Australian Government
Department of the Prime Minister and Cabinet

BETA

Gather. Grow. Graduate.

Using behavioural insights and technology to help students graduate from university

June 2021

`

Other uses

Enquiries regarding this license and any other use of this document are welcome at:

Managing Director
Behavioural Economics Team of the Australian Government
Department of the Prime Minister and Cabinet
Barton ACT 2600
Email: beta@pmc.gov.au

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Prime Minister and Cabinet, the Department of Social Services or the Australian Government.

Research team

Jacqualine Myint, Eileen Chu, Loren Willis, Bethany Jones, Roxarne Armstrong, David Blackman and Harry Greenwell.

Acknowledgments

Thank you to the Department of Social Services, The Behavioural Insights Team, The Project Factory, Western Sydney University, the University of Newcastle, and the Australian National University for their support and valuable contribution towards this project. We thank all staff who were involved at our partnering universities for their courage and fortitude throughout this project. Special thanks also to our university executive sponsors, Liz Burd and Michael Burgess.

Thank you to:

- The Project Factory: Lynette Reeves, Matt Purchase and Stephanie Watson
- Academic advisors: Katherine Reynolds and Jolanda Jetten
- University of Newcastle: Mark Kirby and Chantelle Wilcox
- Western Sydney University: Danielle Poultney, Natalie Bradbury and Lona Moitra

The trial was pre-registered on the BETA website and the American Economic Association registry:

https://behaviouraleconomics.pmc.gov.au/projects/strengthening-students-resilience

https://www.socialscienceregistry.org/trials/5500

Who?

Who are we?

We are the Behavioural Economics Team of the Australian Government, or BETA. We are the Australian Government's first central unit applying behavioural economics to improve public policy, programs and processes.

We use behavioural economics, science and psychology to improve policy outcomes. Our mission is to advance the wellbeing of Australians through the application and rigorous evaluation of behavioural insights to public policy and administration.

What is behavioural economics?

Economics has traditionally assumed people always make decisions in their best interests. Behavioural economics challenges this view by providing a more realistic model of human behaviour. It recognises we are systematically biased (for example, we tend to satisfy our present self rather than planning for the future) and can make decisions in conflict with our own interests.

What are behavioural insights and how are they useful for policy design?

Behavioural insights apply behavioural economics concepts to the real world by drawing on empirically-tested results. These new tools can inform the design of government interventions to improve the welfare of people.

Rather than expect people to be optimal decision makers, drawing on behavioural insights ensures policy makers will design policies to go with the grain of human behaviour. For example, people may struggle to make choices in their own best interests, such as saving more money. Policy makers can apply behavioural insights to preserve freedom, but encourage a different choice – by helping people to set a plan to save regularly.

Contents

Executive summary

BETA built and tested a new mobile app to help university students gather important support groups, grow their resilience and stay on track to graduate.

Graduating from university is a critical step to starting a career. University students can risk transitioning into unemployment or long-term welfare dependence if they do not complete their studies. Students who experience disadvantage are particularly at risk. Strategies to help university students graduate can help achieve positive individual, educational and broader economic outcomes.

BETA identified two key behavioural insights related to university non-completion rates for disadvantaged students: (i) social groups impact feelings of belonging (ii) and negative attributions affect resilience.

Supported with funding for the Strengthening Student Resilience project from the Try, Test and Learn Fund – an initiative of the Australian Government Department of Social Services (DSS), BETA designed a mobile app, informed by those behavioural insights, called Grok. The aim of Grok was to improve university grades and increase completion rates by growing student resilience.

Grok encouraged students to connect with important social groups to boost resilience. Grok reminded students of their self-worth and value, both within and external to the university context. Grok also challenged negative attributions by showing students alternative ways of thinking, normalising difficulties and offering practical tools to build confidence.

BETA conducted a randomised controlled trial with Grok during the first semester of 2020 at two Australian universities with 4,463 student downloads. BETA evaluated the impact of access to Grok on academic performance, completion rates, wellbeing and feelings of belonging at university.

No significant differences were found in any of the outcomes, for both disadvantaged students and non-disadvantaged students. Student engagement with Grok appeared to be a key issue, as download numbers were high but ongoing app usage was low. For the students who did complete Grok activities, there was no measurable change in their wellbeing or completion rates, but the feedback from interviews and surveys was predominately positive.

Increasing university completion rates is difficult and complicated. Mobile apps are highly accessible but it is challenging to maintain user engagement. BETA could make Grok accessible to other researchers for further testing to better understand its impact under different circumstances and its potential for broader application.

Why?

Students who drop-out of university are at risk of welfare dependency

In 2018 about 800,000 Australian domestic students started a bachelor degree (DESE Higher Education Statistics uCube, 2018) but around one in five are likely to leave university without a degree (Norton, Cherastidtham & Mackey, 2018). Early student dropout, the proportion of students who did not return in second year to their bachelor degree, is also trending up (Cherastidtham, Norton & Mackey, 2018). Students who do not complete their tertiary education risk transitioning into unemployment. Students from a disadvantaged background who do not complete their studies are at particular risk of long-term welfare dependence.

A range of behavioural factors contribute to university non-completion

The higher education policy system is complex and there are many factors determining longer-term outcomes for students. A combination of socio-economic, demographic, and personal factors contribute to non-completion of university degrees[1]. Behavioural insights cannot overcome these underlying structural issues, but can overcome the many behavioural barriers leading to non-completion.

Research indicates motivation, persistence, time put into study and knowledge of university expectations can lead to degree completion (Cherastidtham et al., 2018). Students who have good relationships with their peers and who are more socially engaged in university life are more likely to complete their degree (Cherastidtham et al., 2018). Students who persistently fail subjects can find it hard to 'bounce-back' and ultimately choose to drop out.

Addressing behavioural barriers could improve completion rates

BETA worked in partnership with the Behavioural Insights Team (BIT), who were funded by DSS through the Try, Test and Learn Fund for the Strengthening Students' Resilience project. This partnership sought to increase young students' engagement in education by using behavioural insights and technology to develop resilience and skills to complete their studies.

As completion rates have remained relatively unchanged since the 1970s, despite many policy changes, an innovative solution is needed (DESE, 2017). The primary focus is on young people at risk of dropping out of their tertiary studies. There are many reasons why a student is considered 'at-risk' of non-completion. For the purposes of this report, a student is considered at-risk if they belong to or identify with a 'disadvantaged group' or cohort. A disadvantaged group can be any group of people at higher risk of social exclusion, discrimination or cultural bias than the general population.

[1] We recognise for some students, non-completion of a degree is not the 'wrong' outcome. Other pathways like deferring study, going to TAFE or choosing a vocational course may be more aligned with their needs and interests.

What we did: Behavioural design

BETA conducted extensive research to identify behavioural barriers of student completion.

Students from disadvantaged groups find it difficult to 'bounce back' from university setbacks, leading to declining grades and non-completion

All students are likely to experience setbacks during their degree, such as receiving a poor grade or not being chosen for selective opportunities. Students who tend to have academic success are those with **resilience - the ability to 'bounce back' quickly after a setback**.

BETA identified two key behavioural insights from social psychology research to explain why students from disadvantaged groups could struggle with resilience at university.

Behavioural insight 1

The social groups we belong to influence how we see ourselves and how we behave – but they can help or hinder

Students who identify with a disadvantaged group can question their belonging at university. Seemingly innocuous cues associated with university can trigger feelings of self-doubt and small academic setbacks can be seen as proof they do not belong.

The social groups with which we identify have a big impact on beliefs about who we are and how we behave. These groups can be small, like family and school friends, or they can be large, like national identity and religious groups. An individual can be part of many groups simultaneously and each group has certain characteristics setting them apart.

University can be a stressor for students who identify with disadvantaged groups, as the context can bring out existing insecurities and **the idea they 'don't belong'**. They may identify with groups not traditionally well represented in higher education, such as an ethnic minority, or they may be aware of negative stereotypes about the ability of their group.

Constant thoughts about not belonging at university can lead to stress, feelings of isolation and **can adversely affect academic performance** (Walton & Spencer, 2009). A student who feels they do not belong may also struggle with typical student behaviours, such as attending class. This can hinder their ability to participate effectively at university, potentially leading to non-completion.

Behavioural insight 2

Negative attributional biases diminish the ability to 'bounce-back'

Attributions are assumptions about the underlying cause of our own or others' behaviour. Students from disadvantaged groups tend to have insecurities about their place at university and are more likely to make negative assumptions about their academic performance.

Attributions do not always reflect reality and can be 'coloured' by perspective, a phenomenon known as **attribution bias**. For example, if a student fails an exam, a helpful (positive) attribution would be:

'I probably failed because I didn't study enough and I was tired. For the next exam, I will schedule my time better and get more rest' ← The student recognises they **can do better next time** and the setback is temporary.

An unhelpful (negative) attribution would be:

'*I failed because I am not smart enough. I am a bad student. People like me don't belong at university'* ← The student believes they **cannot do better next time**, and the setback is due to unchanging personal traits.

How a student thinks about university setbacks will affect how quickly they overcome them. Misattributing setbacks to personal traits makes it harder to learn from mistakes. This can harm a student's ability to recover and could **lead to non-completion**.

BETA identified solutions for each behavioural insight to improve resilience

Behavioural insight 1

The social groups we belong to influence how we see ourselves and how we behave – but they can help or hinder

Solution 1: Increase feelings of social connection with a range of social groups

Remind students of *other* social groups they belong to and create opportunities for students to feel connected with important social groups to bolster resilience.

The more groups a student feels connected to, the more likely they will have a group able to support them after a setback. If a student is disadvantaged by having fewer established connections at university, they can draw on social groups outside of university as a 'social safety net' and to gain perspective. The same situation, like receiving a poor grade, can appear less stressful to students who perceive they have more social support (Griffin, Steptoe & Cropley,1999). **Social belonging** can be increased by asking students to:

- **Create a visual map of important social connections**

Asking students to draw (either on paper or electronically) a 'map' of all their groups and how those groups connect to each other can help create a sense of belonging and feeling of personal value (Epton, Harris, Kane, van Koningsbruggen & Sheeran, 2014). The process of

mapping and the final product can act as a reminder of all the groups they can draw on for support and the ample social resources they have to cope in stressful situations.

- **Set aside time to reflect on important relationships**

Reminders of positive group characteristics and belonging can boost personal resilience and feelings of self-worth. Simply thinking or writing about the people we care for and why they are special has been shown to improve academic performance in disadvantaged groups (Shnabel, Purdie-Vaughns, Cook, Garcia, & Cohen, 2013).

- **Regularly interact with groups both inside and outside of university**

Encouraging students to build university ties can slowly increase feelings of belonging. This should be balanced with messages about maintaining communication with groups beyond university. Knowing one has a diverse, strong safety net can make developing new student behaviours and meeting new university friends feel less daunting.

Behavioural insight 2

Negative attribution biases diminish the ability to 'bounce-back'

Solution 2: Offer alternative explanations for difficulties and provide practical tips

Challenge negative attributions by showing students alternative ways of thinking, normalising difficulties and emphasising setbacks are learning opportunities.

If a student believes academic setbacks are a result of traits they cannot change, there is little motivation to learn from mistakes. Providing timely, understandable information can minimise the tendency to internalise setbacks. Ways to increase positive student attributions include:

- **Provide alternative explanations of behaviour and outcomes**

Providing explanations about why a *situation* is challenging makes it harder for students to hold negative attributions about their ability. Informing students others struggle with transitioning to university can help students realise setbacks are normal, situational and temporary.

- **Encourage students to look for evidence of how they are 'worthy'**

Reflecting on unique values to build self-worth can stop negative attributions 'sticking'. Encouraging students to think about past academic successes and personal strengths can help them realise they have overcome setbacks before and are capable of doing it again (Sherman et al., 2013).

- **Give students practical tools to overcome new challenges**

Teaching students self-care practices and academic` skills can give them confidence to tackle challenges. Understanding all students need to practice and grow ability over time means students are less likely to react defensively when they receive negative feedback.

What we did: App design

BETA designed a novel mobile app based on behavioural insights to increase student resilience.

