Summary report:

Piloting methods to develop better evidence on student support

November 2023

CONTENTS

1. INTRODUCTION

TASO's aim is to identify, collate and share high-quality evidence on what works to widen access and improve student outcomes in higher education. This means that we want to develop a better understanding of which activities cause better outcomes for students from disadvantaged and underrepresented groups.

In 2021 we launched two new research themes following a sector consultation and prioritisation exercise:

- Mental health and disability

- Employment and employability

Since then TASO has commissioned evidence reviews under both themes to identify evidence on which approaches are most effective at addressing inequalities. Our report on 'What works to reduce equality gaps for disabled students' suggests there is a need for more and better evaluation of interventions to address inequalities for disabled students in HE, and also identified a number of evaluation challenges specific to this group. Similarly, our report on 'What works to reduce equality gaps in employment and employability' also found the evidence base is relatively weak in terms of that which can help us understand impact, and particularly when it comes to students from disadvantaged or underrepresented backgrounds.

To build the evidence base, in May 2022 TASO put out a call for higher education providers (HEPs) to submit case studies of interventions they are delivering in these areas which showed evidence of promise and would benefit from robust evaluation. As a result of this call, four HEPs were partnered with an independent evaluation organisation to conduct pilot evaluations of their interventions. These evaluations aimed to generate Type 2 evidence and scope out the feasibility of Type 3 evaluation, based on the Office for Students standards of evidence.

Type 1 – Narrative: there is a clear narrative for why an activity may be effective, and this is often based on findings of other research or evaluation.

Type 2 – Empirical Enquiry: data suggests that an activity is associated with better outcomes for students.

Type 3 – Causality: methods are used which demonstrate that an activity has a 'causal impact' on outcomes for students.

The overarching aim of the project was to learn more about the approaches which could be used to develop these sorts of evidence by piloting evaluation methods which are less commonly used in the sector. By cascading the lessons learnt from these pilots to a wider audience, TASO hopes to stimulate more and better evaluation of similar interventions which can help us build a stronger evidence base on 'what works.'
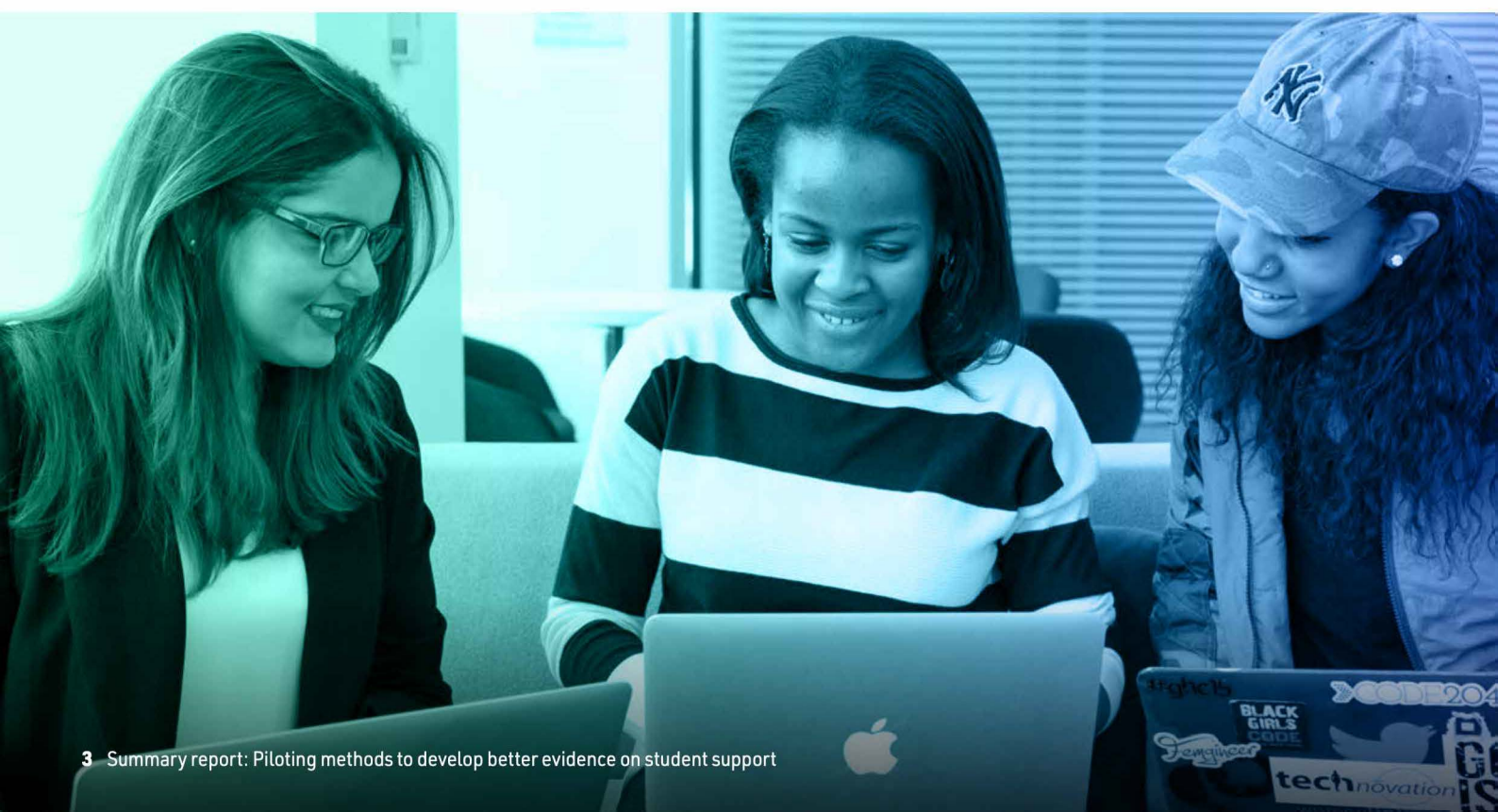
2. EVALUATION PROJECTS UNDERTAKEN

