2011). The degree to which methodologies in education research are able to account for and overcome selection bias, is therefore of prime importance.

Similar approaches have been taken by Arceneaux et al (2006). Our study is also similar to Belot and James (2016), who consider which schools select to be part of field experiments. Our setting is unusual though, in that it considers post-16 learners in Further Education college settings in England. These learners attended full-year qualifications in maths and English. In an attempt to improve basic skills levels in the UK, students who failed their GCSEs (A* - C or 9 - 4) at age 16 are now required to retake maths and/or English qualifications. Our sample consists of both students aged 16 – 18 (who are mandated into learning) and aged 19+, but the vast majority fall within the former category (84%). Further Education college learners typically have more autonomy to choose whether or not they want to take part in an intervention than school-aged pupils and hence may be more prone to selection bias.

Additionally, empirical comparisons between popular education methodologies are rare (see Belot & James, 2014; 2016), even though the issue of selection bias is at the heart of our ability to assess the effectiveness of education policy and practice.

In this short paper, we consider an unusual case – an RCT in which consent for an intervention is administered through an opt-in process (whereby participants must give formal consent to be treated[55]), but data is collected on an opt-out basis (whereby participants' outcome data is collected by default, unless participants choose to opt out of data sharing). This comparatively novel situation gives us a dataset in which assignment to treatment is random for the consenting subset, but we possess outcome measures for those who did not self-select. This allows us to attempt non-experimental analysis of the effect of our intervention, and to compare this with our experimental evidence.

The experiment evaluated here is the "Study Supporter" component of the Behavioural Insights Team's large scale RCT in further education colleges, carried out in 2015-2016. The Behavioural Insights Team is a social purpose company part owned by the UK Government's Cabinet Office, having previously been a part of the cabinet office itself. It is considered to be the first organisation set up explicitly to apply behavioural science to public policymaking. The majority of the team's interventions are evaluated using Randomised Controlled Trials . The study described here was conducted as a part of a grant from the Department for Business, Innovation and Skills (BIS), to apply behavioural science to support further and adult education. This study is reported elsewhere in general by Hume et al. (2018a, 2018b), and specifically in Groot, Sanders, & Rogers (2017), and as such we do not detail the intervention in question in much depth here. For exposition, participants in the intervention are asked to nominate two family members, peers, or other key individuals in their social network as "study supporters". Maths and English tutors at participating Colleges were instructed to introduce the research study in the first weeks of the academic year, and invite learners to sign up either via an online

[55] In this trial, 'treated' indicates that (1) the learner consented to take part in the trial, and (2) was assigned to the treatment group. The nominated study supporters of treated individuals received weekly text messages about maths or English class. 'Consenting' indicates that the learner consented to take part, but was subsequently assigned to the control condition. Hence, their nominated supporter did not receive any text messages over the course of the academic year.

link or by completing a short paper-based survey. Learners were told that participation is entirely voluntary. Those who signed up were subsequently individually randomised into treatment and control, where study supporters of learners in the treatment group are sent short weekly text messages encouraging them to support the learner's learning goals, for the duration of the academic year.

The outcome of interest is class attendance in the treated subject, either maths or English. Both attendance and attainment are relevant policy-outcomes, as attendance is a better predictor of academic achievement than any other known predictor, such as study habits or standardised test scores (Credé, Roch & Kieszczynka, 2010).

We find significant evidence of selection bias where consent is not controlled for. Non-experimental attempts to reduce this bias do not substantially do so.

Data

Our data contain a subset of the approximately 18,000 observations from the randomised reported in Hume et al. (2018). In a randomised controlled trial, participants who have consented (in this case through an opt-in process), are randomly assigned to one of at least two conditions – a treatment (or intervention) group and a control group. Random assignment, with sufficiently large samples, ensures that the two groups are similar in terms of both observed and unobserved characteristics, and hence any ex post differences between the two groups can be said to have been caused by the treatment. Specifically, we are concerned with participants in that trial who attended one of 4 colleges- Great Yarmouth College (GYC), Lakes College West Cumbria (LCWC), Uxbridge College (UC), and West Hertfordshire College (WHC). In total, this dataset comprises 6089 participants – students attending the colleges who did not opt out of data collection. For all participants, we also have an indication of whether they consented to take part in the "study supporter" intervention, and for these participants, whether they were randomised to receive the intervention, or to be in the control group. The intervention consists of a series of text messages sent to a "study supporter" nominated by the student, providing information and updates about their studies. Only a subset of the total number of learners taking full-time maths or English qualifications opted in to take part, as shown in Table 1.

Table 1: Participant Numbers, by College

College	Total number of learners	Consenting (control)	Assigned (treatment)	Opt-in rate (across both conditions)
GYC	794	38	21	7.4 %
LCWC	1707	256	135	22.9%
UC	2842	359	244	21.1%
WHC	746	90	51	18.9%
Total	**6089**	**743**	**451**	**17.6%**

Covariates of interest include age (in years), gender, course type (whether the participant is studying for functional skills[56] or GCSE), and subject (Maths or English). For a subsample of participants we also possess data on age (3914 participants), and gender (3915 participants). By analysing data fields relating to the name of participants' courses, we are able to infer whether a participant is taking an English course (1688 participants) or a maths course (4302 participants), or both (98 participants). For each participant we are unable to identify whether they are studying English of Maths.

By a similar approach, we attempt to identify whether participants are taking Functional Skills (FS) courses, or GCSEs. This process produces a set of 2848 participants identified as doing GCSEs only, 3162 doing Functional Skills training only, 78 doing both, and 12 where this is not identified.

[56] Functional skills qualifications are have equivalence with GCSEs, but are more embedded with vocational studies, than the academic route of GCSEs. A maths FS qualification for example, might require a learner to learn about invoicing or ordering stock, and an English FS qualification might teach about writing CVs and punctuation.

Analysis

We now proceed to our analysis. To begin, we consider the extent to which our observed covariates are correlated with participants' decisions to consent to the trial. Table 2, below, contains a set of regression models in which the consent binary variable is regressed on each of our covariates in turn. Regression analysis, here using a linear prediction model, allows us to estimate the relationships between variables, holding other things constant, and is the standard approach to analysing data from randomised controlled trials. Where data are largely missing, as with age and gender, missing values are coded as 0 and a binary "was missing" indicator is used to avoid these individuals biasing the findings, without excluding them from analysis. Figures in this table can be interpreted as the difference in (in this case) consent rates when a particular variable (in the leftmost column) increases by one. For example, the value in model (1) for Functional Skills is -0.082, shows that a participants studying Functional Skills classes, is 8.2% points *lower* than for those studying GCSEs. In the case of age, in model (4), we can see that a 1 year increase in the participant's age is associated with a 0.1% point increase in consent rates.

The regression constant (at the bottom of each column) shows the rate of consent when the active variables in that model are equal to 0 – so, for example, that that consent rate for students studying GCSEs is 16.6% points.

Table 2: Observable characteristics of participants and the probability of consenting (Linear Prediction Model)

	(1)	(2)	(3)	(4)
Functional Skills	-0.082***			
	(0.008)			
Subject: Maths		-0.011		
		(0.009)		
Gender: male			0.005	
			(0.011)	
Gender missing			-0.112***	
			(0.011)	
Age				0.001
				(0.002)
Age missing				-0.098*
				(0.039)
Constant	0.166***	0.130***	0.160***	0.146***
	(0.006)	(0.008)	(0.008)	(0.038)
N	6089	6089	6089	6089

As can be seen above, participants who do not report their gender are significantly more likely to consent than those participants who do, while participants taking Functional Skills courses are significantly less likely to consent, compared to participants on GCSE qualifications. There is no clear relationship between subject taken, age, or gender and consent, at least in this sample. For the purposes of this data, we can therefore say that selection, if it is occurring, is on non-observables.

We now proceed to analyse the effect of our treatment under different models, in table 3. As in table 2, we make use of linear regression models to enable our analysis to identify the estimated relationship of interest - in this case, between the treatment and attendance rate. Although the models used here are slightly more complicated than those above, they can be interpreted in the same way as those in table 2. We use intention to treat (ITT) analysis, and so an individual is said to be treated in this analysis if they consented and were randomly assigned to treatment. Column 1 of table 3 reports the results of our the primary analysis of the randomised trial, with an estimation for selection bias. In this column, we estimate a simple linear regression of attendance rate on participants consent status, their treatment assignment and a set of college fixed effects. Column 2 conducts the “naïve”, non-experimental analysis by regressing attendance rate on whether or not a participant both (1) consented to treatment and (2) received it. In Column 2 – 4, the sample size is reduced by our removing those participants who consented but were randomized into the control condition.