BETA chose a mobile app because university students have higher engagement levels with mobile apps, more so than a website or desktop app (Meeker, 2019). Mobile apps are highly available and easily accessible, with 98% of 18-25 year olds owning a smartphone (Iqbal, 2020). Apps are customisable to the needs of the user and the university, creating a unique, personalised experience.

BETA reviewed existing mobile apps across the categories of time-management, goal-setting, wellbeing, study skills and social connection. This was critical to understand current gaps in the market. We found no available app combined the key behavioural solutions identified earlier as critical to the target group, nor focused on an educational experience tailored to the Australian context.

BETA worked with an app developer to create an app using behavioural insights with gamified elements to encourage student engagement. Gamified mobile apps are more successful at engaging and motivating students than non-gamified apps (Hamari, Koivisto & Sarsa, 2014) and lead to better behavioural outcomes, such as increased academic performance (Pechenkina, Laurence, Oates, Eldridge & Hunter, 2017).

Figure 1 provides an overview of how the app design and content was refined with extensive user testing and feedback across three Australian universities.

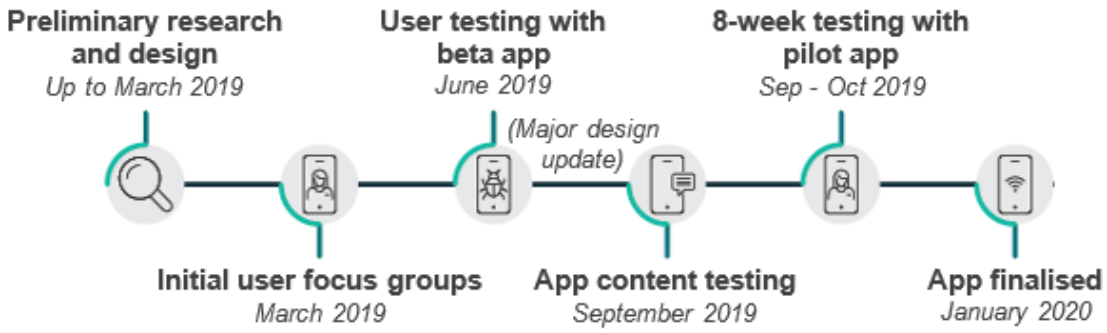


Figure 1: Main stages of the app design.

BETA created Grok[2], an interactive Zen garden to 'grow' student resilience

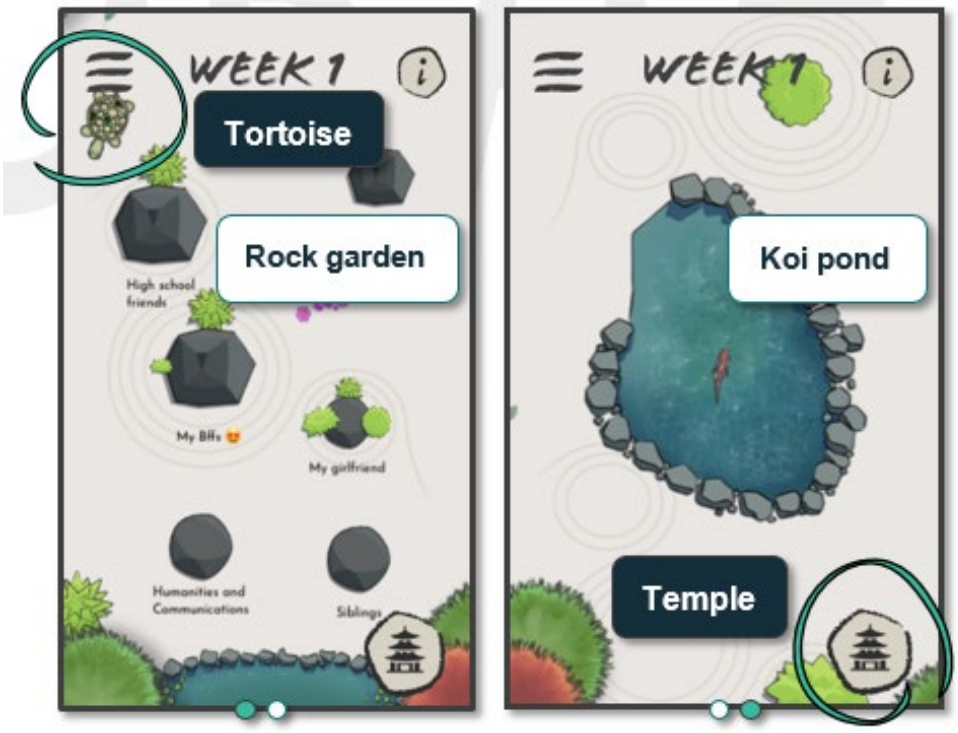
The garden was an analogy for personal growth and nurturing of the self to become more resilient. This concept resonated with students during user testing:

> *'I enjoyed using the app as it reminded me to be mindful and grow my garden as well as myself.'*
>
> *'The Zen garden gave me a living representation of my goals being met, kind of like keeping a pet.'*

The **content** focused on the two key behavioural solutions identified to improve resilience:

1. **Increase feelings of social belonging**: **The rock garden** represented important social connections.
2. **Provide alternative explanations for difficulties and deliver practical tips**: The **tortoise** represented reflections on self-worth, self-care practices and healthy student behaviours. The **Koi pond** represented visualisations of academic success, practical study tips and tools, and explanations of typical student behaviours. The **temple** represented key wisdom, providing alternative explanations for setbacks at university and research facts about the importance of resilience.

The four areas housed different activities released on a weekly basis over the 16-week semester. Students set their own pace navigating the app and choosing activities of interest.



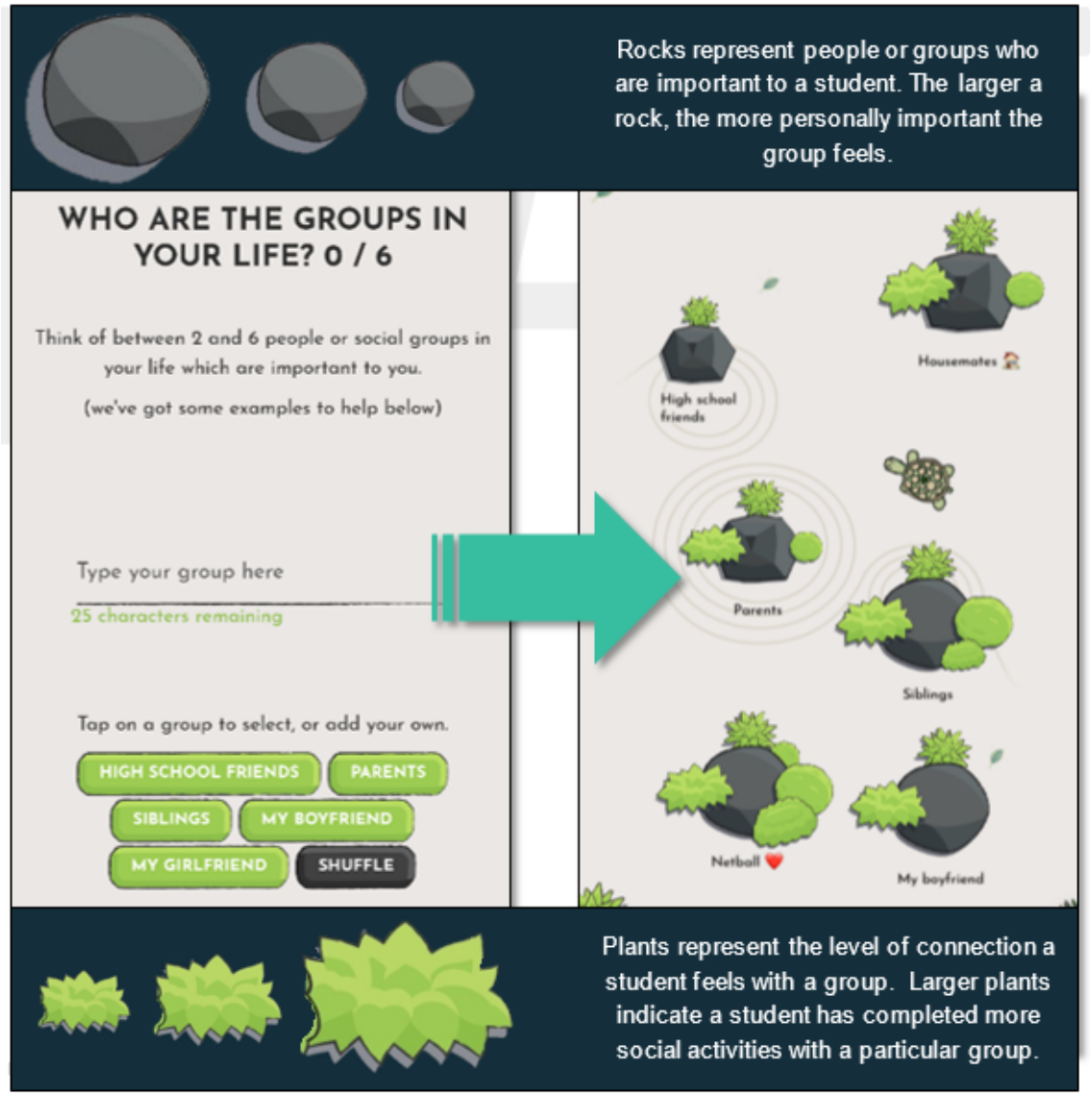
[2] Grok means to understand something deeply. Writer Robert A. Heinlein coined the term in his book 'Stranger in a Strange Land'. To 'grok' something meant to be 'merged with' or to have become a part of something bigger through deep understanding. Today, the term is popular in IT and means to understand something intuitively.

Grok offered a way to increase feelings of social connection with more social groups

Rock garden – A visual map of important social connections

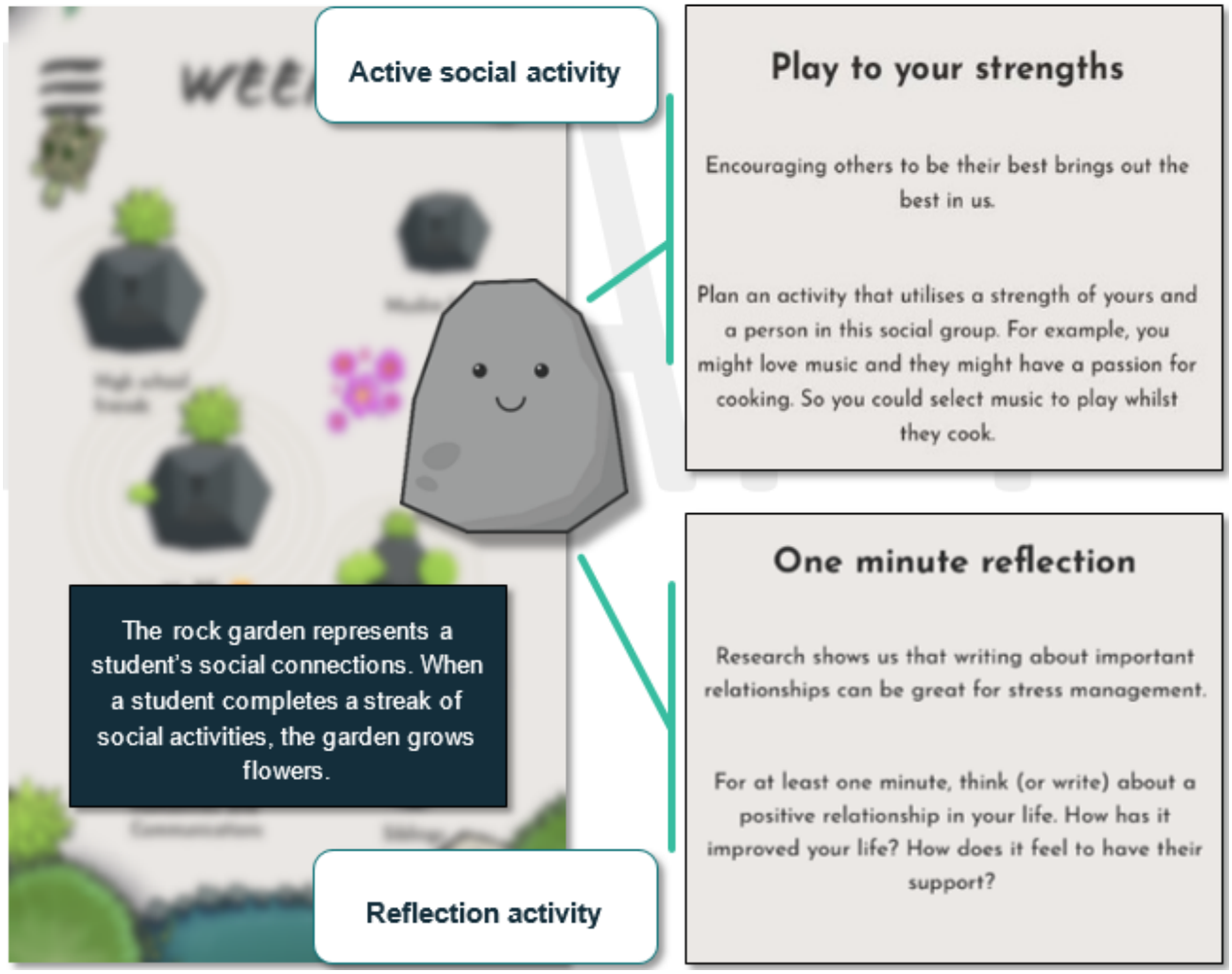
The rock garden is core to Grok's design and is a **visual map of important social connections,** symbolising a student's **social world** within and outside of university. Students create their personalised social maps by nominating 2 - 6 important people or social groups, represented by rocks in the garden. The size of the rock signified the **relative personal importance** of social groups (i.e. the larger the rock, the more important a group feels).