TASO funded four HEPs to participate in this project; University of Exeter (evaluating their Access to Internships – or A2i – programme); the University of Brighton (evaluating their mentoring programmes); London School of Economics (LSE) (evaluating disabled students career appointments); and University of Central Lancashire (UCLan) (evaluating their Student Wellbeing Ambassadors – or SWA – programme). These HEPs were paired with independent evaluation organisations (the University of Cambridge or SQW) to help design and run the evaluations.

All of the evaluations took a mixed-methods approach, combining qualitative and quantitative data. As per the aims of the project, every provider tested out an evaluation method which is generally under-utilised in the sector, resulting in three of the HEPs conducting a quasi-experimental evaluation using Propensity Score Matching (PSM) and one HEP running a Contribution Analysis.

A summary of the evaluation approaches and findings for each HEP are presented in Table 1 on the following page. More detailed findings and recommendations from the independent evaluation organisations can be found in the accompanying Analysis Reports.

Table 1: A summary of the evaluation approaches and findings for each HEP

University / Independent evaluation organisation	Intervention evaluated	Evaluation approach being tested	Key findings
University of Exeter / SQW	**Access to Internships (A2i)** A2i provides funding for up to 140 hours of work, and covers the costs to employers of employing an intern. A2i is a targeted intervention to address inequalities in graduate outcomes and therefore, to be eligible, students must be from a WP background.	Contribution Analysis	Based on insights from survey responses from interns and employers, interviews with interns and university staff, and regression analysis of student outcomes the Contribution Analysis suggests it is likely the availability of A2i has generated additional internships that would not otherwise have taken place. The regression analysis indicates that the scheme is associated with improved employability outcomes for its participants over and above:, i), students participating in other internships at six months after graduation and, ii), students not participating in any internships. Interns are motivated to engage in A2i for a range of reasons, including a desire to gain experience, knowledge and connections within a specific occupation and to improve their job prospects. Interns and employers feel the scheme is well run and delivered.
University of Brighton / SQW	**Mentoring programmes** Mentoring programmes match students with trained volunteer professionals who support the students in working towards their goals and objectives. The aims of the intervention are to support students to succeed on their course and after they leave the university.	PSM	Mentoring is not statistically significantly associated with continuation, degree attainment or graduate outcomes. Mentoring is statistically significantly associated with an increased likelihood of progression in the year that a student undertook mentoring. This effect persists in the first and second course years. Mentees are motivated to engage in the mentoring programmes for a range of reasons, including a desire to access career guidance and advice, develop existing or new skills, and sometimes to overcome loneliness. Mentees mostly felt they had achieved their goals and were generally satisfied with the structure and content of their mentoring sessions. Feedback on mentors was positive, suggesting that the matching process was working well.
LSE / University of Cambridge	**Disabled students career appointments** LSE offer bespoke career support to disabled students. These bespoke, individual, career appointments look to support disabled students' transition into employment in a way that develops their understanding of their strengths and the range of support and accommodations that they may want to seek from potential employers beyond graduation.	PSM	The process evaluation offers evidence in support of most of the ways in which the career appointments were expected to work. Disabled students welcomed and saw the bespoke career appointments as useful. There was evidence that the career appointments could lead to the desired changes around increased understanding and acceptance of students' own disability, including gaining confidence to disclose their disability and advocate for themselves, and increased confidence to request adjustments. Respondents also offered evidence that the appointments supported them in navigating job applications, applying for more jobs than they would have otherwise, being better prepared for recruitment processes, and generally improving their self-belief. The impact evaluation explored employment-related outcomes, including employment, earnings, and general job satisfaction. Due to small sample sizes, the analysis was not able to identify any statistically significant effects of the bespoke career appointments on the above outcomes.

University / Independent evaluation organisation	Intervention evaluated	Evaluation approach being tested	Key findings
UCLan / University of Cambridge	**Student Wellbeing Ambassadors (SWA)** The SWA Programme aims to improve the on-campus experiences of disabled students by providing them with appropriate tailored support, starting with first building an awareness of what this support would look like, then offering this support to enrolled students, enabling them to navigate campus and academic life more straightforwardly; and equipping ambassadors, who take one the paid support role and receive training, with better disability awareness and employment-relevant skills.	PSM	Compared to similar disabled students not engaged with the SWA Programme, disabled students who did engage performed as well in terms of their continuation into the second year of their degree, degree completion, and, for completers, their probability of securing a high (1st class or 2:1) degree classification. Also compared to similar students, students engaged with the SWA Programme as ambassadors were more likely to continue into their second year and more likely to complete their degrees; and performed as well as the comparison group in relation to their probability of securing a high degree classification. Support beneficiaries and ambassadors both reflected positively on their engagement with the programme: the former noted increased confidence and access that supported an overall positive university experience, and practical support that enabled them to engage more fully with campus life; the latter highlighting building skills of likely relevance to their future work and increasing their awareness of and empathy for their disabled peers.

