Student Success Evaluation Framework: Determining causality in activities to improve attendance and attainment.

Alexandra De La Torre and Ellen Dowie

Student Success, University of Kent

October 17, 2023

Abstract

This paper provides an overview of the Student Success Evaluation Framework as a mechanism to assess and identify the outcomes and areas of impact of Student Success interventions on students' attainment and engagement. Delivered across academic schools, these interventions are examined in the context of the Student Success Programme, an institutional research and practice initiative to address attainment and continuation differentials at the University of Kent. The paper provides insights into the rationale behind the Student Success Programme, the institutional factors that have influenced the requirement for an evaluation framework of its implementation and delivery, the challenges faced, and journey undertaken so far to develop, pilot and embed this evaluation framework in the mainstreaming of Student Success across the University of Kent academic divisions. To this end, this paper contributes to the higher education endeavours to assess outcomes and impact of initiatives to reduce and address awarding and continuation gaps, by providing a case study and a model of evaluation that combines process and impact evaluation through the application of Theory of Change, mathematical testing and Contribution Analysis as the main pillars of an innovative evaluation model.

Keywords: Evaluation, Success, Outcomes, Theory of Change, Attainment, Contribution Analysis

1. Background

Considering the increasing focus on racial inequities in society, not least with the Black Lives Matter movement and the murders of George Floyd and Breonna Taylor, there is additional scrutiny on the educational disparities that exist for different ethnic groups. One of these is the awarding (or attainment) gap in Higher Education Institutions (HEI's), metricised by the proportional difference in First and 2(I)s awarded compared to the overall number of awards distributed. Another is the continuation gap which is the proportional difference in progress from the first year of study compared to the overall entrants. Current data indicates that the national attainment gap is at 17.4% and for continuation 4.7%, both of which are statistically significant (The Office for Students 2022). Controlling for entrance qualifications does not eradicate these gaps (The Office for Students, 2018); further supporting that the issue is that of race. Given this information, the government of the United Kingdom, through the Office for Students (OfS) as the regulator, endeavours to reduce and eliminate these gaps, for which there are targets.

In England, HEI's are required to respond to the OfS with plans to address their most prevalent gaps and to report on the work that the institutions are doing to reduce, and ultimately eradicate, these gaps. Each

institution is required by the OfS to produce an Access and Participation Plan (APP) that documents the strategy, targets, and evaluation mechanisms to assess the effectiveness of the APP to tackle attainment and continuation differentials.

There are a number of reasons why HEI's are interested in reducing these gaps, some of which are detailed here. Firstly, the morality of having an educational experience that is equitable for all and not advertently or inadvertently benefitting one group over another. Another is the potential for encouraging more students to attend the University by acknowledging and working on greater equality and equity. There is also the opportunity for the critical assessment of teaching and learning practices that ensure the best provision exists, using new research around pedagogy and equality. Since the University of Kent is registered with the OfS under the "Approved (fee cap)" category, it is required to have an approved APP in force for each year that it charges higher fees (The Office for Students, 2018, p. 83) and to take all reasonable steps to comply with the provisions of the plan.

In addition, there is a requirement for evaluation, for which the OfS supplies a guide, with the standards of evidence expected when presenting evaluation outcomes and when making claims of impact through the development of three types of evaluation. These are: Type 1 Narrative Evaluation, using Theory of Change (ToC) as a means for process evaluation; Type 2, Empirical Enquiry evaluation, to assess impact; and Type 3, Causality evaluation to demonstrate a cause-effect (The Office for Students, 2019). To properly account for reductions in these gaps, HEI's are required to provide evidence-based evaluations of interventions that are used; a subsidiary benefit of this is that it creates a repository of effective and efficient changes in the administration of degrees that contribute to narrowing the gaps. Given the size of some of these gaps, finding the most effective way to reduce them is essential. To this end, the Student Success Central Team at the University of Kent have developed an evaluation framework that meets these monitoring and evaluation requirements, through a joint approach of ToC, mathematical testing and Contribution Analysis (CA).

Organisations and networks such as such as TASO (Transforming Access and Student Outcomes in Higher Education) and the NERUPI network (Evaluating and Researching University Participation Interventions) have been established to facilitate the sharing of good practice and assistance to HEI's with evaluation support, given the complexity of assessing impact and measuring progress towards achieving APP targets. In addition, there are many challenges in determining causality when working within the higher education (HE) sector, that need to be taken into consideration, for example: navigating ethical implications of using randomised controlled trial groups and randomised experimental methodologies, determining extraneous variables when many factors contribute to a student's ultimate degree award, collecting data where systems are perhaps not designed to support intense evaluation, to name a few.

In 2018, the OfS challenged HEI's to achieve the following targets in terms of addressing attainment and continuation gaps (The Office for Students, 2020)

- Eliminate the unexplained[1] gap in non-continuation between the most and least represented groups by 2024-25, and eliminate the gap entirely by 2030-31
- Eliminate the unexplained gap in degree outcomes (1sts or 2:1s) between white students and black students by 2024-25, and eliminate the gap entirely by 2030-31
- Eliminate the gap in degree outcomes (1sts or 2:1s) between disabled students and non-disabled students by 2024-25.

[1] "Some of the factors that contribute to the non-continuation and attainment gaps are structural, such as entry qualification, subject of study or age of students. However, once such structural factors are taken into account, there remain significant unexplained differences which are referred to as unexplained gaps." (The Office for Students, 2020)

Within the overarching targets that the OfS has set, there are specific targets for each institution contained within individual APP's. The mathematical analysis, which is referred to later in this paper, is conducted for the metrics where there are success targets for the University of Kent.