The **strength of connection** to the group is symbolised by the **plants around a rock** (i.e. the more plants around the rock, the stronger the connection feels). The plants grow as a student completes social activities. Students are encouraged to 'grow' their social connections by completing two broad types of social connection activities.



Active social activities focused on **regular interaction and quality communication** to maintain and build connections. The activities often involved interaction with at least one other person in either a virtual or in-person setting. The content was updated on a regular basis to ensure adherence to COVID-19 advice. The goal was to strengthen existing social connections and build the student's support network by encouraging regular contact.

Reflection tasks focused on **setting aside time to reflect on important relationships.** They could be completed alone and asked students to visualise, categorise or write about important social relationships and group values. The reflection tasks were easy and brief so students could complete them regularly and in periods of stress. The goal was to allow students to think deeply about their groups and encourage them to focus on positive group characteristics.

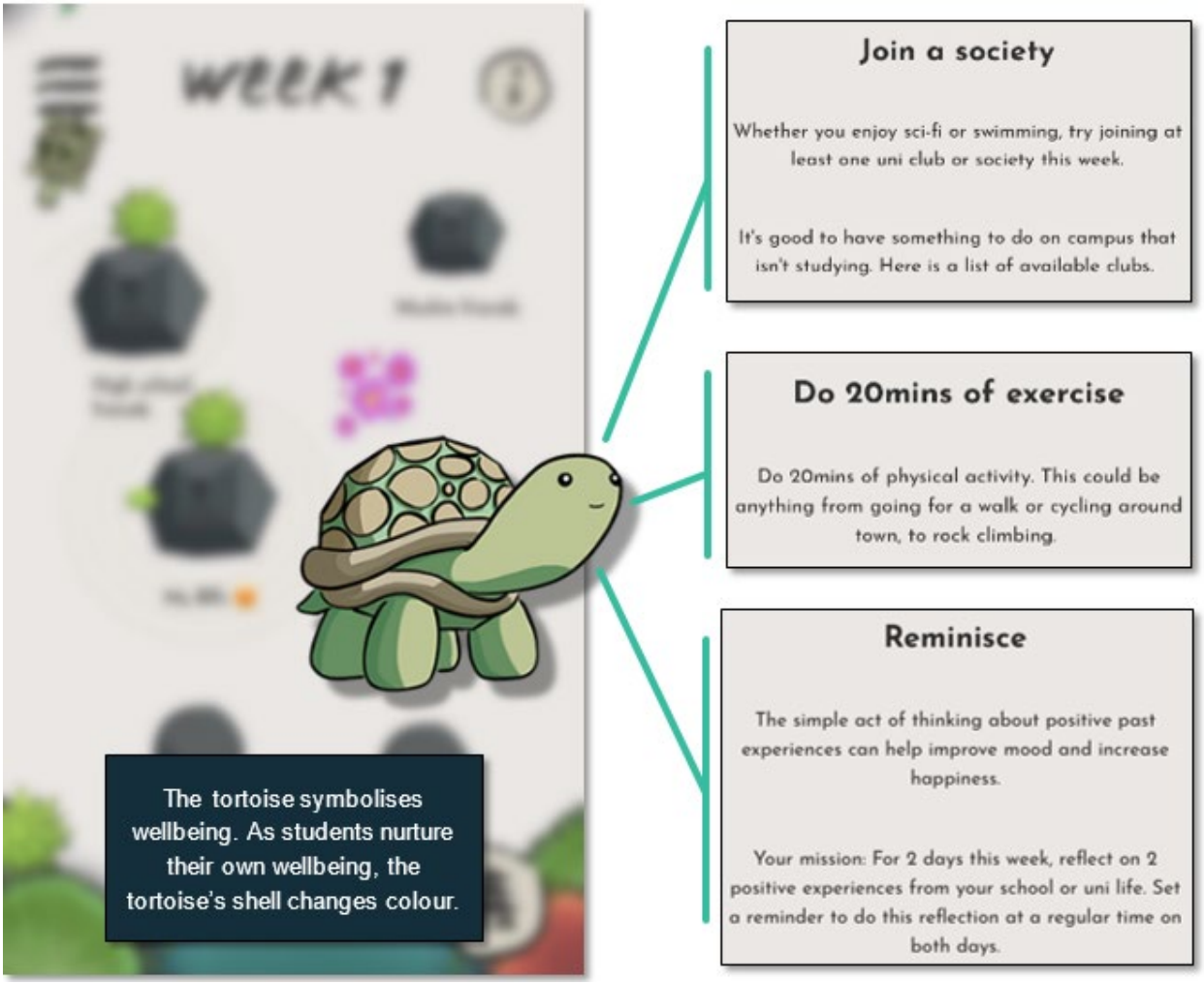


Grok provided alternative explanations for difficulties and delivered practical tips to increase resilience

1. Wellbeing activities – Practising self-care

The wandering tortoise acted as a wellbeing guide and housed the weekly wellbeing activities. The aim of the wellbeing activities was to offer evidence-based strategies for dealing with stress and skills to increase resilience. The activities focused on **encouraging students to look for evidence of how they are 'worthy'** and **give students practical tools** to build resilience through self-care practices.

The activities were a mix of individual reflections to increase feelings of self-worth, campus activities to encourage a healthy transition into a new environment, physical wellbeing activities (such as exercising and eating healthily) and psychological wellbeing exercises (such as practising mindfulness and gratitude).



2. Academic Skills – Improving study behaviour

The Koi pond represented academic studies and housed the weekly academic skills. The aim of the academic skills was to give **students practical tips** to decrease informational barriers and increase their ability to follow degree requirements. By providing advice on how to prepare for a successful semester, Grok offered ways to improve the university student experience and **enhance academic performance**.

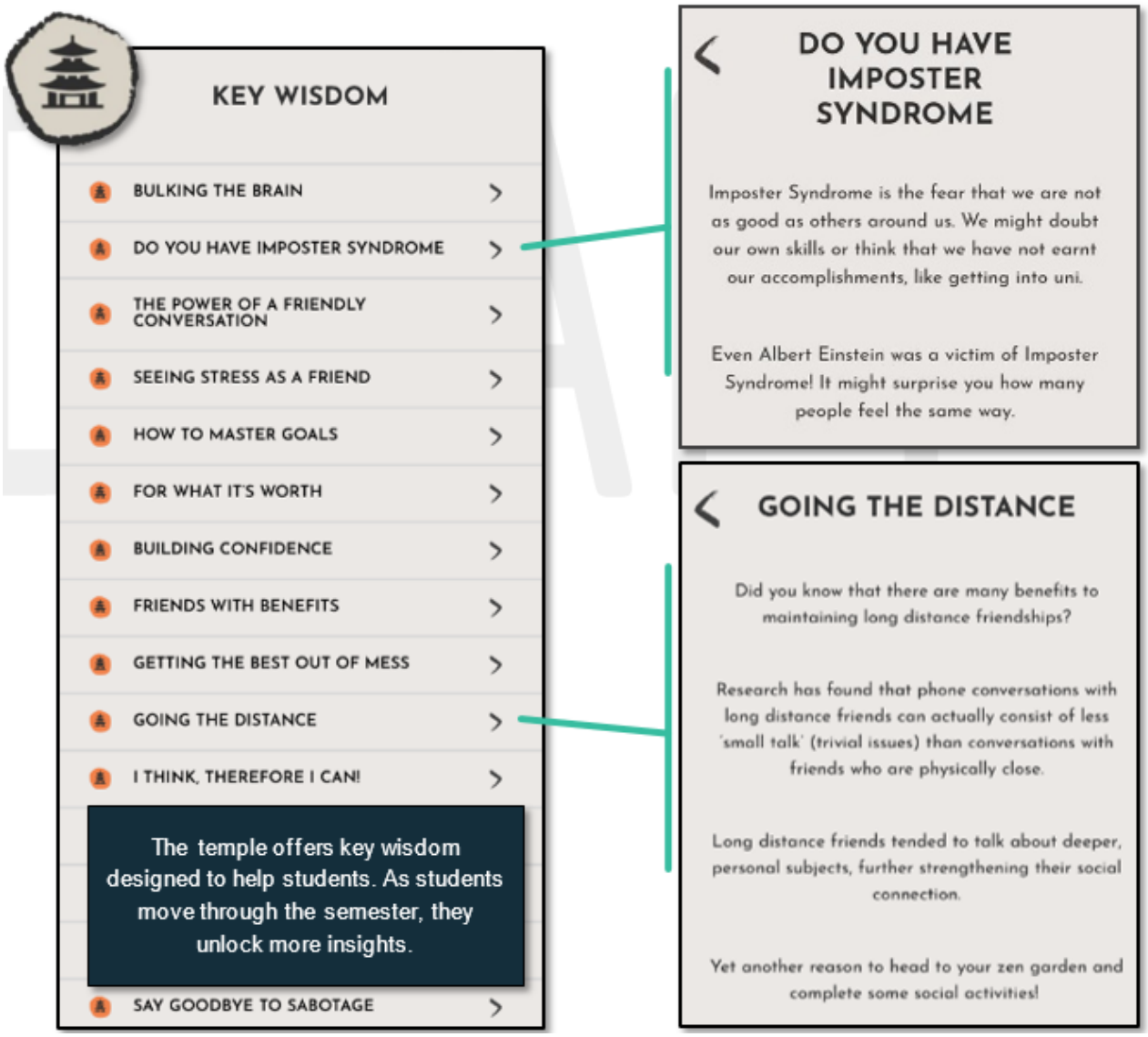
The academic skills also tried to strengthen connection to the 'university student' identity and encouraged active participation on campus, by providing tools to navigate typical university student behaviours.

The activities were a mix of individual reflections to remind students of past successes, practical study skills (such as how to calculate assignment timelines) and campus information (such as how to book a library room), to slowly introduce students to expected behaviour. When COVID-19 restrictions were imposed, the activities were updated to encourage online engagement with student resources.