Recognizing the naiveté of the model estimated in column 2, the model in column 3 includes our full set of covariates – age, gender (and binary indicators where these are missing), course type (FS vs. GCSE), and subject (maths vs. English). Finally, in Column 4 we attempt a simple propensity score match model. All of our covariates are used to predict selection into treatment in a first stage regression, and participants in the treatment group are matched with participants in the control group with the same probability of being selecting into treatment. These matched participants are then used in a second stage regression to evaluate the impact of the intervention. There are limitations to the use of this technique here. Although we have a control pool that is much larger than our treated sample, we have relatively few covariates with which to predict treatment, and so it is likely that the same technique applied with better data would yield better results.

Table 3: Estimated Effects of Study Supporter Under Different Inclusion Criteria

	(1) (RCT)	(2) (Naïve)	(3) (Naïve + Covariates)	(4) (Propensity Score Matching)
Consent	3.872*			
	(1.819)			
Treatment	4.763*	8.446***	7.037***	7.023***
	(2.269)	(1.509)	(1.515)	(1.803)
Control	48.790***	48.925***	66.876***	
	(1.072)	(0.421)	(10.018)	
College Fixed Effects	Yes	Yes	Yes	Yes
Covariates	No	No	Yes	First Stage
N	6089	5797	5797	5797

The table above shows considerable evidence of selection bias. In the first column, our experimental analysis produces an estimated treatment effect of a 4.76 percentage points increase in attendance. Participants who consented to be part of the trial, but were randomly assigned not to be treated, had 3.87 percentage points higher attendance than learners who did not consent. Due to randomisation, we argue that this 4.76 percentage points increase in attendance constitutes the true effect of the intervention.

In the naïve model estimated in column 2, we find that the estimated treatment effect is much larger than the true value, because selection is not accounted for through randomisation. In this case, the estimated effect (8.45 percentage points) is 77% larger than the true effect, and is statistically significantly greater than 4.76 percentage points, at the 5% significance level.

The addition of covariates in column 3 slightly reduces this bias, to a 47% over-estimation, and the two results are now significantly different from each other at the 10% significance level.

The use of propensity score matching, in column 4, reduces the bias only very slightly (by one hundredth of a percent of attendance). The multi-stage process used in propensity score matching inflates the standard errors, and so the estimated effect from column 4 is not significantly greater than the true effect. We note that manipulation of the functional form of the first stage model in the propensity score match can produce estimates between 4.5 % points and 9.2% points.

Discussion

In this brief paper we have demonstrated the evidence of considerable selection bias into interventions in a Further Education college setting. On the basis of our experimental evidence, we are able to establish a 'true' effect of our intervention, to which we can compare the estimated effects from three alternative models. We find, perhaps unsurprisingly, that a large portion of the effects estimated by non-experimental methods can be attributed to selection bias, and that the magnitude and direction of this effect are not strongly influenced by more sophisticated analysis.

The implications of this finding are straightforward. First, that far from being a relatively minor problem, selection bias can contribute substantively to our ability to interpret evidence. Secondly, that selection appears, at least on the basis of our limited data, to be on unobservables and hence is more difficult to correct for econometrically. Thirdly, to the extent that this issue is pervasive across other trials, and that different baseline levels of an outcome suggest differential treatment responses, it suggests that using an opt-in means of consent substantively limits to generalise findings to a wider population, compared with an opt-out consent procedure.

Practically are a few concrete considerations from our findings. Principally, when we are seeking to evaluate any intervention or policy change in this space, evaluators need to be mindful of the limitations of the methods that they are using. Although we do not dispute that improvements on our model are possible, this will require substantially richer data than we had in this study. In the context of analysing the impact of widening participation strategies in particular, strategies which involve either controlling for, or matching with, a small set of demographic variables are unlikely to capture the full set of selection into receiving an intervention, and consequently will produce biased estimates of the effectiveness of these policies. With the

Office For Fair Access (OFFA) requiring universities to conduct a specific (non-experimental) analysis of bursary impacts (OFFA, 2017), these are live debates in the sector that warrant additional discussion and methodological consideration to ensure that the impact of what are often expensive interventions is accurately measured.

References

Allcott, H. (2015). Site selection bias in program evaluation. The Quarterly Journal of Economics, 130(3), 1117-1165.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review, 104*(10), 3003-37.

Arceneaux, K., Gerber, A. S., & Green, D. P. (2006). Comparing experimental and matching methods using a large-scale voter mobilization experiment. *Political Analysis*, *14*(1), 37-62.

Athey, S., & Imbens, G. W. (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments* (Vol. 1, pp. 73-140). North-Holland.

Belot, M., & James, J. (2014). A new perspective on the issue of selection bias in randomized controlled field experiments. *Economics Letters*, *124*(3), 326-328.

Belot, M., & James, J. (2016). Partner selection into policy relevant field experiments. *Journal of Economic Behavior & Organization*, *123*, 31-56.

Christ, T. W. (2014). Scientific-Based Research and Randomized Controlled Trials, the "Gold" Standard? Alternative Paradigms and Mixed Methodologies. *Qualitative Inquiry*, *20*(1), 72–80. http://doi.org/10.1177/1077800413508523

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272-295.

Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of Economic Field Experiments* (Vol. 2, pp. 95-322). North-Holland.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.

Groot, Sanders & Rogers (2017) *I get by with a little help from my friends: Two field experiments on social support and attendance in further education colleges in the UK* (Working paper). Retrieved from http://38r8om2xjhhl25mw24492dir.wpengine.netdna-cdn.com/wp-content/uploads/2017/06/Study-Supporter-Working-Paper_2017.pdf

Hume et al. (2018). *Retention and Success in Further Education. A Report of the Behavioural Research Centre for Adult Skills and Knowledge*. The Behavioural Insights Team, London: UK.

List, J. A., & Rasul, I. (2011). Field experiments in labor economics. In *Handbook of labor economics*(Vol. 4, pp. 103-228). Elsevier.

OFFA (2017). Institutional Financial Support Coding Workbook, https://www.offa.org.uk/wp-content/uploads/2016/11/Institutional-Financial-Support-Coding-Workbook.pdf (accessed 28th January 2018)

Sadoff, S. (2014). The role of experimentation in education policy. *Oxford Review of Economic Policy, 30*(4), 597-620.

Schanzenbach, D. W. (2012). Limitations of Experiments in Education Research. *Education Finance and Policy*, *7*(2), 219–232.

The Education Endowment Foundation (2018). *Annual Report 2017*. Retrieved

fromhttps://educationendowmentfoundation.org.uk/public/files/Annual_Reports/ EEF_Annual_Report_2018.pdf

Whitehurst, G. J. (2012). The Value of Experiments in Education. *Education Finance and Policy*, *7*(2), 107–123. http://doi.org/10.1162/EDFP_a_00058

9 REFERENCES

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation, 36*(3), 320-338.

Ahmed, W., Minnaert, A., van der Werf, G., & Kuyper, H. (2010). Perceived social support and early adolescents' achievement: The mediational roles of motivational beliefs and emotions. *Journal of Youth and Adolescence*, *39*(1), 36–46.

Alcott, B. (2013). Predicting departure from British education : Identifying those most at risk through discrete time hazard modelling. *Widening Participation and Lifelong Learning*, *15*(4), 46–65.

Allan, T., Rodger, J., Dodd, M., & Cutmore, M. (2016). *Effective practice in supporting Entry / Level 1 students in post-16 institutions* (2015/17). Report prepared for the Department for Education, United Kingdom.

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *Quarterly Journal of Economics*, *130*(3), 1117–1165.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review, 104*(10), 3003-37.

Anders, J. (2017). The influence of socioeconomic status on changes in young people's expectations of applying to university. *Oxford Review Of Education*, *43*(3), 381–401.

Anders, J. D., Brown, C., Ehren, M., Greany, T., Nelson, R., Heal, J., ... & Allen, R. (2017). *Evaluation of Complex Whole-School Interventions: Methodological and Practical Considerations*. A report for The Education Endowment Foundation. Retrieved from http://discovery.ucl.ac.uk/10037777/1/EEF_CWSI_RESOURCE_FINAL_25.10.17.pdf

Anders, J., Dorsett, R., & Stokes, L. (2018). *The relative effectiveness of blended versus face-to-face adult English and maths learning Research report*. Report prepared for the Department for Education, United Kingdom.

Anderson, M. (2015). Technology Device Ownership: 2015. *Pew Research Reports*, 1–26.

Anderson, N., & Peart, S. (2016). Back on track: exploring how a further education college re-motivates learners to re-sit previously failed qualifications at GCSE. *Research in Post-Compulsory Education*, *21*(3), 196-213.

Angrist & Pischke. (2008). Mostly Harmless Econometrics : An Empiricist 's Companion. *Massachusettts Institute of Technology and The London School of Economics*, 290.