- Reduce the White/Black awarding gap from 27.8% to 8%
- Reduce the White/Asian awarding gap from 13.4% to 4%
- Reduce the White/Other or Mixed Ethnicity awarding gap from 9% to 2%
- Reduce the POLAR4 Quintile 1/Quintile 5 non-continuation gap from 2.3% to 0.5%
- Reduce the Mature/Young non-continuation gap from 6.7% to 3%
- Recue the Mature/Young awarding gap from 5.3% to 1.5%

2. Introduction

The following sections of this paper provide insights into the challenging endeavours to assess outcomes and impact of interventions aimed at reducing awarding and continuation gaps, in the complex context of HE. It explores how the application of a systematic combined formula of process and impact evaluation can and should provide an evaluation framework for targeted interventions aimed to reduce attainment differentials in the context of Higher Education. The model proposed here provides a robust methodology that can be replicated and expanded by evaluation practitioners and academics in this field. The paper also addressees the challenges encountered in a journey, still underdeveloped, into the complex and puzzling exploration of how Student Success interventions can make a difference in reducing inequalities in education outcomes.

The Student Success Evaluation Framework has been developed in the context of the University of Kent's efforts to address and reduce the awarding and continuation gaps affecting underrepresented groups of students across schools and subject areas. Over almost 10 years of implementation, the Student Success Project became a mainstreamed programme with a defining purpose: to identify and to progress areas of research providing insights into the causes and factors contributing to these gaps, while investing in the development of interventions, and mainstreaming research outcomes and good practice. This has required a coordinated approach across the University in terms of the delivery, monitoring and evaluation of interventions delivered across academic divisions, and specific schools, departments, and subject levels. To this end, the Student Success Central Team evolved from having a researcher and an administration officer who were appointed in 2014 and a Project Manager in 2015, to the current team of specialists including Development Officers assisting all academic divisions, Data and Evaluation Manager, System Development Manager and Assistant Managers and administrative team, with an overarching role to provide an implementation framework for Student Success, to support, coordinate, mainstream, and evaluate Student Success interventions across the six academic divisions of the University of Kent.

With a focus on reducing awarding and continuation gaps, targeted interventions[2] have been developed by engaging students most affected by such differentials. In addition, academic, teaching, and support staff have engaged in conversations about the impact of outcomes differentials on students and the University as whole and the Student Success programme has provided resources for them, to actively take part in planning strategies to address these gaps considering the specific context of academic schools, departments, and subject areas. In addition, the Student Success Central Team has also provided academic divisions across the University with the tools and mechanisms to collect and monitor data by

[2] By targeted interventions, we refer to interventions specifically developed to support student groups who are under-represented or disadvantaged in comparison to others. In the case of ethnicity, "Inclusive" interventions are designed to benefit all students but in particular one or more ethnic groups, or "exclusive" interventions that are explicitly designed to benefit one or more minority ethnic groups.

school and subject areas, as evidence to measure the impact of their interventions. This evidence is also required to inform the development and implementation of Student Success Plans when drawing up a strategy to work towards closing the awarding and continuation differentials.

Student Success interventions have not been delivered in isolation from student activities across the University and within specific programmes. These interventions also have contributed, and continue to do so, to other areas of student experience and student support. However, Student Success interventions are distinctive in that these are the operational instrument for the University to meet the institutional OfS targets to reduce its awarding and continuation gaps.

In line with the Student Success aim and purpose, the Student Success Evaluation Framework is explained in this paper by providing insights on the methodology, including the components and types of evaluation embedded in the framework and also its limitations. To this end a detailed account of the ToC and its implication for the framework are explained, followed by a brief overview of the mathematical testing as a method of processing and selecting interventions data, and finalising the section by highlighting the main features and application of CA to determine the causality chains between interventions and the outcomes and impact observed in student engagement and academic outcomes. The paper concludes with a presentation of the main finding of the pilot analysis for the academic year 2019/2020.

3. Methodology

Given the complexity of modelling student attainment and the myriad factors that contribute to a student being awarded a First or 2(I) classification of degree, or continuing within HE, robust quantitative evaluation can be difficult to achieve. We have opted not to use randomised controlled trials (RCTs) due to the ethical implications of providing an intervention to a structurally selected group and not to another. There are also time and cost implications to conduct such an experiment with accuracy. Instead, we allow students to self-select engagement with interventions, while targeting advertising to certain groups most affected by attainment differentials and assume that self-selection is not a contributory factor to increasing grades or attendance. For this evaluation, self-selected students are the comparative group for this analysis.

We attempted many different methods before refining the methodological process to its current point. It quickly became evident that just considering the end of year result for the student was ineffective: we had no control of only providing interventions to students below the First/2(I) threshold, nor could we argue that those below that threshold at the start of the year would not naturally have surpassed it by the end of the year. This led us to use two points of reference: the prior year attendance/attainment and the current year. This framework allows for a mixed-methods approach using both quantitative and qualitative data, through various means, to show the effectiveness of an intervention. It also creates a ‘chain of causality’ subject to how closely the interventions followed the ToC and embeds evaluation from the planning stage of an intervention right through to the conclusion and findings.

4. Theory of Change

The ToC is a model and methodological approach commonly used for process evaluation and a tool for planning and implementation of complex projects. ToC captures the detail of the implementation of a programme by tracing the systematic process to achieving its expected outcomes. By principle, a ToC approach helps to articulate the desired change an intervention is expected to make based on several assumptions that hypothesise how that change can be enabled and how it is going to be measured in terms of impact. As a result, this method of evaluation involves an ongoing monitoring and observation of the activities involved in delivering an intervention. For this reason, the sources of evidence and monitoring mechanisms for data collection are also defined within the planning and process evaluation of ToC. This allows the tracking of inputs, outputs, and outcomes of an intervention over a period. This also requires

an understanding of how interventions are expected to work, what are the causal chain of factors to bring about the change, and the conditions and the wider context in which such interventions take place (Magenta Book, 2020, p24).