3. Key wisdom – Building confidence

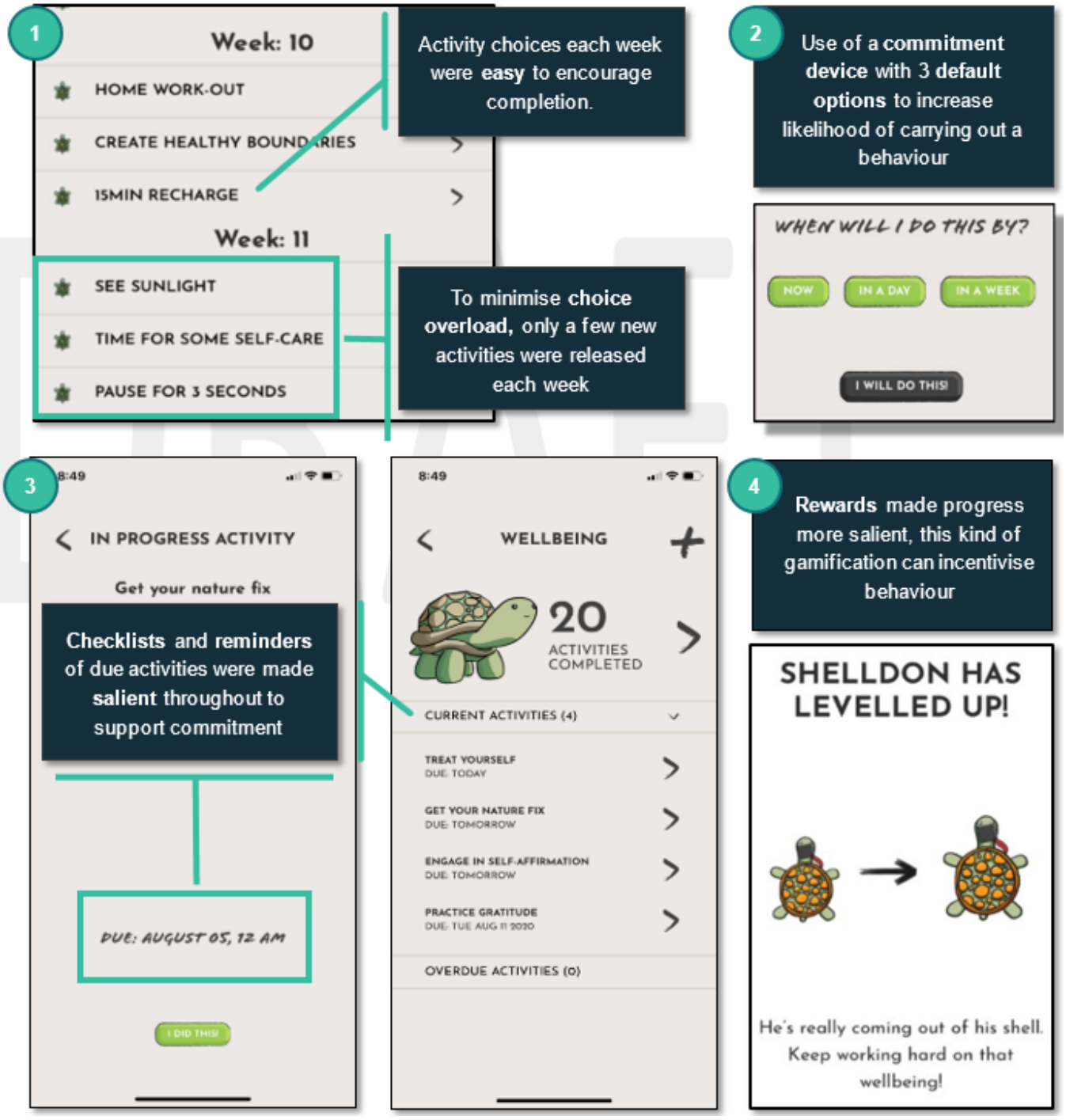
The key wisdom section includes evidence-based behavioural ‘titbits’ unlocked weekly and housed in the garden temple. The aim of the key wisdom section is to **provide alternative explanations of behaviour and outcomes** by informing students about relevant scientific research normalising the stresses and challenges university students typically face.

The key wisdom section draws from a number of research areas in social psychology. It helps to illuminate common university pitfalls, explain the importance of building resilience and offers helpful techniques to overcome behavioural biases often affecting students.



BETA designed Grok with behavioural insights to encourage continued engagement with the activities

In addition to using social psychology to inform the app content, BETA designed the functions of Grok with behavioural insights in mind. BETA used a combination of **commitment devices, reminders** and **gamification** to maximise student participation and engagement.



Each behaviourally informed design feature aimed to increase user friendliness and encourage regular use. The activities were released weekly at specific points during the semester to maximise the relevance of information and avoid overwhelming the students.

Feedback from user testing provided suggestions into which elements students would value the most and what would attract them to Grok. These suggestions were combined with BETA's knowledge of behavioural insights to build an app inspiring long-term positive behaviour change in university students.

Further behavioural principles applied to encourage continued app use

Personalised: Users had autonomy to customise the Zen garden by labelling meaningful groups, adjusting their rocks and choosing activities, creating a sense of ownership and making them more likely to follow through on behaviour change

Attractive display: Displaying study skills and information about common university challenges in an attractive, colourful setting helps to introduce users to new concepts and unfamiliar behaviours in a non-threatening manner

Timely: Weekly content followed the semester schedule so new activities were appropriate and users were prompted at the most relevant time

Time-bound information: Only a few activities were released at a time so the user was not overwhelmed. The user was encouraged to follow the program for at least four weeks, as the time-bound nature made the task seem more achievable

Salient reminders: Notifications acted as salient reminders when novel activities were released at the start of the week and mid-week to encourage ongoing engagement

Easy navigation: After an on boarding tutorial, users had access to an information button housing tips about features of the app. Users are more likely to take action when instructions are easy to understand and action is clear

Reduce friction: Hyperlinks to university specific information, such as maps of campus, and examples, such as template study schedules, removed friction points to make it easier for users to act

Pre-commitment devices: Users clicked a pre-commitment device to pledge they would complete an activity by a specific time and were prompted when activities were overdue, making it more likely they would follow through

Checklists: Checklists organised the user's pledged activities into a manageable display and contributed to a sense of accomplishment when ticking off completed commitments

Incentives: The Zen garden 'levelled-up' after completing a streak of activities which encouraged a sense of achievement, was a visual representation of progress and motivated students to continue completing activities

What we did: Trial design

Grok was evaluated in a two-arm Randomised Controlled Trial (RCT) with undergraduate students at a metropolitan and a regional Australian university from February to June 2020 (Figure 2). We evaluated the impact of access to Grok on key outcomes including academic performance and completion rates. We also surveyed and interviewed students to collect information about their sense of university belonging and wellbeing.

What is a Randomised Controlled Trial (RCT)?

Well-designed randomised controlled trials are the best available method for determining whether policies, programs or services have a specific intended impact. RCTs work by separating people into two or more groups randomly, in a manner similar to flipping a coin. People in the 'treatment' groups are assigned to receive an intervention while people in the 'control' group are not. The control group receives either the business-as-usual experience or nothing. On average, the difference in outcomes between people in the groups reflects the effect caused by the intervention.

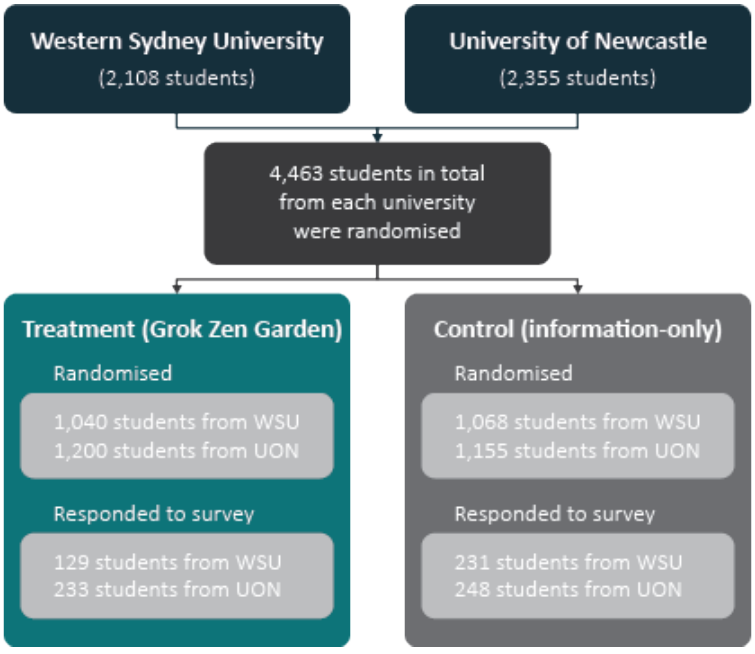


Figure 2: Design of the Grok RCT

Participants

The population of interest for the trial was undergraduate students at the University of Newcastle (UON) and Western Sydney University (WSU) during semester one, 2020. Undergraduate students from all campuses at both universities were invited to participate, though marketing and outreach activities focused primarily on three UON campuses (Callaghan, Ourimbah and NewSpace – a total of over 20,000 students) and four WSU campuses (Parramatta, Parramatta CBD, Penrith and Campbelltown – a total of around 16,000 students). The trial sites focused on institutions with historically high non-completion rates and diverse student populations.

Recruitment

The student engagement teams emailed eligible students an invitation to participate in the trial and posted multiple sign-up links on the internal university social media pages. We prepared an invitation slide for teaching staff to include in their initial lectures and placed an invitation on the student portal. Digital signs and posters were displayed around campus and bookmarks were included in the UON orientation bags and WSU student diaries. BETA members also attended student welcome events on campus and offered small incentives for downloading and registering Grok. These events were interrupted at the UON due to flooding and at WSU due to the COVID-19 pandemic.

Our goal was to recruit at least 3,000 students at each university to detect differences in primary outcomes. We recruited 2,355 UON students and 2,108 WSU students for a **total 4,463 participants in the trial**.

Randomisation

Once registered, participants were randomly assigned ‘on-the-spot’ to treatment or control with a 50% chance of being assigned to treatment. **The treatment version of the app** allowed access to the Zen garden with weekly activities. **The control (information-only) version,** did not contain the Zen garden or activities, but gave students information they would normally receive from their university.

Randomisation took place at the student-level, and occurred when an eligible student successfully downloaded the app, consented to participation in the trial, and entered a valid student identifier and email address. The randomisation was implemented in-app as part of a student’s registration and on boarding process.

Outcomes and hypotheses

There were **two primary outcome measures** for this project:

- **Academic performance**: measured by subject marks out of 100 at the end of semester one.
- **Completion**: measured by the units successfully completed as a proportion of all units a student was enrolled in after census date[3].

Each primary outcome was measured at the level of the study unit and clustered by student. This approach resulted in a dataset of 4,463 participants with a total of 15,259 study units (7,549 in the Zen garden group and 7,710 in the information-only group).

We hypothesised the group with access to Grok would have higher marks, and higher rates of completion compared to the control group.

There were **two secondary outcomes measures**, wellbeing and student identification, captured by a survey at the end of the intervention. The final survey had 853 responses (360 from WSU and 481 from UON). This represents a 19% response rate.

We used the Flourishing Scale (Diener et al., 2010) to measure wellbeing (see Appendix 1). This is an 8-item self-perception measure of optimism, purpose, self-esteem and relationships. We measured student identification using the Three Dimensional Strength of Group Identification Scale (Cameron, 2004) (see Appendix 2). This scale comprises three subscales: **Centrality**, **Ingroup Affect** and **Ingroup Ties**. Centrality refers to the prominence of a social identity to the individual, Ingroup Affect measures the emotional evaluation of group membership and Ingroup Ties refers to the perception of similarity and bonds with other group members.

Impact of the COVID-19 pandemic

On 11 March 2020, the World Health Organisation declared a pandemic coinciding with the early weeks of the trial. The pandemic resulted in campus closures, changes in the delivery of learning to students, grading policies and the timing of examinations. This had a considerable impact on app content and recruitment, but little impact on the trial and analysis plan.

Participating universities also changed the timetabling of the final exams for semester one, 2020. This meant there were extra weeks in the semester, which had implications for the app content, but did not affect trial design. Campus closures and social distancing meant much of the app content had to be re-written to reflect distance learning, and a lack of on-campus interactions (see Appendix 3). The biggest impact from the pandemic on the trial was generalisability. This is addressed in the Limitations section.