These evaluations provide rich case studies of how interventions to support disabled students, and those designed to improve employability/employment can be conducted. Commentary in the Analysis Reports reveals how the findings will impact each HEP. However, the key output is the overarching lessons we can learn from the project. We've harvested these lessons from the recommendations in each Analysis Report and reflections provided by the HEPs themselves. These fall into lessons learnt about evaluation methods, and other lessons from the project, which are covered in turn below.
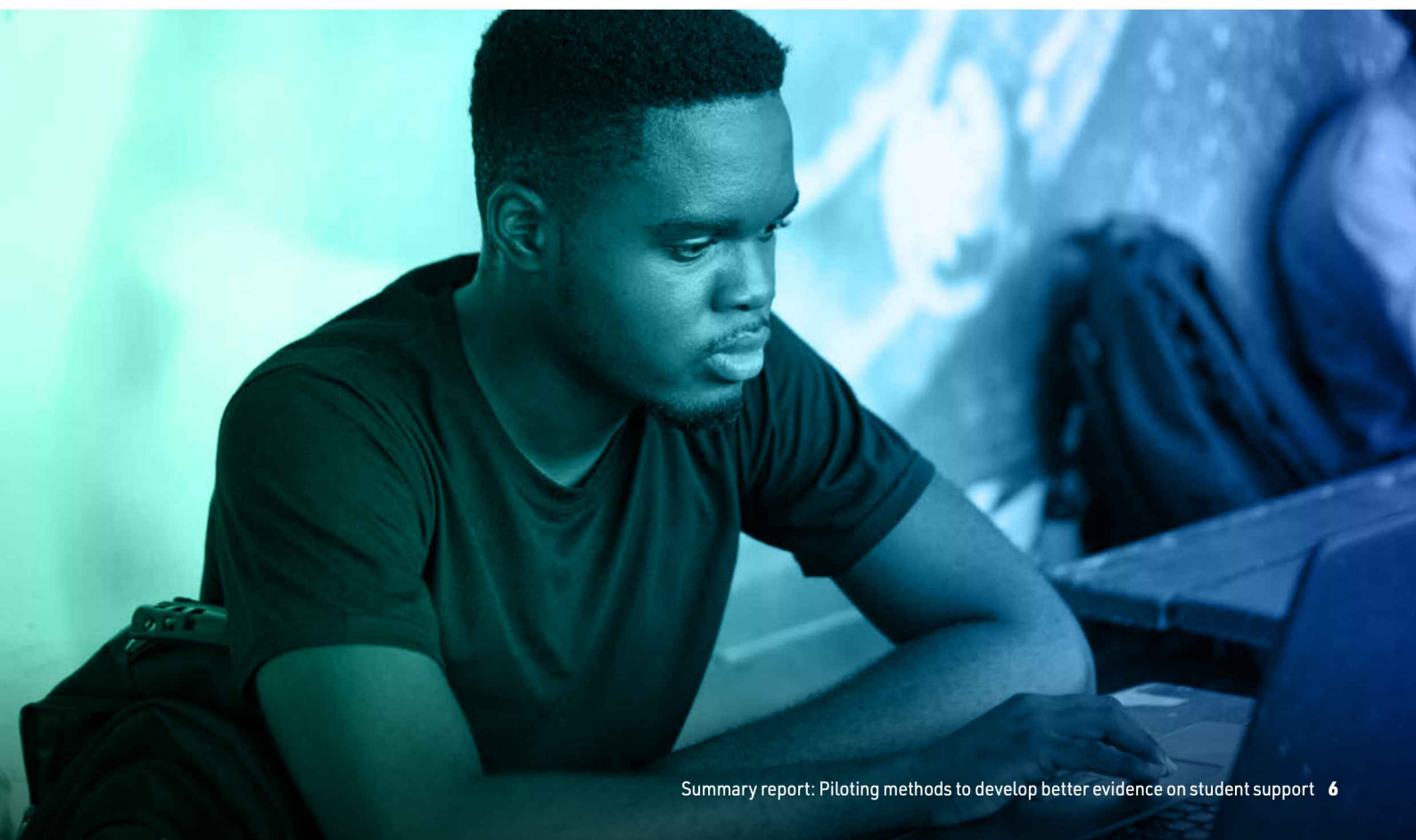
3. LESSONS LEARNT ABOUT EVALUATION METHODS

METHODS WHICH WERE TESTED

Propensity Score Matching (PSM)

What is this approach?

Three of the evaluations piloted PSM using as a method of generating Type 3 evidence. PSM is a quasi-experimental method which matches people who engaged with an intervention (a 'treatment' group) with a group of people who did not engage (a 'comparator' group). The groups are matched based on how likely people are to end up in the treatment group (their 'propensity score'). The aim is to be able to compare quantitative outcomes among the treatment group with a comparator group who would have had a similar probability of participating, based on their observable characteristics (e.g. sex, ethnicity). These sorts of matching techniques are used to try to reduce the problem of selection bias, where students in the treatment group 'look' quite different to the broader student population because there are a range of factors which might make people more (or less) likely to participate. To make valid comparisons between groups of students, we want them to be as similar as possible to one another, in terms of how likely they are to end up experiencing the intervention – PSM tries to achieve this.

