TASO
Annual Conference:
How to Evaluate
8 May
#TasoCon24

TASO

Opening and welcome remarks

Dr Omar Khan, CEO, TASO

#TasoCon24

In conversation: Supporting disabled students in HE

Amelia McLoughlan, Network Director,
Disabled Students UK

#TasoCon24

Impact evaluation: using quasi-experimental designs in HE evaluation

#TasoCon24

Who we are

Sonia Ilie
Associate Professor
University of Cambridge

Mike Kerrigan
Head of Research and Insights (Access and Participation)
Nottingham Trent University

TASO

Overview of session

1. Impact evaluation
2. Quasi-experimental designs (QEDs)
3. A practical example
4. Q&A

TASO

Evaluation should ideally be . . .

Flexible

Practical

Informative
(to others)

Robust

TASO

The challenge of impact evaluation

1. How can we tell that a programme causes **impact?**
2. The challenge: **estimating the counterfactual**
3. The counterfactual usually achieved through **random assignment**
4. If randomisation not possible - **Quasi-experimental designs**

Estimating the counterfactual

While **random assignment** usually generates the strongest counterfactual (the control group),
QEDs can estimate the counterfactual through statistical means.

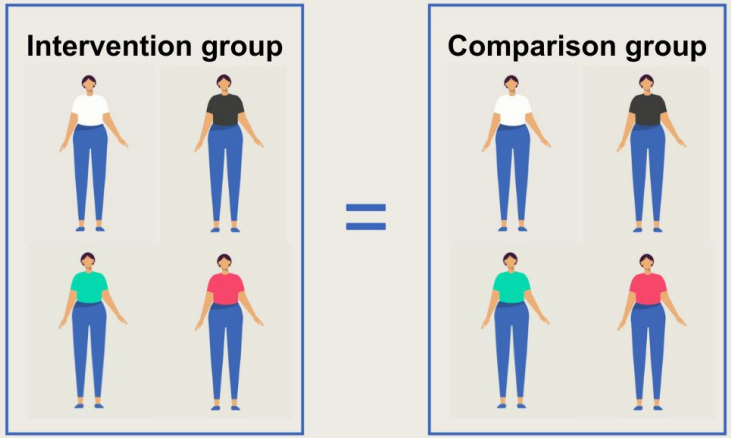
Estimating the counterfactual

The counterfactual in QEDs is not always as strong as that obtained through random variation, because it cannot always avoid self-selection bias completely.

Quasi-experimental designs

Can be deployed creatively, to make use of ‘natural’ or ‘random’ variation in programme implementation.

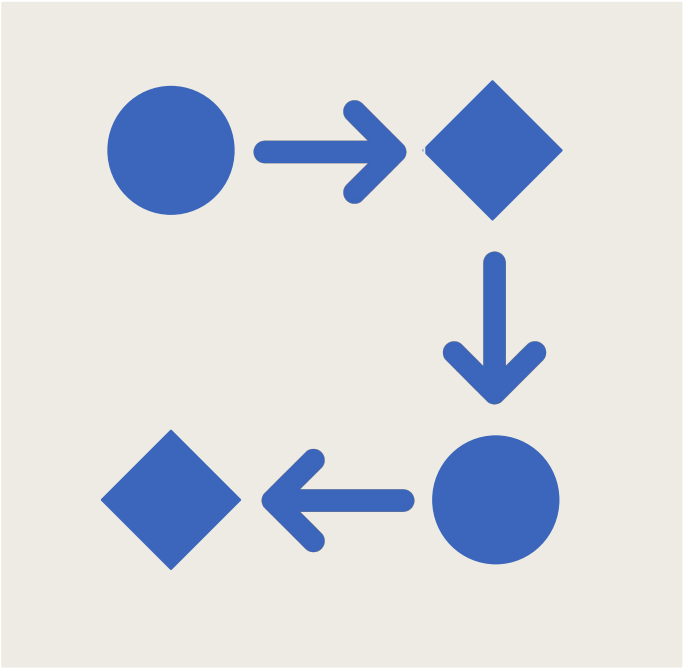
Quasi-experimental designs



Can generate a comparison group as similar as possible to the intervention group, making use of existing data on:

- previous cohorts
- participants just ineligible
- non-participants due to oversubscription
- non-participants due to time considerations?