[3] The census date is when a student's enrolment in a subject is finalised. It is the last date a student may withdraw from a subject without financial penalty. In the autumn semester one, 2020, the census date for the UON was the 20th of March and for WSU it was the 31st of March.

Results

We found no difference in key outcomes between those who had access to Grok and those who did not. We do not know if this is because of the app content or due to low app usage.

There were no differences in academic performance or unit completion rates

To assess the impact of Grok on academic performance we compared average marks and unit completions between the 'information-only' group and the group with access to Grok. We found no meaningful or statistically significant difference in marks or completion rates.

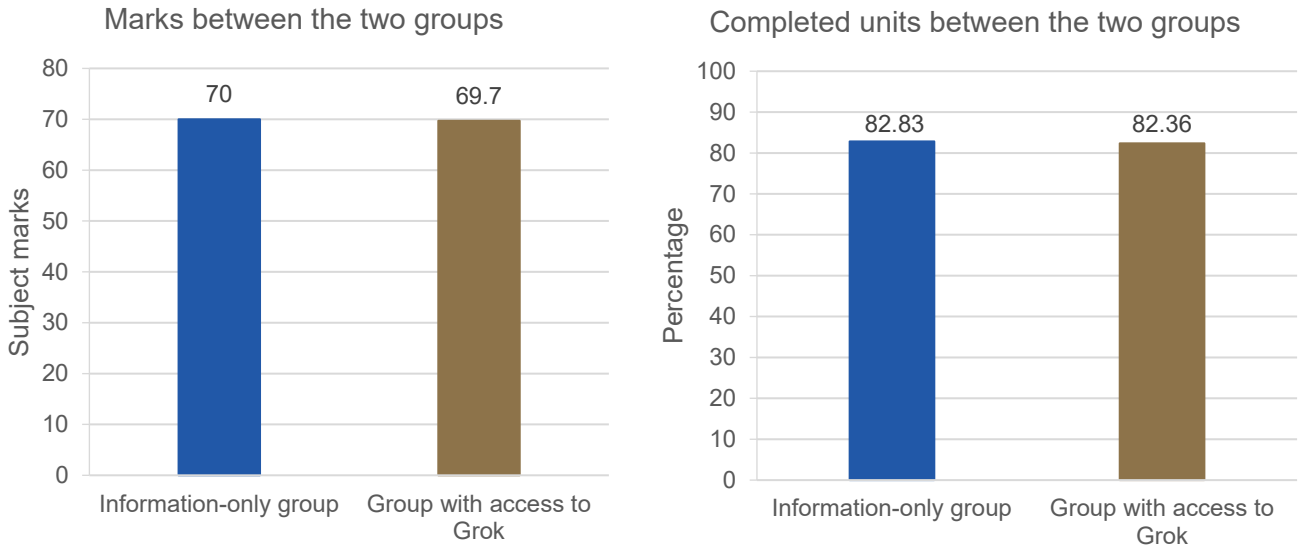


Figure 3. Impact of access to Grok on university marks and completed units

We investigated whether the app was more or less useful for students who were disadvantaged. We compared results across seven markers of disadvantage:

- Aboriginal and/or Torres Strait Islander;
- Attendance type (part-time/ full-time);
- Disability status;
- First in family (where the student is the first in their family to attend university);
- Culturally and linguistically diverse;
- Regional or remote home location; and
- Low socio-economic status.

There were no differences in the outcomes for disadvantaged students and non-disadvantaged students.

The app had no effect on wellbeing or student identification

We measured student wellbeing and student identification in the final survey. We found no evidence showing access to Grok impacted wellbeing or student identification. For the Flourishing Scale, the mean score for the 'information-only' group was 50.4 and for those with access to the Zen garden it was 50.2. On this scale, the maximum score is 56, indicating both groups had many psychological resources. It is possible the intervention was unable to increase these scores any further, as there was limited room for improvement.

On the student identification measures, with a maximum score of 7, means ranged from 4 (ingroup ties) to 4.5 (centrality and ingroup affect). In all three cases, differences between the 'information-only' group and the group with access to the Zen garden were negligible.

There was no relationship between app usage and academic outcomes, but usage was very low

The results above focused on the causal effect of access to the app. We also investigated the association between usage of the app and outcomes. To estimate this we divided people into app users or non-app users. App users were participants who completed at least one app activity. Only 659 participants completed at least one activity. This is about 29% of the treatment group, indicating over 70% of the treatment group did not use the app at all.

When we compared people who used the app with those who did not, usage of the app was not associated with difference in any outcomes. However, even among app users, usage was very low. Of the 659 participants who used the app at all, 481 (73%) completed 10 activities or fewer (Figure 3. - truncated). There were 24 'super users' who completed 50 or more activities in the trial period. This is about one per cent of people in the treatment group.

For app users, we looked at the association between the number of completed activities and marks to determine if there was a relationship between higher usage and outcomes. There was no association, however, at the high usage end there were very few participants.

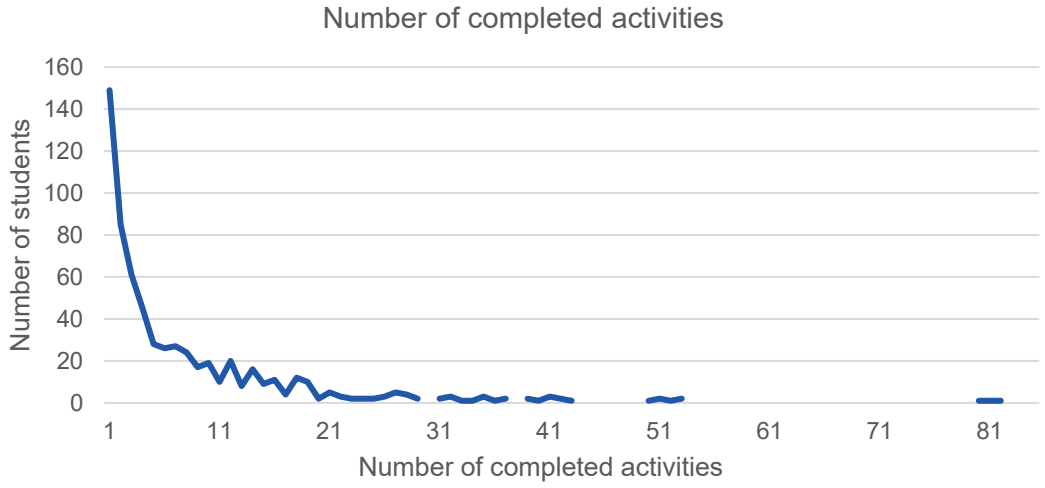


Figure 4. Number of completed activities

When students did engage with Grok, the feedback was largely positive

In the final survey, 317 students in the treatment group wrote about their Grok experience with open-ended responses. We conducted sentiment analysis where each response was broken down into positive, negative or neutral categories. For example, in response to the question 'What did you think of Grok?', 58% of responses were positive, 26% were negative, and 16 % were neutral. Overall, 15% said the app was helpful and 14% liked the concept, but 10% said the app needed to be more engaging and 8% believed the app was confusing.

We also interviewed 13 users to gain a deeper understanding of how they interacted with Grok. Each highlighted different concepts and activities they saw as beneficial. Suggestions for improvement included increasing gamification and having a wider variety of activities.

Although most Grok experiences were positive…

'Grok felt as though it was made for first-year students, very relevant to the university, almost personalised, with touches that went beyond average advice.'

'What I took away was the concept of separating things into different groups so that when something goes wrong, it is only one part of a bigger picture.'

'The app made me feel like I had more control over my start to university. It was an easy way to keep track of all aspects of my life in the least overwhelming way possible.'

'Grok was an app worth downloading - it helps you stay motivated, reflect on your self-identity/wellbeing and stay on track when it comes to academics.'

'I love the idea of a Zen garden that grows as you achieve milestones. I think it's a great way to improve yourself through triggering the habit loop.'

'It deepened my understanding of the value I place in relationships…It gave me a real ease to have it all visually represented.'

'[Grok] made me want to challenge myself…it made me be a lot more social within my cohort when I wouldn't have spoken to them otherwise.'

…some were negative

'Ticking off completed task was satisfying but eventually some reoccurring task became boring and less engaging. More personalisation is needed for engagement.'

'After using it for a few weeks, I noticed that it was not as useful as I initially believed… it was definitely possible to cheat through your goals and I think that made me see the app as invalid'

'What made this frustrating was the fact every activity took several clicks and was pretty slow'

'It had a lot of dependence on social connections I don't have…that aspect of the app was lost on me'

Limitations

The launch of this trial coincided with the beginning of the COVID-19 pandemic

The trial began in mid-February 2020, prior to the first recorded community transmission in Australia, and before COVID-19 was officially declared a pandemic. During the early stages of the launch, the spread of COVID-19 in Australia escalated and university campuses closed. These events mandated an immediate stop to all in-person marketing. At this point, our team was part way through the recruitment and marketing schedule, and were yet to travel to the largest campuses in Sydney.

The ceasing of in-person marketing and campus visits had an immediate and sustained effect on the recruitment strategy for the trial. This led to a smaller trial sample size than what was needed to detect a significant effect on the key outcome variables.

Face-to-face marketing on campus had been highly effective at getting students to download the app and register for the trial. The BETA recruitment team were able to help students troubleshoot technical difficulties and answer any questions in real time. This helped overcome small frictions and gaps between ‘intention’ and ‘action’ a student might have experienced as a result of registering for the trial on their own.

A lack of marketing materials and recruitment presence on campus during key open day events for new undergraduates may have also impacted the effect of online marketing in later weeks. This is because students would have had less prior exposure or familiarity with the initiative when they later saw online promotional material. Even if BETA had been able to maintain our recruitment plans and met the target sample size, the generalisability of any outcomes from 2020 would have been questionable.

When campuses closed in response to COVID-19, it rapidly changed the traditional landscape for learning, and possibly the expected predictors of non-completion. Predictors such as engagement in university life, indexed by behaviours like participating in on-campus events, attending tutorials or meeting new friends on campus, were suddenly not relevant. Instead, the norm for students became online lectures and tutorials, studying from home and social distancing.

An unforeseen challenge was the need to reflect the rapidly changing COVID-19 restrictions in the app activities

Activities due to be released later in the semester, like those encouraging in-person social interaction and exploration of the university campus, were replaced with COVID-19 safe suggestions (see Appendix 3).

Grok’s primary purpose was to foster social connection behaviours on and off-campus. Although the adapted activities tried to keep the spirit of social connection through the use of technology and social media, these adaptations were a compromise to what had been

originally designed. It is possible this undermined the effectiveness of Grok's activities. Original content had also been user tested extensively with university students to ensure maximum relevance and impact. It was not possible to ensure new content had the same relevance for students.

Due to the timing of the trial's launch with the developing pandemic, it did not become immediately apparent Grok's activities required modification. This meant the activities already released in the first four weeks of semester remained unchanged. It is possible students found the earlier activities irrelevant in the new climate and were limited by restrictions, leading to disengagement. Some users expressed this in the final survey:

> *'I used it frequently during the first few weeks of the university semester, but when COVID-19 restrictions occurred and courses went online, I frequented it less.'*

The pandemic had an overarching impact on both the control and treatment groups. Increases in stress related to COVID-19 uncertainties potentially diluted effects of Grok's wellbeing activities, which were designed with the stressors of a 'typical' university semester in mind. The nationwide requirement to socially distance and minimise travel, highlighted the human need for connectivity in unprecedented ways. A highly publicised topic at the start of the pandemic was the importance of keeping strong social networks and regular contact with vulnerable people. Many articles offered social activities to encourage online connection and self-care. As social connection was a key insight of Grok, it is possible the control and treatment group were exposed to similar social activities outside of the app, masking effects Grok might have had.