ToC is associated with an evaluation that puts emphasis on the narrative description of processes and how these generate specific outcomes. In terms of Student Success, the ToC framework has provided a mechanism for decision-making, implementation, monitoring, and evaluation of interventions. In this sense, the Student Success Evaluation Framework provides a tool to be able to describe how an intervention is delivered and determine the factors that have contributed to or prevented its effectiveness in terms of students' academic outcomes.

Following the ToC process, and for the purpose of the application of the Student Success Evaluation Framework, we have traced the systematic processes of implementation of the selected interventions, from the point where the intervention was included in the Student Success School Plan, up to the point where the intervention was completed. For this reason, to be able to carry out the evaluation, we have relied on the data collected and recorded during the process of implementation by the Student Success teams in academic divisions and schools. These monitoring mechanisms, such as the Student Success interventions tracking spreadsheets, the standardized progress reports, and the post-intervention SStaRT[3] (Student Success Resources Toolkit) form, have constituted our main sources of evidence for analysis. We extracted this data to conduct our impact evaluation through mathematical testing, and CA.

Figure 1 summarises the Student Success implementation framework for the purpose of the ToC process evaluation.

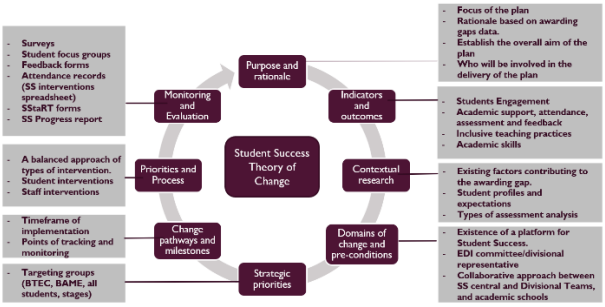

Figure 1 Student Success ToC Process Evaluation

The ToC assumptions that have informed our analysis rest under the core research areas that the Student Success Central Team has investigated since its inception. Through these, the Student Success Central Team has explored the complex dynamics and causes of awarding and continuation gaps, and piloted strategies and mechanisms that can potentially address and reduce inequalities and differentials in students' outcomes. This has resulted in the identification of key areas and themes for research and practice, which have consequently informed the development of interventions and other Student Success activities. These research themes are show in Figure 2.

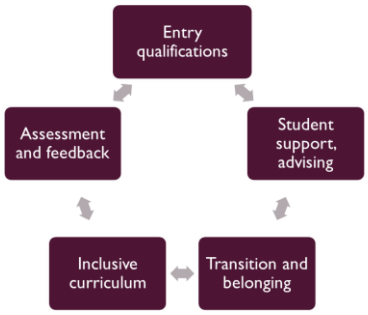

Figure 2 Student Success Research Themes

Based on the outcomes of institutional research in these areas, and the experiences and learning gained over the years of Student Success implementation of activities, interventions, and good practice, we have

[3] The SStaRT form is a mechanism to record practical information about the intervention once it is completed. This information is recorded in a general data set to capture key areas of the intervention and outcomes and serves as a repository of all interventions developed across all academic disciplines. It also provides information for the Student Success Resources Toolkit of good practice, where details of all interventions are provided for use in other schools and departments.

defined our ToC outcomes and assumptions as interventions aimed at:

- supporting and improving students' study skills
- Improving engagement with feedback
- academic advising
- progress monitoring
- improving transition to the university
- improving sense of belonging

which are fundamental for student attainment and continuation.

Therefore, if students from under-represented groups most affected by awarding and continuation gaps are provided with opportunities to engage with these types of Student Success activities, these students will be more likely to achieve **better academic outcomes, increase their general engagement to curriculum and module activities** and complete their degrees on time.

By engaging academic staff across the divisions in activities and interventions to raise awareness of awarding and continuation differentials, developing inclusive teaching practices and diversifying the curriculum, students most affected by awarding and continuation gaps are more likely to engage with more diverse modules and therefore will achieve better academic outcomes and engagement with curriculum activities.

The input of academics and module convenors in decisions around the strategic priorities, and design and delivery of interventions for targeted groups, has been a pre-condition to engender a domain of change, essential when delivering the Student Success implementation framework. Academic engagement in the development of Student Success Plans has been fundamental to ensure that there is a balanced approach when developing interventions, not only for students, but also for academic staff in areas such as inclusive teaching practices and curriculum.

5. Mathematical Testing

To show whether an intervention correlates to a students' attendance or attainment increasing (referred to hence as "metric"), we mapped the changes for every student by academic school and stage on these metrics and removed visible outlier data. We then plotted the distributions of the remaining data. The distributions we found to be approximately normal[4], verified by size ordered plots and Quantile-Quantile plots, thus allowing us to use standardised scores[5] of the data. Quantitative data have been essential to conduct the mathematical testing part of the framework. Individualised student attendance or engagement with interventions allowed us to map the benefits seen. These metrics are students' attainment and attendance.

With the data standardised, it was ready to be analysed. Rather than analysing interventions individually, where often there were not sufficient sample sizes to show a strong conclusion about the effectiveness, we assigned interventions to different codes. We created over 60 granular codes, which we summed to 28 higher-level codes. For instance, there was a granular code such as "Skills workshops", "Skills 1:1's", "Skills exam preparation", but these were also analysed altogether as "Skills". Part of the ability to do this is due to standardising scores so that we could then regroup later as, even though x_1 and x_2 may have different means and standard deviations, Z_1 and Z_2 have the same mean and standard deviation.

Given that we found that all x_k were approximately normal, we had that all Z_k were also approximately normal. To assess the effectiveness of an intervention we then considered attendance at an intervention as group A and those who did not as group B, the combination of these two groups being approximately normal. We can no longer say that the mean and standard deviations of group A and group B are equal to zero and 1 respectively, but we assumed that each group is still approximately normal. We then

[4] Normally distributed variables follow a "bell-curve" shape when viewed as a probability distribution with a symmetry around the mean.