Quasi-experimental designs



Come in a variety of forms and use a variety of statistical approaches.

Therefore, understanding the intervention in detail, through a robust theory of change, is essential.

TASO

Quasi-experimental design approaches

#TasoCon24

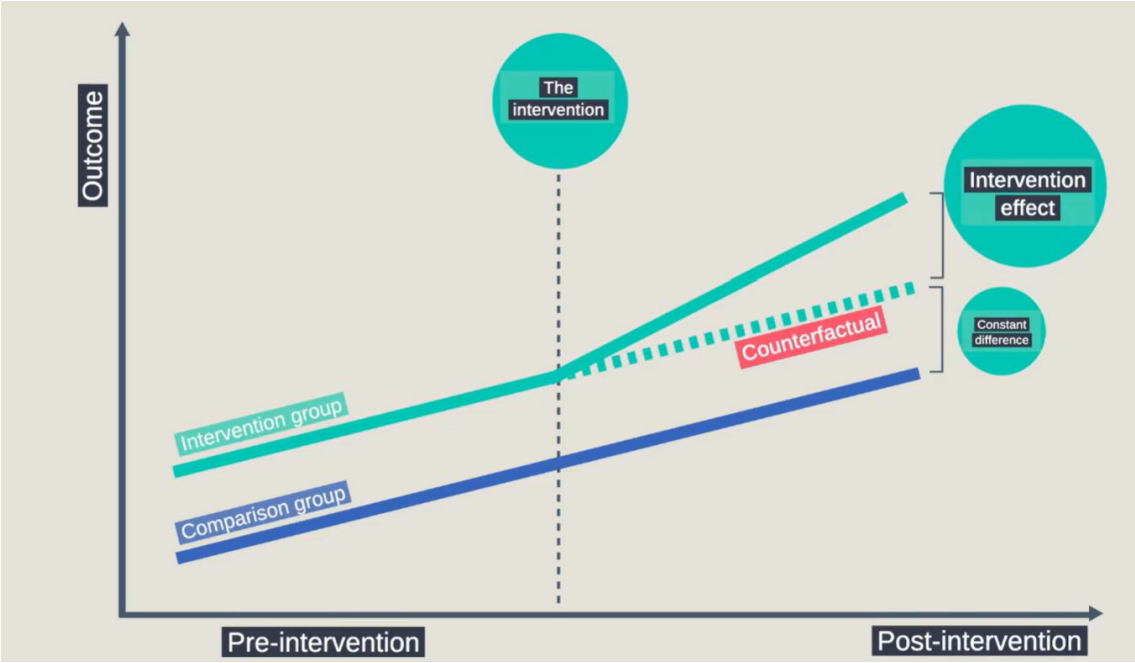
TASO

Before-and-after designs (first difference)

- The weakest of all QEDs, usually **not considered as QED-proper**
- The counterfactual is the weakest, because it assumes that ‘no changes would have occurred in the absence of the programme’; and because it cannot account for any confounding influences, any other factors that may impact on the outcome
- Generate Type 2 evidence, but without attribution of impact

Difference-in-difference (second difference)

A robust, widely-used and well-understood QED approach



TASO

Difference-in-difference

- Provide strong evidence of impact (if present), especially if combined with matching (coming up!)
- They allow for unobserved heterogeneity in between groups
- But data needs to meet several assumptions:
  - The unobserved heterogeneity is time-invariant and
  - The pre-intervention difference in groups is (and would remain) constant
  - An understanding of selection is required