Reflections on how this approach worked in practice

The PSM evaluations offer valuable case studies of how this method can be applied. The key strength of all these examples is that they go further than previous evaluations in taking into account pre-existing differences between students who do, and don't, take part in interventions. For example, at the University of Brighton, one measure of the first year continuation gap between mentees and non-mentees narrows from 2.9 percentage points (95.3% versus 92.4%) to 0.7 percentage points (95.7% versus 95.0%) when we restrict the non-mentees to only the matched group. This reveals that, without matching, simple comparison of continuation could overestimate the size of the difference in outcomes between the two groups because we are not taking into account how they might differ from each other in terms of factors like motivation or need to participate in the intervention.

The PSM approach has identified some evidence of impact; for example, there is some evidence for a positive effect of the University of Brighton's mentoring programme on progression (but not other outcomes) and the SWA programme at UCLan is linked to a positive, statistically significant impact for students acting as ambassador in terms of continuation into the second year and degree completion (but not for degree classification). For the beneficiary analysis (i.e. the impact on disabled students themselves), the results point to no statistically significant results for any of the three outcome measures.

It is important to interpret these results in the context of several considerations and limitations. The key limitation here is that PSM is only able to match students based on observable characteristics recorded in the available datasets. Differences in unobservable characteristics, such as motivation, are unaccounted for in the match, meaning that the method cannot entirely account for the potential self-selection bias in relation to participation in the intervention. The impact of this on evaluation outcomes is demonstrated by an example taken from the UCLan Analysis Report:

"For beneficiaries, engagement with the SWA Programme is predicated both on the disclosure of disability and that disability falling within the scope of what the Programme can support...Disabled students who do not disclose a disability therefore cannot access the Programme (as they are not known). Even if they have disclosed, disabled students' engagement with the Programme is further contingent on how their needs align with the availability of support from the Programme."

Therefore, if we simply compare outcomes for people who do and don't participate in SWA, even if we match them on what we know about their existing disabilities (for which the data is imperfect) there are likely to be reasons that some students are less likely to participate which we cannot pick up in the data.

Understanding the reasons students might end up in the 'treatment' group for any particular intervention is key to understanding the limitations of PSM and interpreting any analysis. Because uptake of the intervention is likely to differ systematically by unobserved characteristics in each of the evaluations in this project, this presents an important caveat to the findings across the piece.

It is also important to note that comparable outcomes between the treatment and comparator groups suggests that those students who participate in the intervention are as likely to achieve the outcome as other students. This is not always a bad thing – particularly if those in the treatment are those most in need of support who may, in the absence of intervention, have worse outcomes than those in the comparator group and our intervention is intended to have a compensatory effect. Ideally, our matching approach would offer us comparable groups which would negate this issue, but the limitations of quasi-experimental analysis means we cannot always be confident that this has happened.

Issues around sample size are also a clear recurring theme in these evaluations. As noted in the LSE evaluation, due to small sample sizes, the analysis was not able to identify any statistically significant effects of the bespoke career appointments on the above outcomes. Lower than hoped sample sizes also present issues in the other evaluations for some outcomes.

Small sample sizes are problematic before they limit the precision of any impact estimate and because they may exacerbate self-selection bias.

Finally, the choice of outcome measure emerges as an important consideration for these evaluations. These evaluations focused on robust, albeit distal, outcomes including continuation, progression, attainment and graduate outcomes. These sorts of outcomes are likely to be subject to influence from a wide range of factors beyond the sorts of interventions tested, particularly those which are recorded months or years after interventions take place. Exploring other, more proximal outcome measures which align with interventions' Theories of Change may be fruitful. Within these parameters, choosing outcome data is available for the largest possible sample of students is also helpful.