[5] For each school and stage we calculate $Z_{ik} = \frac{x_{ik}-\mu_k}{\sigma_k}$ where k denotes the school and stage in question, Z_{ik} is the standardised score of data point i, x_{ik} is the value of said data point, μ_k is the mean of x_k and σ_k the standard deviation of x_k. The distribution of Z_k then has mean 0 and standard deviation 1.

conducted a two-sample t-test with unequal variance (Welch's test) to show if the intervention was effective mathematically. We used a one-tailed test, as it is most imperative to discover if the intervention benefits students rather than it just being different to non-engagement.

Our hypotheses are thus:

Null hypothesis: The mean of the standard scores of students who attended the intervention was less than or equal to the mean of the standard scores of students who did not attend the intervention.

Alternative hypothesis: The mean of the standard scores of students who attended the intervention was greater than the mean of the standard scores of students who did not attend the intervention.

Given we were using a t-test to establish if interventions showed statistical significance of success, small samples presented an issue for the analysis. With sample sizes (i.e. attendance at an intervention code) of 50 students or less, there was not enough data to be confident in the probability value that returned, as the mean calculated would have a large margin for error. However, we included the results in the analysis and conducted modelling to show the power (the likelihood of avoiding a Type II[6] error) of the test.

Students were only included in the evaluation if they had data (either on attendance or attainment) from the prior year to the intervention and the year of the intervention to enable us to compute the change. Additionally, only students whose stage changed between these two data points were included to avoid some of the spurious factors that occurred when there are capped resits if a student was resitting the stage.
It is most common to use a confidence level of 95% for statistical significance; however, we have opted to use a lower threshold of 90%[7]. We premised these reasons upon the mathematical testing only being the first stage of a process; the CA allowed us to show further rigor after an intervention style has "passed" the mathematical testing. Given that this was a natural experiment, we were also aware of the multitude of factors that we have not accounted for, including multiple intervention attendance and other changes that occurred within the academic year, both student life and degree alterations.

There were a number of assumptions made in the mathematical analysis:

(1) The combination of the approximately normally distributed schools and stages was also approximately normally distributed.
(2) The group that engaged and the group that did not engage, which both came from an assumed approximately normally distributed group from (1), were assumed to each be individually approximately normally distributed.
(3) The data was a representative random sample of the population, the population in this case being all future students.
(4) That there was little or no collinearity or multi-collinearity between any of the metrics examined.

These assumptions were accepted as reasonable, and we have verified some of the more spurious suppositions such as the maintenance of normality between engagement and non-engagement with sample checks.

Given that the experiment and control group were not decided by us but by self-selection of participants, we did not construct the analysis to reach a certain power. Nevertheless, it can be useful to assess the power of the codes that showed statistical significance. This was completed using MATLAB 2021 and a simulation of 1,000,000 iterations of randomly generated groups with approximately the same mean and standard deviation as the test groups.

[6] Type II errors are where the test result is not that of significance, but it is a false negative.
[7] This increases our risk of a Type I error of a false positive, however we feel that this is managed by the CA conducted before a conclusion is drawn about the intervention.

In this evaluation, multiple t-tests were conducted on the data to show effect on relevant protected groups. Necessarily, we have considered the use of a correction such as the Bonferroni correction. Having explored on the application of such a methodology (Armstrong, 2014; Perneger, 1998; Armstrong, 2011), we have concluded that we are not concerned about establishing that all tests are significant for a code. Due to this, and the concern of an increased Type I[8] error, we have chosen not to use a correction within this evaluation.

When we had the evidence to accept the alternative hypothesis at least at the 10% level, then we concluded that the intervention was mathematically successful and that there was evidence to suggest that this intervention code should be continued for future years, as it is highly likely that it has contributed to increased attainment or attendance. We then review these interventions under the CA to establish the causality chain.

6. Contribution Analysis

CA is an approach to examine systematically if an intervention has contributed to the observed outcomes by tracing all the stages of the implementation process and by capturing the evidence collected during this process. CA is also a process to verify the theories of change and assumptions made around the expected outcomes of an intervention to find plausible association between activities, outcomes, and impact. A plausible association can be established when CA is developed based on a ToC process and when there is evidence that an intervention has been implemented as planned. Therefore, the causality chain of expected results can be observed and evidenced, and other influential factors that appeared to make a difference or a relative contribution are clearly recognised (H. White and D. Phillips 2012, p42).

To develop the Student Success Evaluation Framework CA was adopted as this approach acknowledges the multiple factors that can influence an outcome of an intervention in a complex setting[9] of implementation (Mayne, 2017) such as in HE. The context and purpose of Student Success interventions made it unsuitable to use randomised control groups or trial methods as an evaluation mechanism, as Student Success interventions have been developed within the curriculum and teaching practices of the university, rather than in a controlled experimental setting. CA, therefore, supplies a mechanism to make causal inferences when it is not possible, practical, feasible or ethical to use control trial methods. This is of relevance when there are ethical implications and methodological constraints to select randomised controlled groups as participants or beneficiaries of an intervention (Mayne, 2017) as explained earlier in the methodology section.

As a result, rather than providing precise attribution estimates of the dimension of the impact of the Student Success interventions on the observed outcomes, our Student Success Evaluation Framework is concerned with both the extent of the change and why the change occurs within the context where the intervention takes place. CA considers the generative perspective on causality to assess whether the intervention has or has not made a difference and explores the contributing factors to demonstrate a causality chain of impact[10].

For this purpose, the CA approach has been adopted by the Student Success Evaluation

[8] Type I errors are where the test result is that of significance, but it is a false positive.