Unrelated to COVID-19, there were technical issues which were not anticipated, negatively impacting student engagement

Behavioural insights have shown reminders are effective in keeping intentions active and initiating behavioural change. Due to software updates after the launch, there was an absence of push notifications and only infrequent in-app reminders. This may have led to low ongoing participation in a busy university semester.

Finally, Grok was initially programmed to only allow access to select undergraduate students at the participating universities. A technical update inadvertently gave access to a wider cohort of students who were not our target, such as postgraduate students. The update also allowed 'multiple registration', where eager students attempted to register multiple times with different student numbers to gain access to alternative versions of the app. There were 515 students randomised more than once. To manage this issue our primary analyses were based on intention to treat principles. Please refer to the technical appendices for more information on these principles.

Discussion and Conclusion

University completion rates have remained largely unchanged for approximately 50 years despite many policy changes. Working with a hard-to-shift cohort challenged BETA to design an innovative and novel intervention.

The response to this challenge was a new gamified app for university students called Grok. Grok allowed students to easily **gather** current and new friends around them during difficult times, **grow** essential skills and knowledge for resilience, so they can ultimately **graduate**.

The results indicated having access to Grok had no impact on completion rates or grades at an individual level. The qualitative data from the final survey and interviews suggested Grok had some positive impact on those who completed app activities. This underscores the engagement limitations of the trial but it also suggests the concept shows promise.

The challenge remains how to translate important behavioural theories and psychology frameworks into a contemporary and everyday context. Mobile apps are appealing because of their ubiquitous and accessible nature, particularly for a student cohort. But the app market is crowded and competitive, with a combined total of over 4.83 million apps available for android and apple users (Statista, 2021). As a result, users are overwhelmed with choice and demand high quality functionality.

As demonstrated by these results, apps are easy to download but do not guarantee regular usage or retention. Data shows the average app loses 77% of mobile users within the first three days after download and 90% of mobile users within the first month (Chen, 2015). This is consistent with our experience of most students disengaging with the app content after download. Mobile apps might not be the most appropriate medium when engagement over a long period is desirable.

The behavioural theories underpinning the app have a long history in research and could help student cohorts in a different format. For example, future researchers may want to take the principles of Grok's Zen garden into a pen and paper setting, or incorporate 'key wisdom' as daily reminders on an existing learning platform. Other mechanisms, such as text messages, emails or in-class modules, could also be examined as viable options for student cohorts.

On balance, the likelihood of being able to make a measurable impact in the short trial timeframe was probably small. The barriers experienced by disadvantaged students are complex and engrained. Any positive benefits of an intervention of this nature may have warranted a longer period of evaluation. It may be beneficial for future researchers to explore longitudinal designs allowing more time for any impacts.

It is also difficult to gauge whether students who were struggling with university downloaded Grok. It is possible we engaged with high achievers who wanted to be involved in campus life, but missed the students who were already disengaged. Other researchers could try combining the app with support programs already in place, or encourage app download after a subject failure. A multi-pronged approach, reaching more students through multiple avenues, might help to capture more at-risk students before non-completion.

Now Grok has been designed and built, there are several avenues of further exploration. One option is to make Grok accessible to other researchers. Since the conclusion of the trial, there has been interest in Grok from other universities in trialling the app with scope for customisation. Future research could help understand the extent to which COVID-19 led to non-significant results in the current trial. Other researchers could also explore the minimum dosage requirements for a student to benefit from using Grok or similar approaches.

The fundamental behavioural insights underlying the design of Grok are also relevant to cohorts beyond university students including high-school students, new migrants or the elderly. Grok may have a role in helping behavioural practitioners, policy makers and researchers with similar behavioural problems across a range of social contexts. Researchers interested in any aspect of Grok should contact BETA.

Appendices

Appendix 1 - Flourishing scale

The Flourishing Scale is an 8-item summary measure of one's self-perceived success in important areas of life, contributing to an overall psychological well-being score. The scale was designed by Diener et al. (2010).

Below are 8 statements with which you may agree or disagree. Using the 1–7 scale (1 – Strongly disagree, 7 – Strongly agree), indicate your agreement with each item by indicating that response for each statement.

I lead a purposeful and meaningful life.

My social relationships are supportive and rewarding.

I am engaged and interested in my daily activities

I actively contribute to the happiness and well-being of others

I am competent and capable in the activities that are important to me

I am a good person and live a good life

I am optimistic about my future

People respect me

Scoring: Add the responses, varying from 1 to 7, for all eight items. The possible range of scores is from 8 (lowest possible) to 56 (highest PWB possible). A high score represents a person with many psychological resources and strengths.

Appendix 2 - Group identification scale

The three dimensional strength of identification scale is a 12-item measure of one's perceived identification with a particular ingroup. The three dimensions include centrality (how self-defining the group is to one's identity), ingroup affect (whether one sees the group in a positive or negative light) and ingroup ties (how connected one feels to other members of the group).

Below are 12 statements with which you may agree or disagree. Using the 1–7 scale (1 – Strongly disagree, 7 – Strongly agree), indicate your agreement with each item by indicating that response for each statement.

Centrality
I often think about being an (ingroup member).
Being an (ingroup member) has little to do with how I feel about myself in general.
Being an (ingroup member) is an important part of my self image.
The fact I am an (ingroup member) rarely enters my mind.

Ingroup Affect
In general I'm glad to be an (ingroup member).
I often regret being an (ingroup member).
Generally I feel good about myself when I think about being an (ingroup member).
I don't feel good about being an (ingroup member).

Ingroup Ties
I have a lot in common with other (ingroup members).
I feel strong ties to other (ingroup members).
I find it difficult to form a bond with other (ingroup members).
I don't feel a strong sense of being connected to (ingroup members).

Appendix 3 - Updated COVID-19 safe app activities

The Grok activities were evaluated in user testing over the course of 2019 and finalised at the start of 2020. Many of the activities encouraged students to engage in on-campus experiences and meet new people at university to increase feelings of belonging. Activities also focused on organising group events or spending one on one time with people as a way of grow existing connections. Other activities encouraged students to explore their surroundings and enjoy nature, as research shows just being outside can improve wellbeing and mood (Tyrväinen, Ojala, Korpela, Lanki, Tsunetsugu, & Kagawa, 2014).

As COVID-19 restrictions were announced and continued to change throughout semester one, the original Grok activities were rapidly updated. With university campuses closed, public venues shut and social distancing practices in place, many Grok activities needed to shift focus. Some of the original activities only required slight changes, but others needed to be completely rewritten. Below is a selection of original Grok activities and the matching activity, updated to be in line with COVID-19 safe practices.

Table 1. Comparison of original and updated wellbeing activities

Original activities (pre COVID-19)	Updated activities (COVID-19 safe)
• **Find a new favourite**: Go to a café on campus you have never been before and order something that is not your usual order. You might just find a new favourite! A list of campus cafés can be found here.	• **Be 100% present:** Have you heard of mindfulness? The idea is to be 100% present in the moment to better manage unwanted thoughts. Try one mindfulness activity this week. Find some great examples on the Headspace website.
• **Get your nature fix:** Research suggests that even 15 minutes of sitting or walking outside, like in a park or bush, can make people feel mentally restored. Go soak up some sun or listen to the rain outside for at least 15 minutes today.	• **Get your nature fix:** Research suggests that even 15 minutes of walking or sitting outside can make people feel mentally restored. Whether you decide to sit on your balcony or do some gardening, spend at least 15 minutes outside today
• **Walk a little extra:** Find a way to walk a little extra this week. Park the car further away, take the stairs or enjoy the scenic walking route to class.	• **Improve your posture**: Sitting at a desk all day can strain your muscles. Try some stretching exercises at least twice this week to help improve your posture and focus. There are some great examples here.

Table 2. Comparison of original and updated social activities

Original activities (pre COVID-19)	Updated activities (COVID-19 safe)
• **Explore**: Think of a nice place (like a cafe, a park or a beach) you'd like to check out and ask someone in your group to go with you. Strength in numbers!	• **Make a new memory**: Sometimes it's fun to remember good times you have spent with someone. But it's also important to make new memories. This week plan something fun with someone and make a new memory. Host an online pizza dinner, house obstacle course or video games tournament. Use your imagination ☺
• **Something eventful:** Invite someone to an upcoming event that you think they will enjoy. Not only do you get to spend quality time with them, but you are also letting them know you are considerate of their interests.	• **Spread the love:** There are a lot of scary, negative news stories circulating online, which can feel overwhelming. But there are also many positives stories of people donating, supporting and helping each other. Find two positive news stories from a trusted, credible site and spread the love! You could post them on social media, share them in a group message or text them to someone.
• **Get a group together**: Organise a group activity that includes someone (or people) from this social group. Group activities can be a great way to reconnect and see many people at once. It could be as simple as grabbing a drink or something a bit different, like an escape room.	• **Have a virtual hang-out**: Organise an online group catch-up. There are heaps of free services you can use, such as the "face-to-face social networking app" Houseparty. Other alternatives include Google Hangout, Skype and Zoom. Try one out this week!

Table 3. Comparison of original and updated academic activities

Original activities (pre COVID-19)	Updated activities (COVID-19 safe)
• **Book a library room**: A great way to complete group work (or study solo) is to book a study room in the library. Go to your library website and book a study room.	• **Set up your space:** Even if you usually prefer to study somewhere like the library, it is important to have a comfortable study space at home or in your room that allows you to be productive. This week, set up your personal study space to be ergonomic and efficient. There is some great advice in this graphic and this 4-minute video.
• **Boost your skills**: Did you know your uni offers ***free*** study sessions with an experienced student, often someone who has done your course before? For more information click here	• **Boost your skills**: Did you know your uni offers ***free*** study sessions with an experienced student, often someone who has done your course before? You can even attend these sessions online. For more information click here
• **Organise a group study session**: Some people can benefit from studying in groups because it gives everyone a chance to think differently. If you have never tried this, see if it works for you. Organise a study session with a friend or a larger group	• **Enlist a friend:** Get a friend or family member to test you on the exam content (either in person or online). It is ok if you make mistakes! Use it as an opportunity to find out what areas you need to focus on.

Appendix 4 - Trial design

This section provides information on the methods underpinning the randomised controlled trial used to evaluate the impact of Grok.

Ethics approval

The project was approved by the University of Newcastle Human Research Ethics Committee (H-2019-0203) and the Western Sydney University Ethics Committee (H13248).

Registration and analysis plan

We registered the trial with both the American Economic Association Social Science Registry (RCT ID no. AEARCTR-0005500) on 25 February 2020 at the commencement of the trial. We also registered on the BETA website 4 March 2020. This was in the recruitment phase of the trial, but before data collection had commenced. There were no changes made to the analysis plan after registration.

Intervention

The intervention tested was a mobile application called Grok. The purpose of Grok was to encourage students to reflect on their social connections to strengthen their confidence and feeling of belonging at university. Additionally, Grok provided a number of activities designed to help student wellbeing, study skills, and knowledge. Grok was designed as a ‘zen’ garden. Completion of activities resulted in growth of plants and characters in the garden. The full details of Grok’s design can be found in the ‘What we did’ section of the report.

Recruitment

The evaluation of Grok was a multisite randomised controlled trial run across two Australian public universities. Western Sydney University is an urban university and the University of Newcastle is regional.

Student engagement teams at the two universities emailed eligible students an invitation to participate in the trial and posted multiple sign-up links on internal university social media pages. An invitation slide was provided to teaching staff to include in their initial lectures and an invitation posted on the student portal. Digital signs and posters were displayed around campus and bookmarks were included in the University of Newcastle orientation bags and Western Sydney University student diaries. BETA members also attended student welcome events on campus and offered small incentives for downloading and registering Grok. In person recruitment was interrupted at the University of Newcastle due to flooding and at Western Sydney University due to the COVID-19 pandemic.

Students who followed the provided sign-up links were taken to either the Apple or Google Play App store depending on their device type. Students were then able to download Grok onto their device. When students first opened Grok, they were asked for their consent to participate in the research project. If consent was provided, then students were asked to provide their student number.