Asmussen, K. (2012). *The Evidence-Based Parenting Practitioner's Handbook*. Abingdon: Routledge.

Athey, S., & Imbens, G. W. (2017a). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3-32.

Athey, S., & Imbens, G. W. (2017b). The Econometrics of Randomized Experiments. In: *Handbook of Economic Field Experiments* (Vol. 1, pp. 73-140). North-Holland.

Austin, P. C. (2009). Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics: Simulation and Computation*, *38*(6), 1228–1234.

Austin, P. C. (2009). Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics-Simulation and Computation, 38*(6), 1228-1234.

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

Bamberger, M. (2012). Introduction to mixed methods in impact evaluation. *Impact Evaluation Notes*, *3*(3), 1-38.

Bamberger, M., Tarsilla, M., & Hesse-Biber, S. (2016). Why so many 'rigorous' evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute. *Evaluation and Program Planning*, *55*, 155–162.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., ... & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives, 31*(4), 73-102.

Barbee, A. P., Cunningham, M. R., Winstead, B. A., Derlega, V. J., Gulley, M. R., Yankeelov, P. A., & Druen, P. B. (1993). Effects of gender role expectations on the social support process. *The Journal of Social Issues*, *49*(3), 175–190.

Barnow, Burt & Greenberg, D. (2016). *What can be learned from multiple trials?* Working paper.

Barrera, M., Jr., Chassin, L., and Rogosch, F. (1993). Effects of social support and conflict on adolescent children of alcoholic and nonalcoholic fathers. *Journal of Personality and Social Psychology, 64*, 602–612.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.

BBC Radio 4 Today. (2016, September). The tekst messages getting students better grades [interview]. Retrieved from https://www.bbc.co.uk/programmes/p047xrbj

Belfield, C. R., & Levin, H. M. (2007). *The price we pay: Economic and social consequences of inadequate education*. Brookings Institution Press.

Belfield, C. R., & Levin, H. M. (2009). High School Dropouts and The Economic Losses from Juvenile Crime in California. *California Dropout Research Project Report*.

Belot, M., & James, J. (2014). A new perspective on the issue of selection bias in randomized controlled field experiments. *Economics Letters*, *124*(3), 326–328.

Belot, M., & James, J. (2016). Partner selection into policy relevant field experiments. *Journal of Economic Behavior & Organization, 123*, 31-56.

Bempechat, J., & Shernoff, D. J. (2012). Parental influences on achievement motivation and student engagement. In *Handbook of research on student engagement* (pp. 315-342). Springer, Boston, MA.

Bennett., R. (2017, October) *Good luck texts boost exam grades*. The Times & The Sunday times. Retrieved from https://www.thetimes.co.uk/article/good-luck-texts-boost-exam-grades-pvrjmn7cj

Bergman, P. (2015). *Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment*. CESifo Working Paper, No. 5391.

Bergman, P. L. (2016) *Technology Adoption in Education: Usage, Spillovers and Student Achievement,* CESifo Working Paper, No. 6101, Center forEconomic Studies and Ifo Institute (CESifo)

Bergman P., & Chan, E. W. (2017). *Leveraging Parents: The Impact of High-Frequency Information on Student Achievement*. Working paper.

Bergman, P., & Chan, E. W., (2017) *Leveraging Technology to Engage Parents at Scale: Evidence from a Randomized Controlled Trial,* CESifo Working Paper, No. 6493, Center for Economic Studies and Ifo Institute (CESifo), Munich

Bergman, P., & Rogers, T. (2017). Is This Technology Useless? How Seemingly Irrelevant Factors Affect Adoption and Efficacy. *SSRN Electronic Journal*.

Berlinski, S., Busso, M., Dinkelman, T., & Martinez, C. (2016). *Reducing parent-school information gaps and improving education outcomes: Evidence from high frequency text messaging in Chile*. Working paper, Poverty Action Lab.

Bettinger, E. P., Terry Long, B., Oreopoulos, P., & Sanbonmatsu, L. (2012). Information in College Decisions. *The Quarterly Journal of Economics*, 1205–1242.

Bettinger, E., Ludvigsen, M., Rege, M., Solli, I. (2017). Increasing Perseverance in Math: Evidence from a Field Experiment in Norway. *Journal of Economic Behavior & Organization*, *114*, 708–715.

Bird, K. A., Castleman, B. L., Goodman, J., & Lamberton, C. (2017). *Nudging at a National Scale: Experimental Evidence from a FAFSA Completion Campaign*. EdPolicyWorks Working Paper.

Bloom, H. S., 1995. Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review, 19*, 547–556.

Bodie, G. D., & Burleson, B. R. (2008). Explaining Variations in the Effects of Supportive Messages A Dual-Process Framework. *Annals of the International Communication Association*, *32*(1), 355–398.

Bolger, N., & Amarel, D. (2007). Effects of social support visibility on adjustment to stress: Experimental evidence. *Journal of personality and social psychology, 92*(3), 458.

Bolger, N., Zuckerman, A., & Kessler, R. C. (2000). Invisible support and adjustment to stress. *Journal of Personality and Social Psychology*, *79*(6), 953–961.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science and Medicine*, *75*(12), 2299–2306.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). The Logic of Event History Analysis. *Event History Modeling: A Guide for Social Scientists*, 7–20. Cambridge University Press.

Bradley, S., & Lenton, P. (2007). Dropping out of post-compulsory education in the UK: An analysis of determinants and outcomes. *Journal of Population Economics*, *20*(2), 299–328.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(May 2015), 77–101.

Brocas, I., Carrillo, J. D., & Dewatripont, M. (2004). Commitment Devices under Self-Control Problems : An Overview. *The Psychology of Economics Decisions*, (1956), 49–65.

Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment Devices. *Annual*

Review of Economics, *2*(1), 671–698.

Burleson, B. R. (2003). Emotional support skills. *Handbook of communication and social interaction skills,* 551-594.

Burleson, B. R. (2009). Understanding the outcomes of supportive communication: A dual-process approach. *Journal of Social and Personal Relationships*, *26*(1), 21–38.

Burleson, B. R., & Goldsmith, D. J. (1996). Chapter 9 – How the comforting process works: Alleviating emotional distress through conversationally induced reappraisals. *Handbook of Communication and Emotion*, 245–280.

Burleson, B. R., & MacGeorge, E. L. (2002). Supportive communication. *Handbook of interpersonal communication, 3,* 374-424.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Cabus, S. J., & De Witte, K. (2012). *School Absenteeism and Dropout, A Bayesian Duration Model.* TIER working paper series.

Campbell, Donald, T. (1969). Reform as Experiments. *American Psychologist*, *24*, 409–429.

Campbell, Evans, M., Tucker, M., Quilty, B., Dieppe, P., & Donovan, J. L. (2001). Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. *Journal of Epidemiology and Community Health*, *55*(2), 132–138.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ: British Medical …*, *321*.

Campbell, M. J., Julious, S. A., & Altman, D. G. (1995). Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ*, *311*(7013), 1145.

Campbell, R., Starkey, F., Holliday, J., Audrey, S., Bloor, M., Parry-Langdon, N., … Moore, L. (2008). An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. *The Lancet*, *371*(9624), 1595–1602.

Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College Student Responses to Web and Paper Surveys: Does Mode Matter? *Research in Higher Education*, *44*(1), 1–19.

Carlson, L. E., Goodey, E., Bennett, M. H., Taenzer, P., & Koopmans, J. (2002). The addition of social support to a community-based large-group behavioral smoking cessation intervention: improved cessation rates and gender differences. *Addictive behaviors, 27*(4), 547-559.

Cartwright, N. (2009). What are randomised controlled trials good for? *Philosophical Studies*, *147*(1), 59–70.

Castleman, B. L., & Meyer, K. (2016). Can Text Message Nudges Improve Academic Outcomes in College? Evidence from a West Virginia Initiative. *EdPolicyWorks*, (43).

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior and Organization*, *115*(April), 144–160.

Castleman, B. L., & Page, L. C. (2016). Freshman Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence. *Journal of Human Resources*, *51*(2), 389–415.

Castleman, B. L., & Page, L. C. (2017). Parental influences on postsecondary decision making: Evidence from a text messaging experiment. *Educational Evaluation and Policy Analysis, 39*(2), 361-377.

Chande, R. (2016). *Essays on Education*. Doctoral dissertation.

Chande, R., Luca, M., Sanders, M., Soon, X., & Borcan, O. (2017). *Increasing attendance and attainment among adult students in the UK : Evidence from a field experiment*. Working paper, The Behavioural Insights Team.

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research*,

Cobb, S. (1976). Social support as a moderator of life stress. *Psychosomatic Medicine*, *38*(5), 300–314.

Coffman, D., & Gilligan, T. (2002). Social Support, Stress, and Self-Efficacy: Effects on Students' Satisfaction. *Journal of College Student Retention 4*(1), 53–66.