[9] Mayne defines "complex settings" as the context of […] interventions that are multi-dimensional in nature, often compromising numerous sub-interventions with feedback loops and numerous influencing factors at play, and involving interactions, synergies and possibly emerging outcomes […] (Mayne, 2017, p54)

[10] "[…] generative causality: seeing causality as a chain of cause-effect events. This is what we see with interventions, a series or several series of causal steps – impact pathways - between the activities of the intervention and the desired impacts. Taking the generative perspective and setting out an impact or contribution pathway is essential in understanding and addressing the contribution made by the intervention. The associated ToC sets out what is needed if the causal links involved, and hence the expected results, are to be realized" (Mayne 2020, p2)

Framework to determine the difference the interventions have made in terms of impact on student engagement and academic outcomes and determine the causality chain between outcomes and impact by responding to the following questions:

1. Did the intervention influence a change, or did the intervention make an important contribution to change?
2. Was the intervention implemented as intended?
3. Has the intervention reached the target group of students?
4. How was the intervention experienced both by those implementing it and those who took part in the intervention?

The CA analysis began with addressing our first research question concerning impact evaluation: What difference has the intervention made? To approach this question, we identified the interventions that appear impactful in terms of the data analysis and mathematically testing. This means that we observed indicative correlations between attending these interventions and increased grades and engagement, in comparison to those students who did not take part in these interventions.

To respond to the process evaluation questions, which are informed by our ToC, we observed if the intervention was implemented as intended, if the intervention reached those students targeted to attend, and how both those implementing it and those who took part in the intervention experienced the process and the outcomes of the intervention. We verified first if the schools that delivered the selected interventions followed each stage of the Student Success ToC planning process to determine the impact pathways that would allow us to establish the plausible associations that inform the contribution causality claims.

Following this step, we gathered the existing evidence of the additional components of the implementation process and the assessment of interventions and recommendations recorded in the school progress reports and the SStaRT forms. When available, we also looked at feedback forms, focus groups, case studies, surveys, etc. concerning the interventions and recorded by the schools in the progress reports.

To assemble the causality chain and to assess the extent to which we were able to make causal contribution claims through plausible associations, we considered the alternative explanations and influencing factors framework proposed by Lemire, Bohni and Dybdal (2012) and adapted the REF (Relevant Explanation Finder) tool provided by these scholars. As a result, we measured the strength of our evidence by tracing the process of implementation, which is embedded in the Student Success ToC, and identified the influential or contributing factors to assess their impact on the observed outcomes resulting from the mathematical testing, against four parameters: certainty, robustness, prevalence, and range.

The level of influence of the intervention in terms of "certainty" was measured by establishing the degree to which the observed outcomes matched the expected outcomes. In this area, we looked at whether the intervention was included in the school plan and specifically within it in the rationale of the plan. Then we assessed if there was an explicit target group of students identified for the intervention, and if the school's own target was met. We also considered any qualitative evidence, if it existed, such as feedback from students or those who delivered the intervention.

The "robustness" assesses if the intervention was a significant contributor of the observed outcomes when looking at the quality of the data collected in the schools' tracking spreadsheets and in the additional monitoring mechanisms aimed at collating qualitative feedback and observations about the intervention. To measure this parameter, we observed whether data on students' attendance was reliable and accurately recorded in the intervention spreadsheets or/and in the SStaRT form. We observed whether at least 10% of students were reached for each of the groups targeted (from the school as a whole) to attend the intervention and if this was reflective of the mathematical impact identified for the school. In this area, we also looked at any qualitative records on challenges, or good practice in terms of meeting the target groups against the impact outcomes observed.

The "prevalence" was measured by looking at the extent to which the intervention influenced other areas of implementation of the plan, or led to other Student Success interventions or activities, as recorded in the progress reports and SStaRT forms. To assess this parameter, we looked at the evidence indicating if the intervention was linked to the school plan indicators explicitly, and if the intervention also linked to other areas of implementation within the plan. We observed if the data and information contained in the school's progress reports made references to the intervention generating outcomes affecting other interventions. We also looked at the data in the intervention spreadsheet to establish if more than one of the intended groups targeted were impacted by the intervention.

The "range" of the interventions measured the extent to which more than one outcome is observed because of the intervention. The mathematical testing indicated which interventions appeared to have an indicative correlation of impact on student attainment or/and student engagement. As these were the outcomes of interest, we assessed if one or both were observed for one single intervention; therefore, these interventions had a greater range of impact. We looked at the evidence in the school progress reports, to see if the school's own initial assessment of impact on these outcomes of interest matched the observed mathematical impact for that specific school. In some cases, we found that the impact observed was not anticipated or expected, as the evidence was absent from the progress report. Other qualitative evidence was also considered in terms of the range of the impact as recorded in the SStaRT form, or from other feedback collected from those who attended the intervention or from those who delivered the intervention.

Once the level of influence of these contributing factors was established through the analysis and scoring based on the evidence against the four parameters, we proceeded to establish the plausible associations that informed the causal-chain and contribution story. This defined the interventions causal pathways to the observed outcomes and impact.

7. Findings

Pilot analysis academic year 2019-2020

The initial exploration and analysis of the interventions recorded by schools in the academic year 2019-2020 resulted in 60 types of interventions. These types of interventions were re-classified into 28 high level codes. All these codes associated to the interventions recorded for this academic year were subject to the mathematical testing analysis. As a result, seven sub codes of interventions appeared to show an indication of correlations between attending the intervention and improving student's attainment and/or attendance. These seven codes comprised interventions within three main categories or high-level codes: Academic Advising, Inspirational speakers, and Skills workshops which were delivered for students in Stage 2 and Stage 3 in nine academic schools[11]:

1. School of European Culture and Languages (SECL)
2. School of Physical Sciences (SPS)
3. School of Sport and Exercise Sciences (SSES)
4. School of Computing (SoC)
5. School of Mathematics, Statistics and Actuarial Science (SMSAS)
6. Kent Business School (KBS)
7. School of Psychology
8. Kent Law School (KLS)
9. School of Social Policy, Sociology and Social Research (SSPSSR)