In sum, PSM offers some promise as an evaluation method which can help the sector move towards a better understanding of causal impact (and Type 3 evidence). However, sample size is a key consideration. Early estimation of sample sizes and preliminary calculations to ensure the analysis will be informative are necessary to ensure that the method is appropriate for the intervention in question. Understanding the reasons why students do, and do not, participate is key to assessing how appropriate PSM is for evaluating specific interventions.
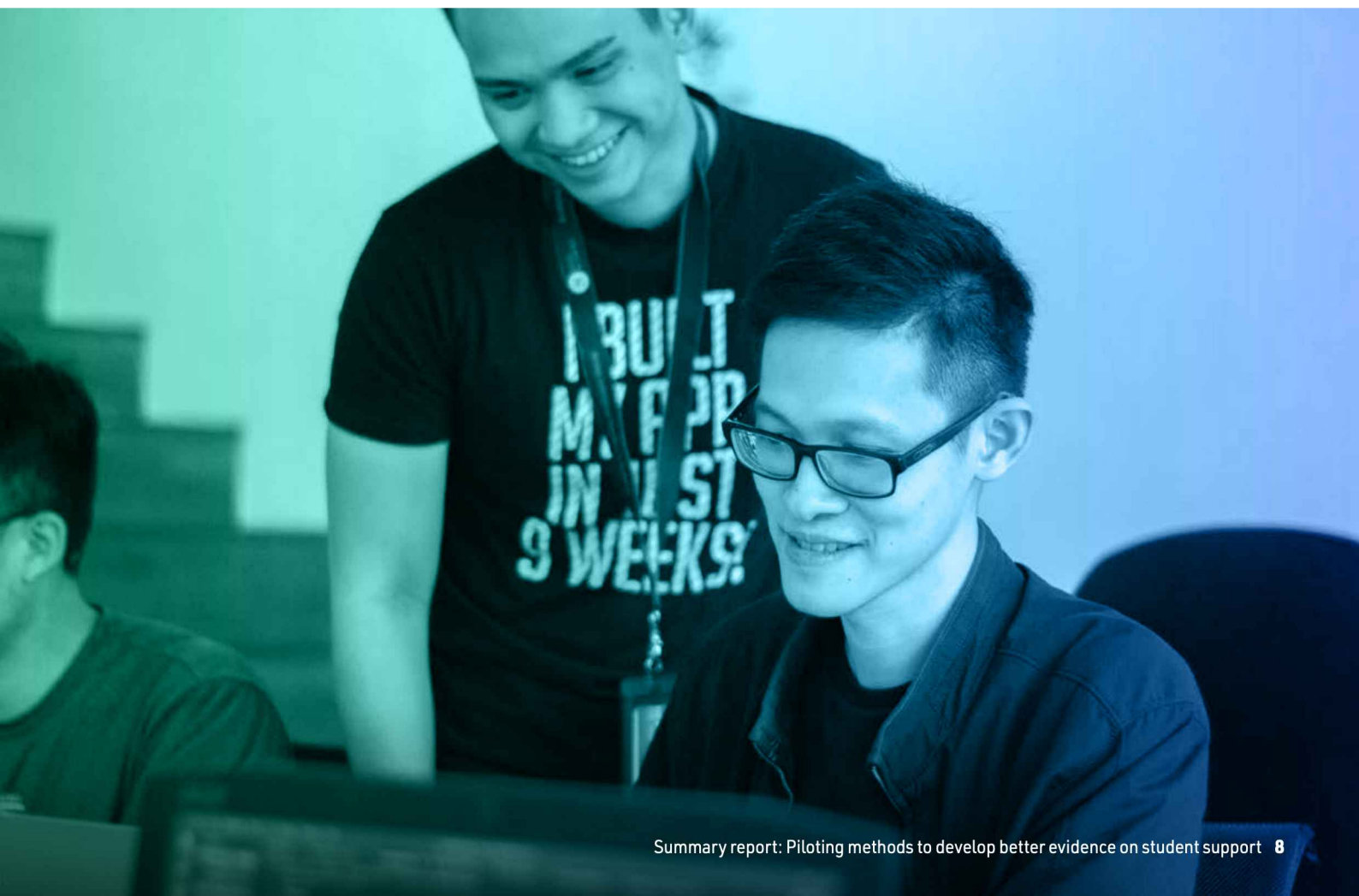
Contribution analysis

What is this approach?

Contribution Analysis explores whether impact can be attributed to an intervention through assessing the contribution that an intervention is making to observed results. To say that an intervention is having an impact, four conditions must be satisfied: plausibility (the intervention is based on a reasoned Theory of Change); fidelity (the intervention is implemented as intended); the Theory of Change is verified (evidence suggests the chain of expected events occurred); and other influencing factors are accounted for (the factors influencing the intervention have been assessed and either shown not to have made a significant contribution or, if they did, the relative contribution is recognised). Contribution Analysis entails a number of steps to assemble and assess evidence against the intervention Theory of Change.

Reflections on how this approach worked in practice

The University of Exeter Analysis Report provides a worked example of a Contribution Analysis to evaluate the A2i internships scheme. It demonstrates the framework which this method provides for triangulating different sorts of quantitative and qualitative evidence to assess whether an intervention's Theory of Change holds true, and concludes there is modest evidence that the A2i intervention generates additional internships.

Broader lessons for the application of this method can be drawn from the evaluation. The HEP partner reflected positively on the mixed-methods approach and the opportunity to use qualitative insights to underpin the Theory of Change, given the many unobservable factors that can influence employment outcomes. However, as for the PSM reported above, small sample sizes can be an issue. In the case of the pilot evaluation, one survey was characterised by lower than hoped response rates. We cannot, by definition, know how non-respondents would feel about an intervention but it is likely that those who are motivated to engage in an evaluation could feel more, or less, positive about an intervention than the average student. Therefore, low sample sizes can bias findings.

The pilot Contribution Analysis here also drew on quantitative analysis to identify a correlation between participation in A2i and improved employability outcomes. However, as noted in the Analysis Report, this does not provide causal evidence of impact due to selection bias (i.e. those who take part in the internships might differ from those who don't, or who take part in different sorts of work experience, on factors we can't observe like motivation or ambition). This issue is likely to be widespread in evaluation of interventions like A2i which combine eligibility criteria with a requirement that students apply to participate.

Both of the limitations above – possible bias and the inability of basic correlational analysis to provide causal evidence – underscore a key consideration for those undertaking Contribution Analysis, which is that it is only as strong as the evidence which goes into it. Each constituent piece of evidence which is assembled must be high-quality to paint a convincing picture overall.

It is also important to note that a great breadth and depth of rich evidence may also be needed to interrogate and underpin every element of an intervention's Theory of Change. The case study presented here highlights the many additional sources of evidence which would ideally be assessed to paint a fuller picture of the contribution of the intervention. It is possible that an evaluator working on this sort of evaluation needs a deeper knowledge of the programme and the context within which it is being implemented than might typically be required in a traditional, impact evaluation, like PSM. For this reason, Contribution Analysis could be just as time-consuming and resource-intensive as data collection in a traditional impact evaluation.