Our pre-analysis plan stated only undergraduates enrolled at Western Sydney University or the University of Newcastle would be eligible to participate in the trial. The universities sent a list of all eligible student numbers approximately weekly for inclusion in the app. After students entered their student number it was checked against this list. If their student number was on the list they were able to enrol in the trial. People who were rejected but believed they were eligible were able to contact BETA for manual override. This happened on 83 occasions typically due to the weekly lag in updates to the list.

Due to technical issues student number verification was turned off mid-trial. Because of this, both international and postgraduate students were included in the final sample. The small number of postgraduate students enrolled in the trial is unlikely to have impacted the findings.

Randomisation

Randomisation took place at the student level. Randomisation was implemented in Grok's Application Programming Interface (API). Once an eligible student successfully downloaded the app, consented to participation in the trial, and entered a valid student number and email address, they were randomly assigned to the treatment group or control group with a 50% chance of being assigned to treatment. Students randomised to the treatment group then had access to Grok's 'zen' garden and associated activities, and those in the control group were provided links to resources.

In total, 4,580 (2,420 from UON, 2160 from WSU) students were randomised. Some of these students withdrew from all units prior to the university census date. We specified that only students enrolled in at least one unit after census date would be included, thus these individuals were removed from the final dataset. The final analytic dataset comprised 4,463 participants.

Power and sample size

These power calculations were originally reported in our pre-analysis plan. Estimates were based on marks for study units undertaken by students at Western Sydney University in semester 2, 2018. To account for the correlation in marks due to individual students taking multiple study units, we clustered our data at the student level when calculating the required sample size. Cluster sizes varied in line with the number of study units a given student undertook in the semester. We estimated the average cluster size as 2.99 study units, with a standard deviation of 1.02 and an intra-cluster correlation coefficient of 0.46 for the completion outcome and 0.68 for the marks outcome.

Subject marks

For the academic performance outcome measure (see Outcomes below for more details), we estimated that we needed 1,842 participants to detect an increase in 2 subject marks (a Cohen's *d* of 0.1) at 80% power (treatment proportion = 0.5, alpha = 0.05, one-tailed test).

Completion

For the completion outcome measure, we estimated that 6,602 participants would be required to detect an increase in the completion rate from approximately 80% to 82% at 80% power (treatment proportion = 0.5, alpha = 0.05, one-tailed test).

Based on these power calculations our aim was to recruit at least 3,000 students at each university for a total study size of 6,000 students. We fell short of this target, and thus our completion outcome was likely underpowered for small effect sizes.

Analysis

For both primary outcome measures (see Outcomes below), we estimated treatment effects and computed p-values and 95% confidence intervals using Ordinary Least Squares (OLS) regression.

Most students had marks and completion indicators for multiple study units. To account for this structure we clustered standard errors at the student level using CR2 standard errors.

We estimated the following specification:

$$Y_{ij} = \alpha + \tau T_i + \beta_1 X_i + \beta_2 X_i T_i + v_i + \omega_{ij}$$

Where Y_{ij} was either a binary variable indicating whether student i completed unit j, or the raw mark received by student i for unit j (out of a possible 100 marks). T_i indicated the treatment status of each student where the coefficient τ gives the average treatment effect, X_i was the set of de-meaned covariates that were collected at baseline, $X_i T_i$ was the interaction of these de-meaned covariates with the treatment indicator, v_i was the student-level error term and ω_{ij} was the study unit-level error term.

The following baseline covariates were included in our primary analysis models: student age (a continuous variable), gender (binary), full-time/part-time status (binary), university (binary), and socioeconomic status (relatively low / relatively high - binary).

We conducted all analyses in R version 4.0.2.

Trial threats

Student withdrawal prior to census

Students usually have until week four to withdraw from units without penalty. Students had access to Grok at the start of the semester, prior to census. To simplify the interpretation of outcomes, we only retained study units that a student was enrolled in after the census date. Students who withdrew from all units prior to census date were excluded from the analysis.

This means that the number of units retained by students after census could have been impacted by the app. However, it seems unlikely that the app could have impacted the choice to drop units prior to census, and if it did occur, it would be rare and unlikely to influence our overall findings.

Multiple randomisations

Our original design minimised the chance of students being randomised more than once due to in-app verification of student numbers. This feature was disabled mid-trial due to technical problems, and as a result, there were 515 students randomised more than once. We found evidence to suggest that some students who were initially assigned to the control group re-enrolled in the trial until they were assigned to treatment. We did not anticipate this issue in our pre-analysis plan. We consulted with members of BETA's Academic Advisory Panel to develop an analytic approach to deal with this issue and implemented their advice.

The implemented approach was to keep our primary analyses as intention-to-treat, as per our pre-analysis plan. However, we took the first randomisation for every student and analysed them in that group regardless of subsequent registration/randomisations.

As a robustness check we re-ran all analyses only including students with a single registration, but also those with multiple registrations where the randomisations were either all control or all treatment. The results from this analysis did not differ from the ITT analysis in a way that would alter our conclusions.

Outcomes

Primary outcome 1 - academic performance

This was measured using subject marks extracted from the university administration system. Marks were expressed as a score out of 100.

Primary outcome 2 - completion

The successful completion of a subject. A full protocol for constructing this outcome from the university administrative data is available in our pre-analysis plan. Both participating universities made changes to their grade systems in response to the COVID pandemic. The modifications were such that if a student failed a unit, that unit did not count towards their grade point average (GPA). This had no impact on the trial, as grades were awarded in the usual manner.

Secondary outcome - subjective wellbeing and student identity

We measured subjective wellbeing with the Flourishing Scale and student identity with the Identity Scale.

Hypotheses

We had two primary hypotheses that we tested to evaluate Grok, these corresponded to our two primary outcomes. We did not adjust hypothesis tests for multiple comparisons, as our outcomes were highly correlated.

Hypothesis one – academic achievement

The average mark in the treatment group will be higher than the average mark in the control group (T > C).

Hypothesis two – completion

The treatment group will have a higher proportion of completed subjects as compared with the control group (T > C).

Both hypotheses were one-sided as we determined that the app was unlikely to reduce marks or completion rates and from an implementation/scaling perspective a null result and a negative result would be treated the same (the app would not be rolled out more broadly).

Missing data

While the completion outcome had no missing data in the final dataset, our academic achievement outcome had 1,867 missing marks (12.2%). Our primary analysis was a complete case analysis. However, as a robustness check, we re-ran analyses involving the academic achievement outcome with an imputed dataset. This dataset was generated using multiple imputation using chained equations (MICE). We assumed that marks data were missing at random with missingness associated with age, gender, attendance type (full time/ part time), university, socio-economic status and whether the unit was completed.

Table 15 in Appendix 5 shows that there were minor differences in equity markers in the imputed dataset (Table 5) and that there was a large difference in marks. This was to be expected, the strongest predictor for missingness was an incomplete unit. Marks were missing because the student withdrew from the unit (effective zero mark) or failed (mark below 50). Therefore imputed marks data are skewed right with higher probability around zero marks and 45 marks.

Despite the difference in the marks distribution in the imputed dataset, this made no meaningful difference to estimated impact of Grok on the academic achievement outcome (Table 16, Appendix 5).

Appendix 5 - Statistical tables

Demographic tables

Table 4 shows the number of students enrolled in the trial and the number of units taken by these students. Tables 5 and 6 provide demographic breakdowns by site and treatment status.

Table 4. Study sample size

	Control	Treatment	Total
University of Newcastle			
Students	1,155	1,200	2,355
Study units	3,891	4,063	7,954
Western Sydney University			
Students	1,068	1,040	2,108
Study units	3,658	3,647	7,305
Total			
Students	2,223	2,240	4,463
Study units	7,549	7,710	15,259

Table 5. Student demographics between sites

Characteristic	UON	WSU	Total
Age (mean)	23.27	23.02	23.15
Gender (% female)	71.63	69.07	70.42
Aboriginal and/or Torres Strait Islander (% Indigenous)	4.63	2.71	3.72
Disability (% with disability)	15.80	7.68	11.92
First in Family (% first in family)	49.34	57.09	51.83
Culturally and Linguistically Diverse (% CALD)	7.22	35.58	20.61
Regional or remote home address (% remote)	18.49	4.94	11.58
SES (% low SES)	25.22	31.93	26.78
Attendance type (% part-time)	10.40	1.33	6.12
Student type (% Domestic)	97.62	90.37	94.20

Note: UON is the University of Newcastle and WSU is Western Sydney University. Denominators may differ due to missing data.

Table 6. Demographic data between trial arms

Characteristic	Control	Treatment
Age (mean)	23.29	23.01
Gender (% female)	68.65	72.19
Aboriginal and/or Torres Strait Islander (% Indigenous)	3.65	3.80
Disability (% with disability)	12.07	11.90
First in Family (% first in family)	52.16	53.67
Culturally and Linguistically Diverse (% CALD)	20.96	20.27
Regional or remote home address (% remote)	12.81	11.97
Low SES (% low SES)	27.22	29.16
Attendance type (% PT)	6.34	5.90
Student type (% Domestic)	93.39	95.00

Note: *A test of joint orthogonality in which the above demographic variables were regressed on a treatment indicator gave the following results: (F (5, 4243) = 2.8922, p <0.04111).*

Primary outcomes

Tables 7 and 8 present the results of the primary analysis. Table 7 is the academic performance outcome (measured using marks) and Table 8 is the completion outcome. These tables correspond to our two primary hypotheses. For more information on how unit completion was defined, see the pre-analysis plan available on the BETA website. As discussed in Appendix 4, some students were randomised multiple times, the analysis presented here is based on a student's first assignment. A robustness check in which students with multiple randomisations were removed is presented in Table 17 and 18.

Table 7. Academic performance

	n	Mean	Effect	95% CI	p-value
Control	5,599	70.0			
Treatment	5,831	69.7	-0.3	(-1.4 – 0.8)	0.281

Note: Covariate-adjusted OLS regression clustered at the student level with cluster robust (CR2) standard errors. The n given in the table is for the number of study units. Study units with missing marks data were removed from this analysis, a robustness check in which these data were imputed is presented in Table 16. P-values are from one-sided tests.

Table 8. Completion

	n	Mean	Effect	95% CI	p-value
Control	7,549	0.1			
Treatment	7,710	0.1	0	(0 – 0)	0.335

Note: Covariate-adjusted OLS regression clustered at the student level with cluster robust (CR2) standard error. The n given in the table is for the number of study units. The completion outcome is binary, we provide a robustness check using logistic regression at Table 19. P-values are from one-sided tests.

Secondary outcomes

The following tables present the analysis of our secondary outcomes. Table 9 presents subjective wellbeing measured using the flourishing scale and Table 10 presents student identity measured using the identity scale. There was no evidence of differences between groups for these outcomes. However, only a subset of individuals in the trial responded to the survey and provided responses to these items. These individuals many not be representative of the broader sample, and the small sample size means we were likely underpowered to detect plausible-sized effects.

Table 9. Subjective wellbeing (flourishing scale)

	n	Mean	Effect	95% CI	p-value
Control	149	50.4			
Treatment	104	50.2	-0.2	(-2.2 – 1.7)	0.411

Note: Analysis was performed with OLS using robust standard error. P-values are from one-sided tests.

Table 10. Student identity (identity scale)

	n	Mean	Effect	95% CI	p-value
Centrality					
Control	458	4.5			
Treatment	346	4.4	0.1	(-0.1 – 0.3)	0.099
Ingroup Affect					
Control	458	4.5			
Treatment	346	4.4	0.1	(-0.1 – 0.3)	0.099
Ingroup Ties					
Control	458	4			
Treatment	346	4	0.0	(-0.2 – 0.2)	0.494

Note: Analysis was performed using OLS regression with a robust standard error. P-values are from one-sided tests.

Subgroups

Table 11 and Table 12 present subgroup analyses for our academic performance and completion outcomes. There was no evidence of any subgroup effects, however, we did not power the study to detect effects in subgroups so these results should be interpreted with caution.