These higher-level codes were subject to CA to establish the causality chain of impact. To this end three tests were undertaken: First, the influential factors in the ToC process for the CA we scored for the following parameters: certainty (max 3), robustness (max 3), prevalence (max 3) and range (max 2). In the process of assigning scores for these

[11] In 2019-2020 these academic schools were part of the Student Success Project and have delivered interventions with sufficient evidence recorded to be included in the evaluation, as the Student Success became mainstreamed the following year 2020-2021 all academic divisions and schools became part of the programme which resulted in a significant increase of Student Success interventions delivered across the University.

parameters it was necessary to carefully verify the existing evidence, both qualitative and quantitative, of the delivery of interventions aligned to the codes by looking at the information recorded in school plans, academic division reports, and SStaRT forms. These scores provided an average per school that was then recalculated into a percentage, which represents the level of influence of the code on the participants of the intervention. Second, we considered the power[12] of the statistical tests conducted. And third, the score of level of influence of the code was multiplied by the power score to give us the final impact rating of the code.

This analysis resulted in the following main findings:

Within the code AA (Academic Advising), the sub code AA2, which compiled interventions concerning structuring the academic advisor system developed by different schools, such as visit to academic advisor, group, or individual meetings, attending first meeting, and so on, had 69% level of influence. The interventions for this sub code found BAME students had improvements in overall attendance with the testing having 57% power and so 39% attendance impact rating. Mature students had improvements on their overall attendance with a 94% power and thus 65% score for attendance impact rating. This means that the strength of the causality chain was strong mainly for mature students concerning improvements in their overall attendance (see Figure 4 for the mathematical testing results) and negligible for BAME students for this specific area of impact.

The sub code IS3 (Study based speakers), that groups the interventions of subject specific Inspirational Speakers, showed a level of influence of 52%. The interventions for this code showed an improvement of attainment for BTEC students with 59% power and so 31% score of attainment impact rating, revealing a weak strength of the causality chain.

In relation to the interventions contained in the code SS1, referring to Skills Workshops, the level of influence of contributing factors was of 70%. This intervention code had improvements on attendance for BAME students with 82% power and so 57% attendance impact rating (see Figure 5 for mathematical testing results) and on Black students with 74% power and so 52% attendance impact rating (see Figure 6 for mathematical testing results). In addition, BAME students also had improvements on attainment with 51% power so 36% attainment impact rating.

Skills Sessions based on interventions for students scoring low on a diagnostic test or for repeaters/re-sitters we grouped in the code SS4. This code obtained 53% level of influence of contributing factors. This intervention had 54% power so a score impact on attainment improvement of 29% for Black students.

The analysis for the general code for Skills, SS, resulted on a 53% level of influence. There were improvements on attendance for the interventions compiled in this code with BAME students (SS1, SS7, SS8) that had 66% power and so 35% impact rating, for Mature students with 97% power and so an impact rating of 51% (see Figure 7 for mathematical testing results), and for Other/Mixed ethnicity students (SS1, SS2, SS4) with a 76% power and so a 40% impact rating. This suggest that the strength of the causality chain was strong for BAME students and medium for Other/Mixed ethnicity and mature students.

Figure 3 Table of FindingsFigure 3 summarises the findings of the CA for each of the codes of interventions.

[12] Statistical power relates to likelihood of a statistical test identifying an effect when there is an effect to identify.

Code	Sub-code	Impact	Level of Influence	Statistical Power	Impact Rating	Strength of the causality chain
SS	SS4	**Attainment**: Black	53%	54%	29%	Weak
	SS	**Attendance**: BAME	53%	66%	35%	Weak
		Mature		97%	51%	Medium
		Other/Mixed		76%	40%	Medium
	SS1	**Attendance**: BAME	70%	82%	57%	Medium
		Black		74%	52%	Medium
		Attainment: BAME		51%	36%	Weak
AA	AA2	**Attendance**: BAME	69%	57%	39%	Weak
		Mature		94%	65%	Strong
IS	IS3	**Attainment**: BTEC	52%	59%	31%	Weak

Figure 3 Table of Findings

When the impact rating is below 50% we determine this as a weak causality chain and therefore cannot confirm that the intervention was the cause of the impact. Values between 50% and 64% produce a medium strength chain, whereby further evidence would be better but we can confirm an argument that the intervention resulted in the impact seen. Finally, values above 65% are determined as strong chains, here we state that the intervention caused the impact.

It is important to acknowledge that although the CA provides a strong proxy analysis of the level of influence of the factors that define the delivery of interventions on the strength of the impact of these interventions, this model has limitations. One of these is that the model has to rely on the existing evidence captured by academic schools in the school plan, school reports and SStaRT, in both forms quantitative and qualitative. Therefore, the absence of this evidence at any or all these mechanisms of data collection affects the overall average of impact as there is low scoring in certain levels of influence where no evidence was found. In these cases, it is evident that our assertions about the impact of the interventions are conditioned by our access to enough robust and quality of data and other information available through the verification process mentioned earlier. The continuing efforts to improve data collection and evidence gathering mechanisms and increasing awareness of the importance of these areas of the implementation process will significantly mitigate the limitations of the model in the future.

Another limitation observed during the CA pilot analysis, concerns the coding system. As interventions change, and innovation is a key feature of the Student Success implementation process, the coding scheme needs to be reviewed year by year to make sure it does reflect accurately the schools and academic divisions responses in terms of the changing dynamics of awarding differentials between students with specific characteristics. However, it is important that by reviewing the coding scheme there is still scope for comparison year by year, therefore the most impactful interventions can be identified to be replicated when it is suitable to do so.