It is also important to note that, in contrast to the PSM examples outlined above, Contribution Analysis cannot quantify the size of an intervention's impact; this may make it difficult to assess the relative effectiveness of interventions evaluated using this method – a key consideration for those seeking to prioritise Access and Participation Plan (APP) programmes for scale-up or improvement.

In sum, the example provided by this project suggests Contribution Analysis can be carried out on employability/employment interventions. This method is well-suited to unpicking the complexity of how interventions might be linked to outcomes, but this example has not provided evidence which would constitute Type 3 causal evidence. TASO is also piloting Contribution Analysis as part of another project testing 'small-n' evaluation methods and will provide a report with further guidance on this topic later in 2023.

Other methods which were scoped out in the evaluations

A key issue emerging in all the pilots was the inability to fully account for selection bias, meaning that those experiencing the intervention and comparison groups are not fully comparable. This issue limits the extent to which these studies can provide strong causal evidence. Other evaluation challenges were also identified. Each of the Analysis Reports contains reflections on alternative evaluation methods which could be used to address these issues, as summarised below.

Randomised controlled trials (RCTs)

A randomised controlled trial (RCT) is a well-established research method which can provide strong causal evidence. In an RCT, students who apply to receive an intervention (e.g. mentoring) are randomly selected to receive a spot (the 'treatment' group) or not (the 'control' group), then outcomes for the two groups are measured and compared. Some scenarios in which an RCT may be most easily embedded in existing programmes are outlined below:

- **If an intervention is oversubscribed.** Limited capacity may offer a natural opportunity to embed a randomised approach, by randomly allocating eligible applicants to receive an intervention. Outcomes are then monitored among this group, and among students who did not get a place because there was not a spot for them. In this scenario, it must be noted that students who applied unsuccessfully could find alternative support elsewhere and this would need to be taken into account in any subsequent analysis.

- **If the intervention can be offered to some students at a later date.** Rather than all students accessing the intervention at once in a 'waitlist' design, students are randomly selected to participate at an earlier (the 'treatment group') or later (the 'waitlist control') date. By comparing outcomes at a time point between the two groups participating you can assess impact. However, this requires that the outcomes being tested are hypothesised to occur in the relative short-time, so that they can be captured appropriately for both groups. This approach is generally suitable for testing if there is an impact on short-term outputs or immediate outcomes as per the Theory of Change.

- **Encouragement to participate can be randomly allocated.** In a randomised encouragement design the treatment group receives active encouragement to take part in the intervention, whereas the control group do not. The encouragement can be a small incentive (e.g., an email or a phone call) that reminds people of their eligibility. This approach may be appealing where there is insufficient resource to provide intensive encouragement to all eligible students. The underlying assumption is that active encouragement increases take-up of the intervention. It is then the impact of receiving encouragement that is evaluated (and its indirect effect on take-up), rather than the direct impact of the intervention itself.

Quasi-experimental approaches

In settings where relevant data is available prior to the start of intervention it may be possible to use pre-intervention cohort(s) as a 'baseline' and apply a difference-in-differences approach. This could also apply when new interventions are being designed: ensuring relevant data exists for cohorts that, by design, had no opportunity to engage with the intervention, would enable such an approach to be undertaken. This could be further bolstered by matching approaches, ensuring that any year-on-year variation in the student make-up, for instance, is accounted for.

Where interventions are being redeveloped or meaningfully changed HEPs could engage in this process of change in a staggered and systematic manner, altering essential aspects of the intervention one at a time, and evaluating the impact of that specific aspect rather than the full intervention using the same evaluation design as above. This would not result in evidence around the totality of the intervention; however, evaluating (consecutive) iterations of the intervention being redeveloped in this manner would, over time, provide a comprehensive and fine-grained understanding of which aspects of the intervention are (most) effective.

Another option would be to exploit changes in eligibility criteria for interventions to compare outcomes before and after the change in eligibility for particular groups. An example of how this might work in practice is given in the University of Exeter Analysis Report: "a review of A2i scheme documentation showed there have been small changes to the definition of widening

participation (WP) [students] in recent years. For example, students with refugee status, those estranged from family support or the care experienced have only been included in the WP definition since the 2019-20 academic year…If the incidence of internships increases among students belonging to a particular group once they become eligible for the scheme (or conversely, if the incidence decreases once a group is no longer eligible for the scheme), this would provide some evidence for the scheme's additionality." The feasibility of this approach would depend on data availability and sufficiently large sample sizes for analysis.

Other approaches to strengthening evaluation

- Given the importance of understanding the factors which affect whether or not students engage in interventions, a number of the Analysis Reports made recommendations around exploring these factors further. This exploration may provide important context to existing findings, or attempts could be made to account for them in matching methods like PSM. Student motivation or need for support may be such factors. When direct measurement of these factors is not feasible, proxies could be explored; for example, attendance at lectures and seminars could be used to proxy for motivation and module

attainment could be used to proxy for student need. Proxies like these are imperfect but could help address some bias in matching methods.

- A number of the Analysis Reports identified potential for further exploration of 'dosage' effects or whether a student had experienced a high/low intensity version of an intervention (for example, intensity could be measured by number of mentoring sessions engaged in, number of internships completed).