Coffman, L. C., & Niederle, M. (2014). *Pre-analysis plans are not the solution, replications might be*. Working paper.

Cohen, S. (1988). Psychosocial models of the role of social support in the etiology of physical disease. *Health psychology, 7*(3), 269.

Cohen, S., Underwood, L.G., & Gottlieb, B.H. (2000). *Social Relationships and Health*. In S. Cohen, Underwood, L.G. & Gottlieb, B.H. (Ed.), Social Support Measurement and Intervention: A guide for health and social scientists. New York: Oxford University Press.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310–357.

Colella, T. J., & King-Shier, K. (2017). The effect of a peer support intervention on early recovery outcomes in men recovering from coronary bypass surgery: A randomized controlled trial. *European Journal of Cardiovascular Nursing*.

Collins, N. L., & Feeney, B. C. (2000). A safe haven: An attachment theory perspective on support seeking and caregiving in intimate relationships. *Journal of Personality and Social Psychology*, *78*(6), 1053–1073.

Collins, N. L., & Feeney, B. C. (2004). Working models of attachment shape perceptions of social support: Evidence from experimental and observational studies. *Journal of Personality and Social Psychology*, *87*(3), 363–383.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class Attendance in College: A Meta-Analytic Review of the Relationship of Class Attendance With Grades and Student Characteristics. *Review of Educational Research*, *80*(2), 272–295.

Creswell, J. W. (2009). Mixed methods Research Design. *SAGE Publications*.

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Third edition. Washington DC: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). Choosing a mixed methods design. *Designing and conducting mixed methods research,* (pp. 53-106). Thousand Oaks, CA: Sage.

Cunha, N., Lichand, G., Madeira, R., & Bettinger, E. (2017). *What Is It About Communicating With Parents?* Working Paper.

Cutrona, C. E. (1991). Stress and Social Support- in Search of Optimal Matching. *Jurnal of Social and Clinical Psychology*, *9*(J), 3–15.

Cutrona, C. E., Cohen, B. B., & Igram, S. (1990). Contextual determinants of the perceived supportiveness of helping behaviours. *Journal of Social and Personal Relationships, 7*(4), 553-562.

Cutrona, C. E., & Russell, D. W. (1990). Type of social support and specific stress: Toward a theory of optimal matching. In B. R. Sarason, I. G. Sarason, & G. R. Pierce (Eds.), *Wiley series on personality processes. Social support: An interactional view* (p. 319-366). Oxford, England: John Wiley & Sons.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, *4*(1), 1–15.

Damschroder, L. J., Aron, D., Keith, R., Kirsh, S., Alexander, J., & Lowery,

J. (2013). Evaluation of large-scale weight management program using the consolidated framework for implementation research. *Implementation Science*, *8*(1), 51.

Damschroder, L., Peikes, D., & Petersen, D. (2013). *Using implementation research to guide adaptation, implementation, and dissemination of patient-centered medical home models*. Mathematica Policy Research.

Day, A. L., & Livingstone, H. A. (2003). Gender differences in perceptions of stressors and utilization of social support among university students. *Canadian Journal of Behavioural Science*, *35*(2), 73–83.

De Boer, M. R., Waterlander, W. E., Kuijper, L. D. J., Steenhuis, I. H. M., & Twisk, J. W. R. (2015). Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 1–8.

Dee, T. S. (2004). Are there civic returns to education?. *Journal of Public Economics, 88*(9-10), 1697-1720.

De Silva, M. J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Patel, V. (2014). Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials*, *15*(1), 267.

De Witte, K., Cabus, S., Thyssen, G., Groot, W., & Van Den Brink, H. M. (2013). A critical review of the literature on school dropout. *Educational Research Review*, *10*, 13–28.

Deaton, A., & Cartwright, N. (2017). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*.

Deci, E. L., & Ryan, R. M. (2000). The " What " and " Why " of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268.

Department for Business Innovation and Skills. (2015). *Measuring the Net Present Value of Further Education in England*. BIS Research Paper.

Department for Education (2012). *Collection: Raising the participation age*. Retrieved from https://www.gov.uk/government/collections/raising-the-participation-age

Department for Education. (2016a). *The link between absence and attainment at KS2 and KS4, 2013/14*. Research report. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/509679/The-link-between-absence-and-attainment-at-KS2-and-KS4-2013-to-2014-academic-year.pdf

Department for Education. (2016b). Main text: SFR15/2016. *Level 1 and 2 attainment in English and maths by students aged 16-18: academic* *year 2014/15*. Retrieved from https://www.gov.uk/government/statistics/level-1-and-2-english-and-maths-16-to-18-students-2014-to-2015

Department for Education. (2017a). *2015 to 2016 education and training NART overall GCSE pass rate and achievement rate*. [Data set]. Retrieved from https://www.gov.uk/government/statistics/national-achievement-rates-tables-2015-to-2016

Department for Education. (2017b). *GCSE and equivalent results: 2016 to 2017 (provisional). Main text: SFR57/2017*. Retrieved from https://www.gov.uk/government/statistics/gcse-and-equivalent-results-2016-to-2017-provisional

Department for Education (2018). *Retention measure, completion and attainment measure and tech level minimum standards tables: SFR03/2018.* [Data set]. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/689932/SFR03_2018_Retention_CA_Tech_level_minimum_Standards_tables.xlsx

Desforges, C., & Abouchaar, A. (2003). The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review (Vol. 433). Nottingham: DfES publications.

Donovan, J., Mills, N., Smith, M., Brindle, L., Jacoby, A., Peters, T., ... Little, P. (2002). Improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *Bmj*, *325*(7367), 766.

Drabble, S. J., & O'Cathain, A. (2015). Moving from randomized controlled trials to mixed methods intervention evaluations. In S. Hesse-Biber & R. B. Johnson (Eds.), Oxford library of psychology. *The Oxford handbook of multimethod and mixed methods research inquiry* (p. 406-425). New York: Oxford University Press.

DuBois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., & Valentine, J. C. (2011). How Effective Are Mentoring Programs for Youth? A Systematic Assessment of the Evidence. *Psychological Science in the Public Interest*, *12*(2), 57–91.

Duflo, E., Glennerster, R., & Kremer, M. (2008). Using Randomization in Development Economics: A Toolkit. *Handbook of Development Economics*, *53*(6059), 68.

Durlak, J. A. (2009). How to select, calculate, and interperet effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928.

Eastwick, P. W., Finkel, E. J., Krishnamurti, T., & Loewenstein, G. (2008). Mispredicting distress following romantic breakup: Revealing the time course of the affective forecasting error. *Journal of Experimental*

Social Psychology, *44*(3), 800–807.

Eccles, J. S. (2007). *Families, schools, and developing achievement-related motivations and engagement*. In J. E. Grusec & P. D. Hastings (Eds.), Handbook of socialization (pp. 665–691). New York, NY: The Guilford Press.

Education Endowment Foundation (2012). *Annual Report 2011-2012*. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Annual_Reports/EEF_AnnualReport_2011-12.pdf

Education Endowment Foundation (2018). *Maths-for-Life. Project Summary*. Retrieved from https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maths-for-life/

Education and Skills Funding Agency (2017). *National achievement rates tables 2015 to 2016*. [Data set]. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/618898/Education_and_Training_Overall_Headline_Achievement_Rates_Tables_201516.xlsx11

Education and Skills Funding Agency (2018). *Guidance. 16 to 19 funding: maths and English condition of funding*. Retrieved from https://www.gov.uk/guidance/16-to-19-funding-maths-and-english-condition-of-funding

Eisenberger, N. I., Taylor, S. E., Gable, S. L., Hilmert, C. J., & Lieberman, M. D. (2009). Neural pathways link social support to attenuated neuroendocrine stress responses, *Neuroimage, 35(*4), 1601–1612.

Enzmann, D. (2015). *Notes on Effect Size Measures for the Difference of Means From Two Independent Groups: The Case of Cohen'sd and Hedges' g*. Working paper, 1–5.

European Commission. (2017). *Education and Training Monitor 2017 - United Kingdom.* Luxembourg: Publications Office of the European Union, 2017. Retrieved from https://ec.europa.eu/education/sites/education/files/monitor2017-uk_en.pdf

Exley, S., & Martin, W. (2017, March). EEF funds major trials to help students 'make the grade' in GCSE resits. Times Educational Supplement. Retrieved from https://www.tes.com/news/eef-funds-major-trials-help-learners-make-grade-gcse-resits

Fan, X., & Chen, M. (2001). Parental Involvement and Studnet's Academic Achievement: A Meta-Analysis. *Educational Psychology Review*, *13*(1), 1–22.