8. Conclusion

In the journey of more than two years to develop and assemble the Student Success Evaluation Framework, there were several lessons learned and challenges we had to overcome during this process. Embarking on this venture meant that we had to reflect on the core aims of Student Success, what the Student Success Central Team meant to achieve and make this explicit from the outset of the Framework. This exercise involved reviewing the implementation framework of Student Success activities and in particular the mechanisms for reporting progress, data collection and monitoring. This revealed some weaknesses that we had to amend, and adjustments were introduced to improve these mechanisms, allowing us to standardise the data collection methods and treat it with more rigor. These preparation steps were necessary for the purpose of the evaluation framework presented here.

During the process of developing the Framework we became more aware than ever of the need to facilitate the data collection and monitoring processes for those responsible for this task at a school level. Whilst doing that, we learnt that we needed to engage in an ongoing conversation with academic and professional services staff around the fundamental idea and principles of evaluation. The existing platform of the Student Success Research and Evaluation Group have been instrumental in supporting and advising this process.

In the transition from one phase of the project to another and moving from a year-by-year funded project to a Student Success mainstreamed programme, we faced many challenges to developing the Framework. One of these challenges was the worldwide COVID pandemic. The COVID restrictions affected our students in ways that we cannot measure yet. The impact of the disruption of classes and moving these to online platforms, and the pressures on their physical and mental health and their families cannot be underestimated. It was a period of significant challenge for the whole University.

Nevertheless, we also learnt that the academic year of 2019-2020, despite the challenges resulting from COVID, was a year of increased attendance of students in interventions. We also observed an increasing number of interventions developed by Student Success teams in the academic schools involved in the programme and receiving Student Success funding. The outstanding work of these teams in such circumstances needs to be acknowledged. The improvement and standardization of reporting, data collection and monitoring mechanisms also paid off. As a result, we were able to use the Student Success outcomes and data from this academic year to pilot and test this evaluation framework. This allowed us to achieve the best approach and rigor for each of its components, the ToC, the mathematical testing, and CA.

The ToC was developed not only as a reflecting mirror of the Student Success implementation framework, but also to evaluate the process of developing interventions meeting the requirement of Type 1 evaluation and narrative. This approach has proven to be effective as we built from this initial stage a solid ground for the development of the mathematical testing that resulted later in the foundations and road map to the CA.

The mathematical testing went through many iterations as it was refined, to ensure that only pertinent and valid data was included in the analysis, thus enabling strong conclusions to be drawn. This provided a strong foundation for empirical analysis required for Type 2 impact evaluation. It is likely that this method will continue to undergo minor alterations to maintain the data integrity, but at this stage it is robust enough to produce the findings illustrated in this paper. It has also led to a greater understanding of how different subject areas grade over different stages, for instance if Stage 2 has lower or a different grade distribution to Stage 3, which has been an invaluable piece of information when considering intervention timing.

The CA was developed at the last stage of assembling the evaluation framework and this resulted from deliberations on how to overcome the significant challenge of embedding Type 3 evaluation (to assert causality) into the framework. We have learned the principles of CA, and the different approaches to understand the plausible associations of contribution factors that determine the chain of causality. This approach and concepts have been applied and developed in contexts different from HE. However, we adopted this approach and saw the value of CA, not only because its analytical tools are informed by a well-defined ToC, but also because CA provided us with a tool to identify which cause-effect questions were being addressed by looking at causal-contribution rather than causal-attribution. We incorporated the approach of influential factors as a means to assert plausible association. However, we extended and enhanced this approach by creating a mechanism to measure the strength of these factors when addressing impact and the validity of the chain of causality.

The Student Success Evaluation Framework is an important addition to educational research as it can be applied to many other institutions aiming to evaluate interventions of several types. While it is most applicable in this paper to Student Success initiatives for the APP, the framework can be applied to any style of intervention where the metrics are normally distributed and have two points of reference. It is also possible to separate by programme of study rather than school as well as many other variations. In terms of evaluation where RCT's are not a viable option, this framework gives an alternative for establishing a causality chain of impact when assessing the outcomes of Student Success interventions.