- An approach to increasing sample sizes is to collect multiple years' of data, but noting that this approach faces limitations, particularly if the intervention evolves over time, or the context in which the students are operating faces significant changes (e.g. the COVID-19 pandemic)

- The pilot evaluations speak to the complexity of interventions being evaluated. Across the piece, refining the Theory of Change in detail, alongside any re-design or re-development of the intervention, is key.

- On disability-specific interventions, evaluations may benefit from efforts to expand the extent to which students disclose a disability, so that these students are more readily identifiable in the data and patterns of intervention-uptake, and outcomes, might be better understood.
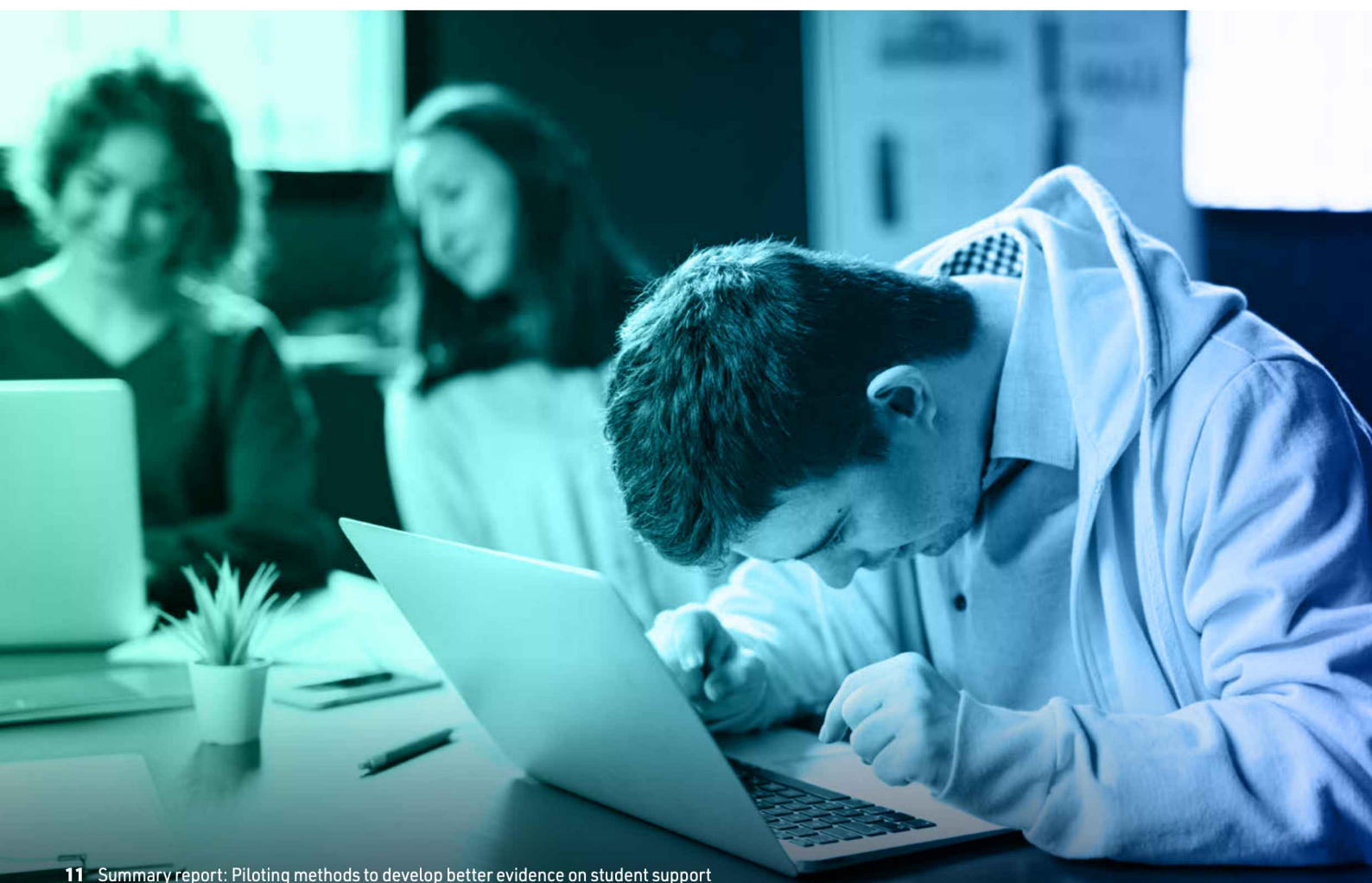
4.   OTHER LESSONS FROM THE PROJECT

The profile and role of evaluation

- HEPs reflected that participating in a TASO-funded project boosted the credibility and profile of evaluation within their institution. This in turn can help foster a wider evaluation culture.

- HEPs appreciated the opportunity to do a 'deep dive' on the evidence for specific interventions. The findings of these evaluations will be used to inform both Access and APP work, but also internal decision-making relating to ongoing provision.

- However, it should be noted that it would be difficult to replicate the level of input which went into these evaluations across all activities covered by APPs. As well as support from a research organisation, funding from TASO facilitated an internal research assistant to work on each of the evaluations, supporting a more intensive evaluation than would be otherwise possible. In the context of limited resources, HEPs reflected that a 'prioritised and proportionate' approach to evaluation is warranted.