Table 11. Subgroups (academic performance outcome)

Academic performance	Treat – Control difference (95% CI)	p-value	Difference across levels (95% CI)	p-value
Aboriginal and/or Torres Strait Islander status				
Aboriginal and/or Torres Strait Islander	0.9 (-5.8 – 7.6)	0.79		
Non-Indigenous	-0.3 (-1.3 – 0.8)	0.633	1.1 (-5.7 – 7.8)	0.752

Table 11. (cont'd)

Academic performance	Treat – Control difference (95% CI)	p-value	Difference across levels (95% CI)	p-value
Attendance Type				
Full-time	-0.2 (-1.2 – 0.9)	0.757		
Part-time	-2.2 (-7.8 – 3.3)	0.433	-2.3 (-7.9 – 3.4)	0.433
Disability				
Disability	-0.8 (-4.5 – 3.0)	0.689		
No disability	-0.2 (-1.3 – 0.8)	0.663	-0.7 (-4.6 – 3.3)	0.736
First in family				
First in family	0.5 (-1.0 – 2.0)	0.528		
Not the first in family	-0.9 (-2.4 – 0.5)	0.211	1.4 (-0.7 – 3.6)	0.195
Gender				
Female	-0.6 (-1.8 – 0.6)	0.304		
Male	0.3 (-1.8 – 2.5)	0.775	1.0 (-1.6 – 3.5)	0.458
Culturally and linguistically diverse				
Culturally and linguistically diverse	-1.4 (-3.8 – 0.9)	0.22		
English speaking background only	0.1 (-1.1 – 1.2)	0.923	-0.5 (-3.5 – 2.4)	0.725

Table 11. (cont'd)

Academic performance	Treat – Control difference (95% CI)	p-value	Difference across levels (95% CI)	p-value
Regional or remote home location				
Regional or remote	1.3 (-1.6 – 4.3)	0.377		
Not regional or remote	-0.5 (-1.6 – 0.7)	0.438	1.4 (-1.7 – 4.5)	0.361
Socio-economic status				
Low SES	0.0 (-2.1 – 2.2)	0.979		
Not low SES	-0.2 (-1.5 – 1.0)	0.709	0.2 (-2.3 – 2.6)	0.889
University				
WSU	-0.7 (-2.2 – 0.8)	0.379		
UON	0.0 (-1.4 – 1.5)	0.948	-0.6 (-2.7 – 1.5)	0.588

Note: Analysis was performed using OLS regression with a robust standard error.

Table 12. Subgroups (completion outcome)

Completion	Treat – Control difference (95% CI)	p-value	Difference across levels (95% CI)	p-value
Aboriginal and/or Torres Strait Islander status				
Aboriginal and/or Torres Strait Islander	0.0 (-0.1 – 0.1)	0.766		
Non-Indigenous	0.0 (0.0 – 0.0)	0.549	0.0 (-0.1 – 0.1)	0.757

Table 12. (cont'd)

Completion	Treat – Control difference (95% CI)	p-value	Difference across levels (95% CI)	p-value
Attendance Type				
Full-time	0.0 (0.0 – 0.0)	0.636		
Part-time	0.0 (-0.1 – 0.1)	0.609	0.0 (-0.1 – 0.1)	0.602
Disability				
Disability	0.1 (0.0 – 0.1)	0.091		
No disability	0.0 (0.0 – 0.0)	0.874	0.1 (0.0 – 0.1)	0.084
First in family				
First in family	0.0 (0.0 – 0.0)	0.587		
Not the first in family	0.0 (0.0 – 0.0)	0.154	0.0 (-0.1 – 0.0)	0.15
Gender				
Female	0.0 (0.0 – 0.0)	0.134		
Male	0.0 (-0.1 – 0.0)	0.355	0.0 (-0.1 – 0.0)	0.062
Culturally and linguistically diverse				
Culturally and linguistically diverse	0.0 (0.0 – 0.1)	0.51		
English speaking background only	0.0 (0.0 – 0.0)	0.777	0.0 (-0.1 – 0.0)	0.915

Table 12. (cont'd)

Completion	Treat – Control difference (95% CI)	p-value	Difference across levels (95% CI)	p-value
Regional or remote home location				
Regional or remote	0.0 (-0.1 – 0.0)	0.72		
Not regional or remote	0.0 (0.0 – 0.0)	0.594	0.0 (-0.1 – 0.0)	0.711
Socio-economic status				
Low SES	0.0 (0.0 – 0.0)	0.871		
Not low SES	0.0 (0.0 – 0.0)	0.8	0.0 (0.0 – 0.0)	0.837
University				
WSU	0.0 (0.0 – 0.0)	0.202		
UON	0.0 (0.0 – 0.0)	0.679	0.0 (0.0 – 0.1)	0.26

Note: Analysis was performed using OLS regression with a robust standard error.

Exploratory analyses

This section presents analyses designed to further explore the null effects found in our primary analyses. Table 13 compares users who used the app to those who didn't (across both arms of the trial) and found there was no difference in outcomes between these groups. Table 14 shows average marks by the number of activities that users completed (among those who used the app) as a proxy for app usage. There is no apparent relationship between app usage and marks although it is hard to draw definitive conclusions from these analyses given the small number of students that used the app multiple times.

Table 13. App users and non-users

	n	Mean	Effect	95% CI	p-value
Academic performance					
Non-users	11,364	69.7			
App users	2,028	70.0	0.3	(-1.4 – 1.9)	0.732
Completion					
Non-users	12,953	0.2			
App users	2,306	0.2	0.0	(0.0 – 0.0)	0.947

Note: Both outcomes were analysed using OLS regression with a robust standard error.

Table 14. Marks by number of completed activities

No. activities completed	Mean marks	No. activities completed	Mean marks
0	69.86	10	69.16
1	68.24	11	78.88
2	72.47	12	76.09
3	68.49	13	72.61
4	72.58	14	69.40
5	71.19	15	68.29
6	68.67	16	67.82
7	62.53	17	78.50
8	72.89	18	67.09
9	72.06	19	73.25
10	69.16	20+	71.61

Robustness checks

We used multiple imputation to impute missing data for our mark outcome and covariates. Table 15 shows the difference on key imputed variables between the original and imputed datasets. Table 16 presents the analysis of the marks outcome using the imputed dataset. There is no notable difference between the analysis using the imputed dataset and our primary analysis.

Table 15. Comparison of complete cases only and imputed datasets

	Complete cases only	With imputed data
Mark (mean)	69.92	64.72
Equity Score (mean)	1.27	1.29
Aboriginal and/or Torres Strait Islander (%)	0.035	0.035
Disability (%)	0.110	0.110
First in Family (%)	0.514	0.526
Low socio-economic status (%)	0.269	0.291
Regional or remote (%)	0.117	0.120

Table 16. Academic performance – with imputed datasets

	n	Mean	Effect	95% CI	p-value
Control	7,549	70.4			(one-sided)
Treatment	7,710	70.1	-0.3	-	0.293

Note: Covariate-adjusted OLS regression clustered at the student level with cluster robust (CR2) standard errors. The n given in the table is for the number of study units. P-values are from one-sided tests.

In Appendix 4, we discuss an issue that occurred where some individuals were randomised multiple times. We dealt with this in our primary analysis by assigning individuals to the first group they were randomised to. Tables 17 and 18 present a robustness check in which all individuals who were randomised multiple times were removed from the dataset. We found no difference between this analysis and our primary analysis.

Table 17. Academic performance (multiple assignments removed)

	n	Mean	Effect	95% CI	p-value
Control	5,137	69.9			(one-sided)
Treatment	5,645	69.5	-0.4	(-1.5 – 0.7)	0.227

Note: Covariate-adjusted OLS regression clustered at the student level with cluster robust (CR2) standard errors. The n given in the table is for the number of study units. P-values are from one-sided tests.

Table 18. Completion (multiple assignments removed)

	n	Mean	Effect	95% CI	p-value
Control	6,937	0.1			(one-sided)
Treatment	7,447	0.1	0	(0 – 0)	0.389

Note: Covariate-adjusted OLS regression clustered at the student level with cluster robust (CR2) standard error. The n given in the table is for the number of study units. P-values are from one-sided tests.

Finally, because our completion outcome was binary, Table 19 presents the primary analysis computed using logistic regression with an average marginal effect calculation.

Table 19. Completion (logistic regression)

	n	Mean	Effect	95% CI	p-value
Control	7,549	-			(one-sided)
Treatment	7,710	-	0	(0 – 0)	0.186

References

Cameron, J.E. (2004) A three-factor model of social identity. *Self and Identity, 3*, 239-262. DOI: 10.1080/13576500444000047

Chen, A. (2015). *New data shows losing 80% of mobile users is normal, and why the best apps do better*. https://andrewchen.co/new-data-shows-why-losing-80-of-your-mobile-users-is-normal-and-that-the-best-apps-do-much-better/

Cherastidtham, I., Norton, A. & Mackey, W. (2018). *Univeristy attrition: what helps and what hinders university completion?* Grattan Institue. https://grattan.edu.au/wp-content/uploads/2018/04/University-attrition-background.pdf

Department of Education, Skills and Employment (DESE). (2018). *Higher Education Statistics uCube*. http://highereducationstatistics.education.gov.au/

Department of Education, Skills and Employment (DESE). (2017). *Final Report - Improving retention, completion and success in higher education.* https://docs.education.gov.au/system/files/doc/other/final_report_for_publishing.pdf

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social indicators research*, *97*(2), 143-156. DOI: 10.1007/s11205-009-9493-y

Epton, T., Harris, P. R., Kane, R., van Koningsbruggen, G. M., & Sheeran, P. (2014). The impact of self-affirmation on health-behavior change: A meta-analysis. *Health Psychology, 34*(3), 187-196. DOI: 10.1037/hea0000116

Griffith, J., Steptoe, A., & Cropley, M. (1999). An investigation of coping strategies associated with job stress in teachers. *British Journal of Educational Psychology, 69*(4), 517-531. DOI: 10.1348/000709999157879

Hamari, J., Koivisto, J., & Sarsa, H. (2014, Jan 6-9). *Does gamification work? A literature review of empirical studies on gamification*. 2014 47th Hawaii international conference on system sciences (pp. 3025-3034). DOI: 10.1109/HICSS.2014.377

Iqbal, M. (2020). *Business of Apps: App Download and Usage Statistics*. Updated 30/10/2020. https://www.businessofapps.com/data/app-statistics/#3.1

Norton, A., Cherastidtham, I., & Mackey, W. (2018). *Dropping out: the benefits and costs of trying university*. Grattan Institute. https://grattan.edu.au/report/dropping-out/

Meeker, M. (2019). *Bond: Internet trends 2019*. https://www.bondcap.com/report/itr19/#view/3

Pechenkina, E., Laurence, D., Oates, G., Eldridge, D., & Hunter, D. (2017). Using a gamified mobile app to increase student engagement, retention and academic achievement. *International Journal of Educational Technology in Higher Education, 14*(1), 1-12. DOI: 10.1186/s41239-017-0069-7

Sherman D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., ... & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: how self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology, 104*(4), 59-618. DOI: 10.1037/a0031495

Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying values-affirmation interventions: Writing about social belonging is a key to buffering against identity threat. *Personality and Social Psychology Bulletin, 39*(5), 663-676. DOI: 10.1177/0146167213480816

Statista Research Department. (2021). *Number of apps avaliable in leading app stores as of 3rd quarter 2020.* https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

Tyrväinen, L., Ojala, A., Korpela, K., Lanki, T., Tsunetsugu, Y., & Kagawa, T. (2014). The influence of urban green environments on stress relief measures: A field experiment. *Journal of environmental psychology*, *38*, 1-9. DOI: 10.1016/j.jenvp.2013.12.005

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*(9), 1132-1139. DOI: 10.1111/j.1467-9280.2009.02417

ISBN 978-1-925364-55-2 Gather. Grow. Graduate. Using behavioural insights and technology to help students graduate from university (Online) (DOCX)

ISBN 978-1-925364-55-2 Gather. Grow. Graduate. Using behavioural insights and technology to help students graduate from university (Online) (PDF)

Australian Government

BETA

Behavioural Economics Team of the Australian Government

General enquiries beta@pmc.gov.au

Media enquiries media@pmc.gov.au

Find out more www.pmc.gov.au/beta