9. Appendix

AA2

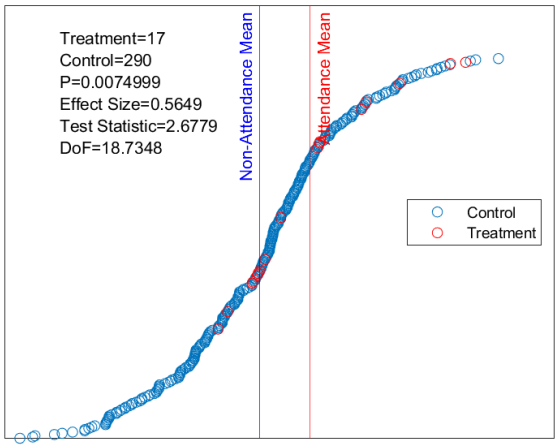

Figure 4 AA2 Mature Attendance Impact

SS1

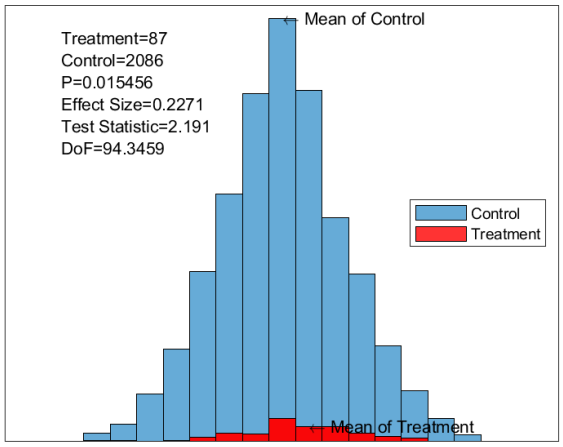

Figure 5 SS1 BAME attendance impact

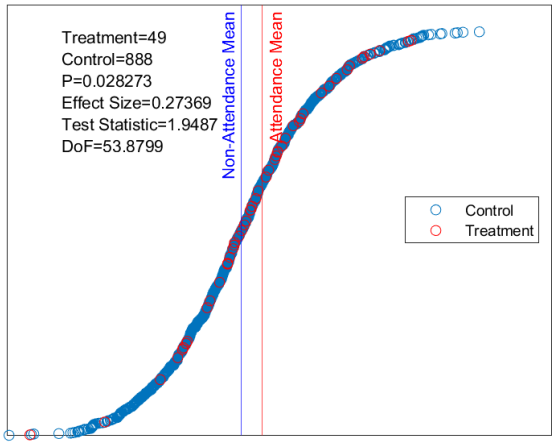

Figure 6 SS1 Black Ethnicity Attendance Impact

SS

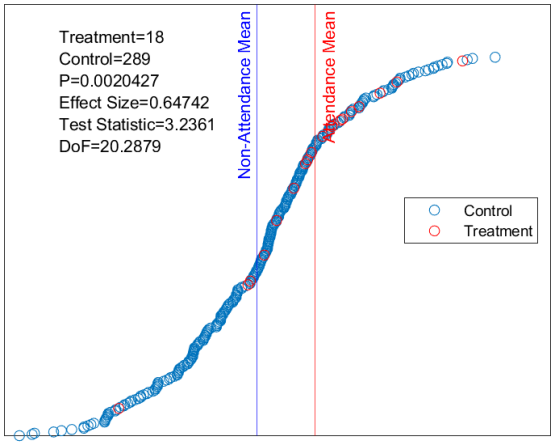

Figure 7 SS Mature Attendance Impact

10. References

Armstrong, R., 2014. When to use the Bonferroni correction. *Ophthalmic Physiol Opt*, Volume 34, pp. 502-508.

Armstrong, R. D. L. D. M. a. G. B., 2011. Statistical guidelines for clinical studies of human vision. *Ophthalmic and Physiological Optics*, Volume 31, pp. 123-136.

Astbury, B. & Leeuw, F., 2010. Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), p. 363–81.

Astbury, F. & Leeuw, B., 2020. Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), pp. 363-381.

Budhwani, S. & McDavid, J., 2017. Contribution Analysis: Theoretical and Practical Challenges and Prospects for Evaluators. *Canadian Journal of Program Evaluation*, 32(1), p. 1–24.

Davies, R., 2018. Representing theories of change: technical challenges with evaluation consequences. *Journal of Development Effectiveness*, 10(4), pp. 438-461.

Delahais, T. & Toulemonde, J., 2012. Applying contribution analysis: Lessons from five years of practice Evaluation. *Evaluation*, 18(3), p. 281–293.

Dybdal , L., Bohni Nielsen, S. & Lemire, S., 2012. Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation*, 18(3), p. 294–309.

Dybdal, L., Bohni Nielsen, S. & Lemire, S., 2010. Contribution analysis applied: reflections on scope and methodology.

Canadian Journal of Program Evaluation, 25(2), p. 29–57.
Funnell, S. & Rogers, P., 2011. *Purposeful Programe Theory: Effective Use of Theories of Change and Logic Models.* San Francisco: Joseey-Bass A Wiley Imprint.
HM Treasuy, 2020. *Magenta Book Central Goverment Guidance on Evaluation.* London: Crown.
Koleros, A. & Mayne, J., 2019. Using Actor-Based Theories of Change to Conduct Robust Evaluation in Complex Settings. *Canadian Journal of Program Evaluation,* 33(3), p. 292–315.
Mayne, J., 2001. Addressing attribution through contribution analysis: using performance measures sensibly. *Canadian Journal of Program Evaluation,* 16(1), pp. 1-24.
Mayne, J., 2012. Contribution analysis: Coming of age?. *Evaluation,* 18(3), p. 270–280.
Mayne, J., 2017. Contribution Analysis: Addressing cause and effect. In: F. K. a. M. M. Schwartz R, ed. *Evaluating the Complex. Attribution, Contribution and Beyond.* New York: Transaction Publishers, pp. 53-95.
Perneger, T., 1998. What's wrong with Bonferroni adjustments. *BMJ,* Volume 316, pp. 1236-1238.
The Office for Students, 2018. *Difference in Student Outcomes.* [Online]
Available at: https://www.officeforstudents.org.uk/data-and-analysis/differences-in-student-outcomes/ethnicity/
[Accessed 09 05 2022].
The Office for Students, 2018. *Securing student success: Regulatory framework for higher education in England.* [Online]
Available at: https://www.officeforstudents.org.uk/media/a32be394-72ba-4975-bd7e-5a9b5d8c0ece/regulatoryframework.pdf
[Accessed 21 04 2022].
The Office for Students, 2019. *Standards of Evidence and Evaluation Self-Assessment Tool.* [Online]
Available at: https://www.officeforstudents.org.uk/media/6971cf8f-985b-4c67-8ee2-4c99e53c4ea2/access-and-participation-standards-of-evidence.pdf
[Accessed 10 05 2022].
The Office for Students, 2020. *Access and Participation Glossary.* [Online]
Available at: https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/access-and-participation-glossary/
[Accessed 20 04 2022].
The Office for Students, 2020. *Transforming Opportunity in Higher Education.* [Online]
Available at: https://www.officeforstudents.org.uk/media/2efcda44-8715-4888-8d63-42c0fd6a31af/transforming-opportunity-in-higher-education.pdf
[Accessed 07 04 2022].
The Office for Students, 2022. *Access and Participation Data Dashboard.* [Online]
Available at: https://www.officeforstudents.org.uk/data-and-analysis/access-and-participation-data-dashboard/
[Accessed 05 05 2022].
van Es, M., Guijt, I. & Vogel, I., 2015. *Hivos TOC Guidelines Theory of Change Thinking in Practice.* The Hague: Hivos.
White, H. & Phillips, D., 2012. Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. *International Initiative for Impact Evaluation* , June, Volume 15, pp. 3-61.