Supporting evaluation practice

- The project prompted participating HEPs to put a sharper focus on causality in their evaluations. That is, by encouraging a scoping of more Type 3 methods, it encouraged HEPs to think more deeply about how they can evidence that their interventions are truly **causing** better outcomes for students. The evaluation process spurred HEPs to engage in further reflection on the many complex factors which might underpin intervention uptake, and which can undermine more straightforward analysis of impact.

- HEPs reflected on how lessons from the evaluation approach undertaken in this project would directly inform other evaluations, either directly (i.e. by replicating the same evaluation approach for other interventions) or indirectly (i.e. by adapting evaluation tools, such as surveys and interview schedules, for other projects).

- One HEP characterised this evaluation as a 'jumping-off point' from which to further refine the Theory of Change and explore more detailed evaluation in specific areas. This is in-line with TASO's recommended approach to evaluation as a cyclical process.

- HEPs spoke favourably about being matched with a specialist research organisation, providing opportunities for learning about resources which are available and strengths and limitations of different methods. One HEP characterised this as a "mutual acknowledgement of each other's expertise and the continuous complementary knowledge exchange."

- HEPs identified multiple other ways in which the project had supported their evaluation practice, for example:
  - All the HEPs now have a Theory of Change for their intervention which articulates how and why the intervention is intended to deliver positive outcomes for students. This is a fundamental building block of any evaluation.
  - Compiling large-scale datasets for analysis helped some HEPs get a better understanding of their existing institutional data, standing them in good stead to conduct further in-house analysis.
  - Across the project, process evaluation findings have identified a number of practical recommendations to support continuous improvement of interventions.

Practical evaluation challenges

- Preparing institutional datasets for analysis can be time consuming. Some of the evaluations used administrative datasets which were never intended for evaluation purposes and extensive data 'cleaning' and linking between different databases was required before analysis could take place. This can add significant burden and complexity to a project.

- In some cases survey/interview responses were low. HEPs reflected on the role that incentivisation can play in promoting higher response rates, and the need for best practice in this area.

- This project utilised external research organisations to pioneer methods which are not commonly used in APP evaluation. To continue work of a similar quality beyond the project timeline, HEPs would need to identify or recruit people within the institution with the knowledge and skills to apply these methods.

- The project highlighted a number of challenges specific to working with an independent research organisation. For example, administrative or legal hurdles can limit the speed or scope of data sharing with external parties. More broadly, it can take significant time for parties from different organisations to build collaborative working relationships and a mutual understanding of project requirements.

5. CONCLUSIONS

The overarching aim of this project was to learn more about the approaches which could be used to develop Type 2 and Type 3 evidence by piloting evaluation methods which are less commonly used in the sector. This project has generated four case studies of HEPs trialling methods to evaluate interventions to support disabled students and to improve employment/employability outcomes for students from disadvantaged/ underrepresented groups.

The resulting evaluations provide valuable case studies of how propensity score matching (PSM) and Contribution Analysis might be used in this context. This report has provided reflections on how these approaches worked in practice, and a summary of other evaluation methods which were scoped out as possible improvements and extensions to the existing evaluations.

As noted by one of the participating HEPs, participating in an evaluation of a project which you work on is "both scary and a privilege". We are grateful to the HEPs for participating in these pilots to help us test and refine the methods for use in the wider sector. We hope these recommendations and reports serve as a stimulus for continued use of more advanced evaluation practice on APP activity and beyond.

The Enhanced Theory of Change for strand two displays the barriers and solutions for students from WP backgrounds converting their intention to complete a sandwich course into a successful placement and completion of a sandwich course.

Naturally, there is overlap between the two Enhanced Theories of Change and stakeholders aiming to evaluate and improve their sandwich course provision for WP students should consider both, using a 'pick and mix' approach to build their own Enhanced Theory of Change unique to their context.

6. RECOMMENDATIONS

Recommendations on evaluation methods

- Propensity score matching (PSM) offers some promise as an evaluation method which can help the sector move towards a better understanding of causal impact (and Type 3 evidence). However, sample size is a key consideration. Early estimation of sample sizes and preliminary calculations to ensure the analysis will be informative are necessary to ensure that the method is appropriate for the intervention in question. Understanding the reasons why students do, and do not, participate is also key to assessing how appropriate PSM is for evaluating specific interventions.

- Contribution Analysis is well-suited to unpicking the complexity of how interventions might be linked to outcomes, but this pilot has not provided evidence which would constitute Type 3 causal evidence. TASO is also piloting Contribution Analysis as part of another project testing 'small-n' evaluation methods and will provide a report with further guidance on this topic later in 2023.

- HEPs should explore these, and the other, methods discussed in this report as they strengthen their evaluation practice.

Other recommendations

- Conducting evaluations of the scope and calibre detailed in this report is time-consuming and resource-intensive. HEPs should invest in further evaluation capacity to facilitate stronger evaluation practice across their APP portfolio to ensure they are providing students with the best possible support.

- Where evaluation teams do not have internal capacity to deliver analysis of the sort outlined in this report, larger HEPs should explore whether academic colleagues are able to provide support.

- HEPs should work to improve internal data linking and accessibility to facilitate analysis of institutional data using a range of methods. TASO is currently running a project to support this aim and will produce a practical guide to how data infrastructure can be used within and across HE providers to enable tracking and evaluation of activities and outcomes in 2024.

- HEPs should explore the ethical use of participant compensation in their evaluations to improve rates of data collection. TASO provides advice on this topic in our Research Ethics Guidance.