Feeney, B. C. (2004). A secure base: Responsive support of goal strivings and exploration in adult intimate relationships. *Journal of Personality* *and Social Psychology*, *87*(5), 631–648.

Feeney, B. C., & Collins, N. L. (2015). A new look at social support: A theoretical perspective on thriving through relationships. *Personality and Social Psychology Review, 19*(2), 113-147.

Fereday, J., & Muir-Cochrane, E. (2008). Demonstrating Rigor using thematic analysis: A hybrid approach of inductive and deductive coding. *International Journal of Qualitative Methods*, *5*(1), 80–92.

Figlio, D., Karbownik, K., & Salvanes, K. G. (2016). *Education research and administrative data*. In Handbook of the Economics of Education (Vol. 5, pp. 75-138). Elsevier.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117–142.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). I*mplementation Research: A Synthesis of the Literature. Tampa,* FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Fjeldsoe, B. S., Marshall, A. L., & Miller, Y. D. (2009). Behavior Change Interventions Delivered by Mobile Telephone Short-Message Service. *American Journal of Preventive Medicine*, *36*(2), 165–173.

Forman, J., & Damschroder, L. (2007). Qualitative Content Analysis. *Advances in Bioethics*, *11*(07), 39–62.

Frey, E., & Rogers, T. (2014). Persistence: How Treatment Effects Persist After Interventions Stop. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 172–179.

Froiland, J. M. (2011). Parental autonomy support and student learning goals: A preliminary examination of an intrinsic motivation intervention. *Child & youth care forum, 40*(2), p. 135-149).

Frontier Economics. (2017). *Further Education Workforce Data For England. Analysis of the 2015-2016 Staff Individualised Record (SIR) data*. Report prepared for the Education Training Foundation. Retrieved from http://www.et-foundation.co.uk/wp-content/uploads/2017/06/FE-workforce-data-2017-FINAL.pdf

Furman, W., & Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child development, 63*(1), 103-115.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148–162.

Gegenfurtner, A., & Vauras, M. (2012). Age-related differences in the relation between motivation to learn and transfer of training in adult continuing education. *Contemporary Educational Psychology*, *37*(1), 33–46.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. New York: WW Norton.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638.

Giordano, P. C. (2003). Relationships in Adolescence. *Annual Review of Sociology*, *29*(1), 257–281.

Givertz, M., & Segrin, C. (2014). The Association Between Overinvolved Parenting and Young Adults' Self-Efficacy, Psychological Entitlement, and Family Communication. *Communication Research*, *4*(8), 1111–1136.

Gleason, M. E., Shrout, P. E., & Bolger, N. (2008). Receiving support as a mixed blessing: evidence for dual effects of support on psychological outcomes. *Journal of Personality and Social Psychology*, *94*(5), 824–838.

Goodenow, C., & Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *The Journal of Experimental Education, 62*(1), 60-71.

Gottlieb, B. H., & Bergen, A. E. (2010). Social support concepts and measures. *Journal of Psychosomatic Research*, *69*(5), 511–520.

Green, M. (2017, November). Global education stakes rise for a system threatened by skills crisis. The Financial Times. Retrieved from https://www.ft.com/content/3df1d370-a9c7-11e7-ab66-21cc87a2edde

Greene, J. O & Burleson, B. R. (2003). *Handbook of communication and social interaction skills*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Groot, A. S. B., Sanders, M. T., & Rogers, T. (2017). *I get by with a little help from my friends: Two field experiments on social support and attendance in further education colleges in the UK*. Working paper. London: The Behavioural Insights Team.

Guay, F., Ratelle, C. F., & Chanal, J. (2008). Optimal learning in optimal contexts: The role of self-determination in education. *Canadian Psychology/Psychologie Canadienne*, *49*(3), 233.

Guryan, J., Kim, J. S., & Park, K. H. (2016). Motivation and incentives in education: Evidence from a summer reading experiment. *Economics of Education Review, 55*, 1-20.

Hale, C. J., Hannum, J. W., & Espelage, D. L. (2005). Social Support and Physical Health: The Importance of Belonging. *Journal of American*

College Health, *53*(6), 276–284.

Hall, A. K., Cole-Lewis H., B. (2015). Mobile Text Messaging for Health: A Systematic Review of Reviews. *Annu Rev Public Health*, *36*(20), 393–415.

Hanley, P., Chambers, B., & Haslam, J. (2016). Reassessing RCTs as the 'gold standard': synergy not separatism in evaluation designs. *International Journal of Research & Method in Education*, *39*(3), 287–298.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review, 73*, 103-130.

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving Student Outcomes in Higher Education: The Science of Targeted Intervention. *Annual Review of Psychology*, *69*(1).

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature, 42*(4), 1009-1055.

Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: how 'out of control' can a randomised controlled trial be? *BMJ (Clinical Research Ed.)*, *328*(7455), 1561–1563.

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality and Quantity*, *43*(1), 59–74.

Helsen, M., Vollebergh, W., & Meeus, W. (2000). Social support from parents and friends and emotional problems in adolescence. *Journal of youth and adolescence, 29*(3), 319-335.

Hesse-Biber, S. (2012). Weaving a Multimethodology and Mixed Methods Praxis Into Randomized Control Trials to Enhance Credibility. *Qualitative Inquiry*, *18*, 876–889.

Hesse-Biber, S. (2013). Thinking Outside the Randomized Controlled Trials Experimental Box: Strategies for Enhancing Credibility and Social Justice. *Mixed Methods and Credibility of Evidence in Evaluation. New Directions for Evaluation*, *138*, 49–60.

Hicks, A. M., & Diamond, L. M. (2008). How was your day? Couples' affect when telling and hearing daily events. *Personal Relationships*, *15*(2), 205–228.

Higgins, S., & Katsipataki, M. (2016). Communicating comparative findings from meta-analysis in educational research: some examples and suggestions. *International Journal of Research and Method in Education*, *39*(3), 237–254.

Higton, J., Archer, R., Dalby, D., Robinson, S., Birkin, G., Stutz, A., & Duckworth, V. (2017). *Effective practice in the delivery and teaching of English and Mathematics to 16-18 year olds*. Department for

Education.

HM Treasury. (2018). *The Green Book. Central Government Guidance On Appraisal And Evaluation.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf

Hogan, B. E., Linden, W., & Najarian, B. (2002). Social support interventions: Do they work?. *Clinical psychology review, 22*(3), 381-440.

Hohwü, L., Lyshol, H., Gissler, M., Jonsson, S. H., Petzold, M., & Obel, C. (2013). Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. *Journal of medical Internet research*, *15*(8).

Holland, P. W. (1986). Statistics and causal inference. Journal of the *American statistical Association, 81*(396), 945-960.

Hooker, E. D., Campos, B., & Pressman, S. D. (2018). It just takes a text: Partner text messages can reduce cardiovascular responses to stress in females. *Computers in Human Behavior, 84*, 485-492.

Hoover-Dempsey, K. V., Walker, J. M., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., & Closson, K. (2005). Why do parents become involved? Research findings and implications. *The elementary school journal*, *106*(2), 105-130.

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science, 241*(4865), 540-545.

Howland, M., & Simpson, J. A. (2010). Getting in under the radar: A dyadic view of invisible support. *Psychological Science*, *21*(12), 1878–1885.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research, 15*(9), 1277-1288.

Hulleman, C. S & Barron, K. E. (2015). Motivation interventions in education: Bridging theory , research , and practice. *Handbook of educational psychology*.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–2.

Hume, S., O'Reilly, F., Groot, B., Chande, R., Sanders, M., Hollingsworth, A., Ter Meer, J., Barnes, J., Dodd, S., & Soon, Z. (2018a). *Improving engagement with maths and English courses: Insights from Behavioural Research*. Research report, Department for Education UK.

Hume, S., O'Reilly, F., Groot, B., Selley, E., Barnes, J., Soon, Z.,Chande, R., Sanders, M., (2018b). *Retention and Success in Further Education*. Research Report, The Behavioural Insights Team.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in educational settings: An introductory handbook. *Education Endowment Foundation*, 1–32.

Hupkau, C., Mcnally, S., Ruiz-Valenzuela, J., & Ventura, G. (2016). Post-Compulsory Education in England: Choices and Implications, *20*(240), 42–57.

Hyde, J. S., Canning, E. A., Rozek, C. S., Clarke, E., Hulleman, C. S., & Harackiewicz, J. M. (2017). The role of mothers' communication in promoting motivation for math and science course-taking in high school. *Journal of Research on Adolescence, 27*(1), 49-64.

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics, 98*(4), 701-712.

Imbens, G. W., & Rubin, D. B. (2015). Assessing Overlap in Covariate Distributions. *Causal Inference for Statistics, Social, and Biomedical Sciences An Introduction*, 309–336.