TASO

A side note: natural experiments

- Very similar logic to that of both Diff-in-Diff and experimental (RCT) designs.
- Common logic in education - the intervention is:
  - assigned systematically at the level of school/HEP/region/etc., and also
  - assigned randomly at the level of the individual
- This can isolate the impact of the intervention being evaluated

TASO

Matching approaches

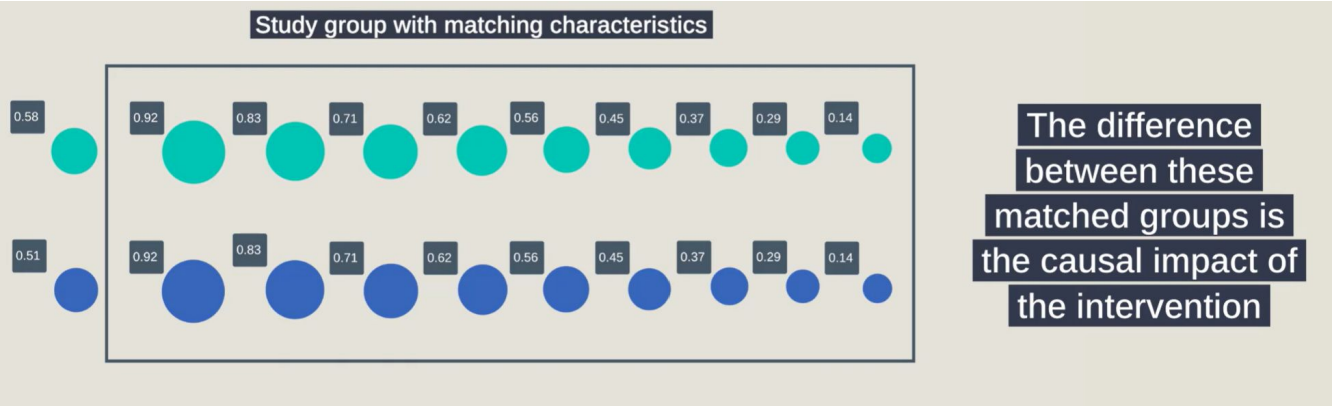
- The most common is Propensity Score Matching (PSM)
  - can generate a good counterfactual but doesn't always work as planned and therefore generates strong Type 2 evidence but not always Type 3.

- Other matching approaches include:
  - matching to a given distribution (in terms of frequency)
  - synthetic control
  - case-control matching

Propensity Score Matching (PSM)

Looks to generate a comparison group similar to the intervention group in terms of individuals' probability to take up an intervention

This propensity is calculated as a function of a (large set of) background characteristics