Impetus- The Private Equity Foundation. (2017). *Life after school: confronting the crisis*. Retrieved from https://impetus-pef.org.uk/assets/publications/Report/2017-03-14_Impetus-PEF-Confronting-Crisis.pdf

Jacobson, K., & Crockett, L. (2000). Parental monitoring and adolescent adjustment: An ecological perspective. *Journal of Research on Adolescence*, *10*(1), 65–97.

Jamal, F., Fletcher, A., Shackleton, N., Elbourne, D., Viner, R., & Bonell, C. (2015). The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. *Trials*, *16*(1), 466.

Jerrim, J. (2018). *A digital divide? Randomised evidence on the impact of computer-based assessment in PISA. Working* paper, Centre for Education Economics.

Jeynes. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement. *Urban Education*, *42*(1), 82–110.

Jeynes, W. H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student Academic Achievement. *Urban Education*, *40*(3), 237–269.

Kalil, A. (2015). *Inequality begins at home: The role of parenting in the diverging destinies of rich and poor children*. Families in an era of increasing inequality (p. 63-82). Springer: Cambridge.

Kaskutas, L. A. (2009). Alcoholics Anonymous effectiveness: Faith meets science. *Journal of addictive diseases, 28*(2), 145-157.

Kaul, M., & Lakey, B. (2003). Where is the Support in Perceived Support? The Role of Generic Relationship Satisfaction and Enacted Support in Perceived Support's Relation to Low Distress. *Journal of Social and Clinical Psychology*, *22*(1), 59–78.

Kirk, M. A., Kelley, C., Yankey, N., Birken, S. A., Abadie, B., & Damschroder, L. (2016). A systematic review of the use of the Consolidated Framework for Implementation Research. *Implementation Science*, *11*(1).

Klein, R. A., Ratliff, K. a., Vianello, M., Adams ., R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152.

Kolenovic, Z., Linderman, D., & Karp, M. M. (2013). Improving Student Outcomes via Comprehensive Supports: Three-Year Outcomes From CUNY's Accelerated Study in Associate Programs (ASAP). *Community College Review*, *41*(4), 271–291.

Kraft, M. A., & Dougherty, S. M. (2013). The Effect of Teacher-Family Communication on Student Engagement: Evidence From a Randomized Field Experiment. *Journal of Research on Educational Effectiveness*, *6*(3), 199–222.

Kraft, M. A., & Monti-Nussbaum, M. (2017). Can Schools Empower Parents to Prevent Summer Learning Loss? A Text Messaging Field Experiment to Promote Literacy Skills. *The ANNALS of the American Academy of Political and Social Science*., 1–50.

Kraft, M. A., & Rogers, T. (2015). The Underutilized Potential of Teacher-to-Parent Communication: Evidence from a Field Experiment. *Economics of Education Review*, *47*, 49–63.

Kuha, J., & Mills, C. (2018). On Group Comparisons With Logistic Regression Models. *Sociological Methods & Research*, 004912411774730.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, *112*(2), 443–478.

Lakey, B., & Cohen, S. (2000). Social support theory and measurement. *Social Support Measurement and Intervention: A Guide for Health and Social Scientists*.

Lakey, B., & Orehek, E. (2011). Relational regulation theory: a new approach to explain the link between perceived social support and mental health. *Psychological Review*, *118*(May 2011), 482–495.

Larson, M., Cook, C. R., Fiat, A., & Lyon, A. R. (2018). Stressed Teachers Don't Make Good Implementers: Examining the Interplay Between Stress Reduction and Intervention Fidelity. *School Mental Health*.

Lewin, S., Glenton, C., & Oxman, A. D. (2009). Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *BMJ, 339,* b3496.

Lewis, J., & Ritchie, J. (2003). *Qualitative research practice: A guide for social science students and researchers*. New York: Sage.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry* (Vol. 75). Sage.

List, J. A., & Rasul, I. (2010). *Field Experiments in Labor Economics*. National Bureau of Economic Research (Cambridge, MA) Working Paper No. 16062.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, *355*(6325), 584–585. 8

Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *American Statistician*, *54*(3), 217–224.

Lyon, A. R., Cook, C. R., Brown, E. C., Locke, J., Davis, C., Ehrhart, M., & Aarons, G. A. (2018). Assessing organizational implementation context in the education sector: Confirmatory factor analysis of measures of implementation leadership, climate, and citizenship. *Implementation Science*, *13*(1), 1–14.

Machin, S., McNally, S., & Ruiz-Valenzuela, J. (2018). *Entry Through the Narrow Door : The Costs of Just Failing High Stakes Exams*. Working paper, CVER series.

Mackinnon, J. G., & White, H. (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators With Improved Finite Sample Properties*. *Journal of Econometrics*, *29*, 305–325.

Maisel, N. C., & Gable, S. L. (2009). The paradox of received social support: The importance of responsiveness. *Psychological Science*, *20*(8), 928–932.

Malecki, C. K., & Demaray, M. K. (2003). What Type of Support Do They Need? Investigating Student Adjustment as Related to Emotional, Informational, Appraisal, and Instrumental Support. *School Psychology Quarterly*, *18*(3), 231–252.

Malecki, C. K., & Demaray, M. K. (2006). Social Support as a Buffer in the Relationship between Socioeconomic Status and Academic Performance. *School Psychology Quarterly*, *21*(4), 375–395.

Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G., & Pawson, R. (2013). Realist RCTs of complex interventions - An oxymoron. *Social Science and Medicine*, *94*, 124–128.

Mason, J. (2002). Linking qualitative and quantitative data analysis. In *Analyzing qualitative data* (pp. 103-124). Routledge.

Maxwell, J. A. (2009). Designing a Qualitative Study. *The SAGE Handbook of Applied Social Research Methods*, 214–253.

May, S., & West, R. (2000). Do social support interventions (" buddy systems ") aid smoking cessation? A review. *Tobacco Control*, *9*, 415–422.

Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2015). *Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together (PACT) Intervention*. NBER Working Paper.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC Medical Research Methodology*, *7*(1), 30.

McConnell, B., & Vera-Hérnandez, M. (2015). *Going beyond simple sample size calculations: a practitioner's guide*. Institute for Fiscal Studies, IFS Working Paper W15/17.

McIntosh, S., & Vignoles, A. (2001). Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers, 53*(3), 453-481.

McNally, S., & Wyness, G. (2017). *Education and Skills: The UK Policy Agenda*. CEP Performance Paper No'CEPEA041, Centre for Economic Performance, London School of Economics.

Meeus, W., Oosterwegel, A., & Vollebergh, W. (2002). Parental and peer attachment and identity development in adolescence. *Journal of Adolescence*, *25*(1), 93–106.

Merrett, F. (2006). Reflections on the Hawthorne effect. *Educational Psychology, 26*(1), 143-146.

Michie, S., Johnston, M., Francis, J., Hardeman, W., & Eccles, M. (2008). From Theory to Intervention: Mapping Theoretically Derived Behavioural Determinants to Behaviour Change Techniques. *Applied Psychology*, *57*(4), 660–680.

Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making. *Perspectives on Psychological Science*, *3*(4), 324–338.

Miller, S., Davison, J., Yohanis, J., Sloan, S., Gildea, A., & Thurston, A. (2017). Texting Parents: Evaluation Report and Executive Summary. *Education Endowment Foundation.*

Mood, C. (2010). Logistic regression: Why we cannot do what We think we can do, and what we can do about it. *European Sociological Review*, *26*(1), 67–82.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., ...

Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *Bmj*, *350*, h1258–h1258. https://doi.org/10.1136/bmj.h1258

Morris, S. P., Edovald, T., Lloyd, C., & Kiss, Z. (2016). The importance of specifying and studying causal mechanisms in school-based randomised controlled trials: lessons from two studies of cross-age peer tutoring. *Educational Research and Evaluation*, *22*(7–8), 422–439.

Morrison, N. (2017, October). *The Little Nudge That Makes A Big Difference To Student Grades*. Forbes. Retrieved from https://www.forbes.com/sites/nickmorrison/2017/10/18/the-little-nudge-that-makes-a-big-difference-to-student-grades/

Nelson, R. M., & DeBacker, T. K. (2008). Achievement Motivation in Adolescents: The Role of Peer Climate and Best Friends. *The Journal of Experimental Education*, *76*(2), 170–189.

Newman-Ford, L., Fitzgibbon, K., Lloyd, S., & Thomas, S. (2008). A large-scale investigation into the relationship between attendance and attainment: a study using an innovative, electronic attendance monitoring system. *Studies in Higher Education, 33*(6), 699-717.

Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal*, *332*(7538), 413–416.

Ofcom (2017). Facts and figures. Retrieved from http://media.ofcom.org.uk/facts/

Oreopoulos, P., Brown, R. S., & Lavecchia, A. M. (2017). Pathways to Education: An Integrated Approach to Helping At-Risk High School Students. *Journal of Political Economy*, *125*(4), 947–984.

Oreopoulos, P., & Petronijevic, U. (2017). Student coaching: How far can technology go? *Journal of Human Resources, 1216-8439R*.

Organisation for Economic Co-operation and Development. (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD Publishing.

Osam, E. K., Bergman, M., & Cumberland, D. M. (2017). An integrative literature review on the barriers impacting adult learners' return to college. *Adult Learning*, *28*(2), 54–60.

Pakter, A., & Chen, L. L. (2013). The Daily Text: Increasing Parental Involvement in Education with Mobile Text Messaging. *Journal of Educational Technology Systems*, *41*(4), 353–367.

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental*

Health Services Research, *38*(1), 44–53.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm. Policy Ment. Health*, *42*(5), 533–544.

Paluck, E. L. (2010). The Promising Integration of Qualitative Methods and Field Experiments. *The ANNALS of the American Academy of Political and Social Science*, *628*(March), 59–71.

Pan, J., Zaff, J. F., & Donlan, A. E. (2017). Social Support and Academic Engagement Among Reconnected Youth: Adverse Life Experiences as a Moderator. *Journal of Research on Adolescence*, 1–17.

Patrick, K., Raab, F., Adams, M. A., Dillon, L., Zabinski, M., Rock, C. L., ... & Norman, G. J. (2009). A text message–based intervention for weight loss: randomized controlled trial. *Journal of medical Internet research, 11*(1).

Pawson, R. (2004). Simple Principles for The Evaluation of Complex Programmes. *An Evidence-Based Approach to Public Health and Tackling Health Inequalities: Practical Steps and Methodological Challenges*, 1–36.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: an introduction*. ESRC Research Methods Programme, Working Paper Series.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146.

Peters, J., Langbein, J., & Roberts, G. (2016). Policy evaluation, randomized controlled trials, and external validity—A systematic review. *Economics Letters, 147*, 51-54.

Piko, B. F., Luszczynska, A., & Fitzpatrick, K. M. (2013). Social inequalities in adolescent depression: the role of parental social support and optimism. *The International Journal of Social Psychiatry*, *59*(5), 474–81.

Plano Clark, V. L., Schumacher, K., West, C., Edrington, J., Dunn, L. B., Harzstark, A., ... Miaskowski, C. (2013). Practices for Embedding an Interpretive Qualitative Approach Within a Randomized Clinical Trial. *Journal of Mixed Methods Research*, *7*(3), 219–242.

Porter, N. (2015). CROSSING THE LINE: Improving Success Rates Among Students Retaking English And Maths GCSEs. *A Policy Exchange Policy Bite*. Retrieved from https://policyexchange.org.uk/wp-content/uploads/2016/09/crossing-the-line.pdf

Pomerantz, E. M., Moorman, E. A., & Litwack, S. D. (2007). The How, Whom, and Why of Parents' Involvement in Children's Academic

Lives: More Is Not Always Better. *Review of Educational Research*, *77*(3), 373–410.

R. Burke Johnson and Anthony J. Onwuegbuzie, Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(14), 14–26.

Rafaeli, E., & Gleason, M. E. J. (2009). Skilled Support Within Intimate Relationships. *Journal of Family Theory & Review*, *1*(1), 20–37.

Ranney, M. L., Pittman, S. K., Riese, A., Ybarra, M., Huang, J., Spirito, A., & Rosen, R. (2018). Mobile Health Intervention Development Principles: Lessons from an Adolescent Cyberbullying Intervention. *Proceedings of the 51st Hawaii International Conference on System Sciences.*

Ravallion, M. (2012). Why Don' t We See Poverty Convergence ? *American Economic Review*, *102*(1), 504–523.

Reeve, J. (2012). *A self-determination theory perspective on student engagement*. In Handbook of research on student engagement (pp. 149-172). Springer, Boston, MA.

Richburg-Hayes, L., Anzelone, C., Dechausay, N., Datta, S., Fiorillo, A., Potok, L., ... & Balz, J. (2014). *Behavioral economics and social policy: Designing innovative solutions for programs supported by the Administration for Children and Families*. OPRE Report No. 2014 16a

Richman, J. M., Rosenfeld, L. B., & Bowen, G. L. (1998). Social support for adolescents at risk of school failure. *The Social Worker*, *43*(4), 309–323.

Robinson, C., & Lee, M. G. L. (2017). Reducing Student Absenteeism in the Early Grades by Targeting Parental Beliefs. *SSRN Electronic Journal*.

Rogers, T., Milkman, K. L., & Volpp, K. G. (2014). Commitment devices: Using initiatives to change behavior. *Jama*, *69*(9), 525–527.

Rogers, T., Duncan, T., Wolford, T., Ternovski, J., Subramanyam, S., & Reitano, A. (2017). 'A Randomized Experiment Using Absenteeism Information to "Nudge" Attendance'. *Education Week*, *36*(21), 5.

Rogers, T., & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour, 2*(5), 335.

Rogers, T., & Frey, E. (2015). Changing Behavior Beyond the Here and Now. *The Wiley Blackwell Handbook of Judgment and Decision Making*, 723–748.

Rogers, T., Duncan, T., Wolford, T., Ternovski, J., Subramanyam, S., & Reitano, A. (2017). A Randomized Experiment Using Absenteeism Information to" Nudge" Attendance. REL 2017-252. *Regional Educational Laboratory Mid-Atlantic.*

Rosenfeld, L. B., Richman, J. M., & Bowen, G. L. (1998). Low social support among at-risk adolescents. *Social Work in Education*, *20*(4), 245–260.

Rosenfeld, L. B., Richman, J. M., & Bowen, G. L. (2000). Social Support Networks and School Outcomes: The Centrality of the Teacher. *Child & Adolescent Social Work Journal*, *17*(3), 205–226.

Rotheram-Borus, M. J., Tomlinson, M., Gwegwe, M., Comulada, W. S., Kaufman, N., & Keim, M. (2012). Diabetes buddies: peer support through a mobile phone buddy system. *The Diabetes Educator, 38*(3), 357-365.

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association*, *100*(469), 322–331.

Rueger, S. Y., Malecki, C. K., & Demaray, M. K. (2010). Relationship Between Multiple Sources of Perceived Social Support and Psychological and Academic Adjustment in Early Adolescence: Comparisons Across Gender. *Journal of Youth and Adolescence*, *39*(1), 47–61.

Rueger, S. Y., Malecki, C. K., Pyun, Y., Aycock, C., & Coyle, S. (2016). A meta-analytic review of the association between perceived social support and depression in childhood and adolescence. *Psychological Bulletin*, *142*(10), 1017–1067.

Rumberger, R., Addis, H., Allensworth, E., Balfanz, R., Bruch, J., Dillon, E., ... & Newman-Gonchar, R. (2017). Preventing Dropout in Secondary Schools. *Washington, DC.*

Rumberger, R. W., & Lim, S. A. (2008). Why students drop out of school: A review of 25 years of research. *California Dropout Research Project, 15, 1-3*.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction*, *42*, 95–103.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78.

Sandelowski, M. (1996). Focus on qualitative methods: Using qualitative methods in intervention studies. *Research in Nursing & Health*, *19*(4), 359–364.

Sanders, M. T., & Groot, A. S. B., (*forthcoming*). How important is selection bias in educational research? Evidence from a field experiment. *Widening Participation and Lifelong Learning*.

Sanderson, I. (2000). Evaluation in Complex Policy Systems. *Policy Research Institute*, *6*(4), 433–454.

Schanzenbach, D. W. (2012). Limitations of Experiments in Education Research. *Education Finance and Policy*, *7*(2), 219–232.

Schochet, P. Z., & Chiang, H. (2009). Technical Methods Report: Estimation and Identification of the Complier Average Causal Effect Parameter in Education RCTs. *NCEE 2009-4040. National Center for Education Evaluation and Regional Assistance*.

Schulz, R. (1976). Effects of control and predictability on the physical and psychological well-being of the institutionalized aged. *Journal of Personality and Social Psychology*, *33*(5), 563–573.

Schulz, R., & Hanusa, B. H. (1978). Long-term effects of control and predictability-enhancing interventions: Findings and ethical issues. *Journal of Personality and Social Psychology*, *36*(11), 1194–1201.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine*, *8*(1), 18.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental for Generalized Designs Causal Inference. *Handbook of Industrial and Organizational Psychology*, *223*, 623.

Shrout, P. E., Herman, C. M., & Bolger, N. (2006). The costs and benefits of practical and emotional support on adjustment: A daily diary study of couples experiencing acute stress. *Personal Relationships*, *13*(1), 115–134.

Simoni, J. M., Franks, J. C., Lehavot, K., & Yard, S. S. (2011). Peer interventions to promote health: Conceptual considerations. *American Journal of Orthopsychiatry*, *81*(3), 351–359.

Sinclair, B., Mcconnell, M., & Green, D. P. (2012). Detecting Spillover Effects: Design and Analysis of Multilevel Experiments. *American Journal of Political Science*, *56*(4), 1055–1069.

Sirin, S. R., Gupta, T., Ryce, P., Katsia, D., Suárez-orozco, C., & Rogers-sirin, L. (2013). Understanding the role of social support in trajectories of mental health symptoms for immigrant adolescents ☆. *Journal of Applied Developmental Psychology*, *34*, 199–207.

Smith, M. (2013). Evidence-based education: is it really that straightforward?. *The Guardian*. Retrieved from https://www.theguardian.com/teacher-network/2013/mar/26/teachers-research-evidence-based-education

Smythe-Leistico, K., & Page, L. C. (2018). Connect-Text: Leveraging Text-Message Communication to Mitigate Chronic Absenteeism and Improve Parental Engagement in the Earliest Years of Schooling. *Journal of Education for Students Placed at Risk (JESPAR)*, *0*(0), 1–14.

Soenens, B., & Vansteenkiste, M. (2005). Antecedents and outcomes of self-determination in 3 life domains: The role of parents' and teachers' autonomy support. *Journal of Youth and Adolescence*, *34*(6), 589–604.

Song, J., Bong, M., Lee, K., & Kim, S. (2015). Longitudinal investigation into the role of perceived social support in adolescents' academic motivation and achievement. *Journal of Educational Psychology*, *107*(3), 821–841.

Spiegel, D., Morrow, G. R., Classen, C., Raubertas, R., Stott, P. B., Mudaliar, N., ... & Riggs, G. (1999). Group psychotherapy for recently diagnosed breast cancer patients: a multicenter feasibility study. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer, 8*(6), 482-493.

Stice, E., Ragan, J., & Randall, P. (2004). Prospective relations between social support and depression: differential direction of effects for parent and peer support? *Journal of Abnormal Psychology*, *113*(1).

Straker, L. M., Fenner, A. A., Davis, M. C., & Hagger, M. S. (2014). Theoretical underpinnings of a need-supportive intervention to address sustained healthy lifestyle changes in overweight and obese adolescents. *Psychology of Sport and Exercise, 14*(6), 819-829.

Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'Getting it'for the white working class. *Research Papers in Education*, *29*(2), 131–171.

Swain, J. (2005) 'Beyond the daily application': motivations for adults attending numeracy classes. *Research in Post-Compulsory Education, 10,* 305-323.

Swan, M. (2006). Learning GCSE mathematics through discussion: what are the effects on students?. *Journal of Further and Higher Education, 30*(3), 229-241.

Tamres, L. K., Janicki, D., & Helgeson, V. S. (2002). Sex differences in coping behavior: A meta-analytoc revoew and an examination of relative coping. *Personality and Social Psychology Review*, *6*(1), 2–30.

Tashakkori, A., Teddlie, C., & Johnson, B. (2015). *Mixed Methods. International Encyclopedia of the Social & Behavioral Sciences* (Second Edition, Vol. 15). Elsevier.

Taylor, S. E. (2011). Social support: A review. *The Handbook of Health Psychology*, 192–217.

Thoits, P. A. (1986). Social support as coping assistance. *Journal of Consulting and Clinical Psychology*, *54*(4), 416–423.

Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior, 52*(2), 126–141.

Thoits, P. A. (1995). Stress, Coping, and Social Support Processes: Where Are We? What Next? *Journal of Health and Social Behavior*, *35*(1995), 53.

Tilbrook, H. E., Hewitt, C. E., Aplin, J. D., Semlyen, A., Trewhela, A., Watt, I., & Torgerson, D. J. (2014). Compliance effects in a randomised controlled trial of yoga for chronic low back pain: A methodological study. *Physiotherapy*, *100*(3), 256–262.

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, *45*(1), 89–125.

Torgerson, C. & Torgerson, D. J. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. Palgrave Macmillan UK.

Tymms, P. (2012). *Interventions: experiments*. In Research Methods and Methodologies in Education. Arthur, J., Waring, M., Coe, R. & Hedges, L.V. Sage. 137-140.

Uchino, B., Cacioppo, J., & Kiecolt-Glaser, J. (1996). The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms and implications for health. *Psychological Bulletin*, *119*(3), 488–531.

Ussher, J., Kirsten, L., Butow, P., & Sandoval, M. (2006). What do cancer support groups provide which other supportive relationships do not? The experience of peer support groups for people with cancer. *Social science & medicine, 62*(10), 2565-2576.

Van Belle, S., Wong, G., Westhorp, G., Pearson, M., Emmel, N., Manzano, A., & Marchal, B. (2016). Can 'realist' randomised controlled trials be genuinely realist? *Trials*, *17*(1), 1–6.

Van der Steeg, M., van Elk, R., & Webbink, D. (2015). Does intensive coaching reduce school dropout? Evidence from a randomized experiment. *Economics of Education Review*, *48*, 184–197.

Vasquez, A. C., Patall, E. A., Fong, C. J., Corrigan, A. S., & Pine, L. (2016). Parent Autonomy Support, Academic Achievement, and Psychosocial Functioning: a Meta-analysis of Research. *Educational Psychology Review*, *28*(3), 605–644.

Wallace, S. (2013). When you're smiling: exploring how teachers motivate and engage learners in the further education sector. *Journal of Further and Higher Education*, (September), 1–15.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*(6023).

Wang, M., & Eccles, J. S. (2012). Social support matters: longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development*, *83*(3), 877–95.

Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of Educational Psychology*, *90*(2), 202–209.

Wentzel, K. R., Battle, A., Russell, S. L., & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology*, *35*(3), 193–202.

Wentzel, K. R., & Caldwell, K. (1997). Friendships, Peer Acceptance, and Group Membership: Relations to Academic Achievement in Middle School. *Child Development*, *68*(6), 1198–1209.

Wentzel, K. R., Russell, S., & Baker, S. (2016). Emotional support and expectations from parents, teachers, and peers predict adolescent competence at school. *Journal of Educational Psychology*, *108*(2), 242–255.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal Inference and Generalization in Field Settings: Experimental and Quasi-Experimental Designs. In *Handbook of research methods in social and personality psychology* (Vol. 1104, pp. 40–84).

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's Perspectives on Causal Inference. *Psychological Methods*, *15*(1), 18–37.

Wheater, R., Burge, B., Sewell, J., Sizmur, J., Worth, J., & Williams, J. (2013). The International Survey of Adult Skills 2012 : Adult literacy, numeracy and problem solving skills in England, (139), 161.

Wheater, R., Burge, B., & Sewell, J. (2013). *The International Survey of Adult Skills 2012: Adult literacy, numeracy and problem solving skills in Northern Ireland*. Belfast: Department for Employment and Learning.

White, H. (2013). The use of mixed methods in randomized control trials. *New Directions for Evaluation, 2013*(138), 61-73.

White, I. R., Horton, N. J., Carpenter, J., & Pocock, S. J. (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. *Bmj*, *342*(7803), 910–912.

Whittemore, R., Rankin, S. H., Callahan, C. D., Leder, M. C., & Carroll, D. L. (2000). The peer advisor experience providing social support. *Qualitative Health Research*, *10*(2), 260–276.

Williams, J., Hadjivassiliou, K., Marvell, R., Green, M., & Newton, B. (2017). *Effective curriculum practice at below Level 2 for 16/17 year olds July 2017*. The Department for Education.

Williams, P., Barclay, L., & Schmied, V. (2004). Defining social support in context: a necessary step in improving research, intervention, and practice. *Qualitative Health Research*, *14*(7), 942–960.

Wilson, T. D., & Gilbert, D. T. (2005). Affective forcasting: Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131–134.

Wolf, A. (2015). *Heading for the precipice: Can further and higher education funding policies be sustained?*. London: Policy Institute, King's College London.

Xu, C. W. (2017). Essays in Education, Peer Effects and Decision Making (Doctoral Dissertaton), University of Michigan. Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/138504/wjxu_1.pdf?sequence=1&isAllowed=y

Yeager, D. S., & Dweck, C. S. (2012). Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. *Educational Psychologist*, *47*(4), 302–314.

Yeager, D. S., Dahl, R. E., & Dweck, C. S. (2018). Why interventions to influence adolescent behavior often fail but could succeed. *Perspectives on Psychological Science, 13*(1), 101-122.

York, B. N., & Loeb, S. (2014). *One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers*. NBER Working Paper No. 20659.

York, B. N., Loeb, S., & Doss, C. (2018). One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers. *Journal of Human Resources*.