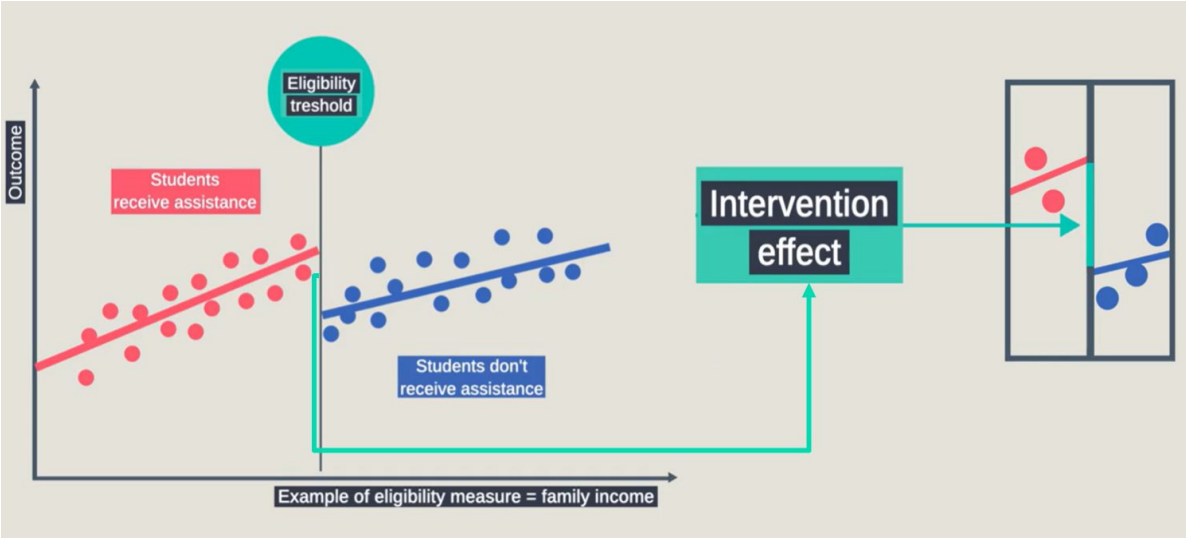
TASO

Propensity Score Matching (PSM)

- Requires the empirical identification of factors associated with intervention take up and that unobserved factors do not affect participation (conditional independence)
- Requires ‘good matches’ (common support) – this does not always happen
- Depends on very large number of analytical decisions which affect the strength of the counterfactual and therefore of the causal inference

Regression discontinuity designs

Makes use of eligibility thresholds to compare the outcomes of those immediately either side of the threshold



TASO

Regression discontinuity designs

Can generate a strong counterfactual but have substantial assumptions and data needs

- There needs to be a suitable cut-off point for receiving the intervention
  - This cut-off or threshold is not always very precise
- Typically requires existing administrative data
  - This reduces the need for data collection
  - *But* the amount of data required is high, especially around the threshold, and for those not eligible (sometimes difficult!)
- It can only tell you about the impact on individuals closest to the threshold, not those further away – ‘local treatment effect’

Many other QEDs exists

- Instrumental variable analysis (IV)
- Interrupted time series analysis
- Other matching techniques

All require specialist technical input

Institutional Data Use Project

Exploring QEDs with institutional data: practical guide

Mike Kerrigan, Nottingham Trent University

#TasoCon24

Methods explored

1. ~~Regression discontinuity design~~
2. ~~Difference in difference~~
3. Propensity score matching / case control matching
4. Logistic regression

Institutional data sets

1. Enrolment and outcomes data (HESA return)
2. NTU student dashboard
3. Participants of post-entry activities (2022/23 ECAs)
4. Importance of identifying influential covariates

Case control matching (CCM)

1. Used in observation studies – reduce selection bias
2. Pairs participants with non-participants
3. On factors known to influence the outcome
4. Can apply tolerances – sample size v good matches

Propensity score matching (PSM)

1. Estimates probability that students will participate*
2. Based on combination of characteristics (covariates)
3. Pairs participants with non-participants (PSM score)
4. Again, tolerances can be applied (different to CCM)

* Even though, as this is retrospective, we already know who participated!

Four stages in propensity score matching

1. Estimate propensity scores based on the covariates
2. Match individual participant with a ‘partner’
3. Check for balance in mean propensity scores
4. Estimate intervention effect and interpret results

Source: betterevaluation.org

Some reflections (before results)

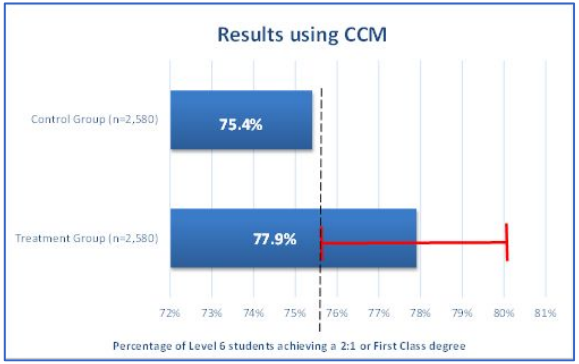
- CCM & PSM not too difficult to do with statistical software
- But they can be difficult to do well
- Do *not* just shove all your data into a statistical package
- Consider the factors that influence participation and outcome
- Consider a regression analysis to inform covariates
- Consider the Level of study (if Level 6, can use Level 5 attainment as a covariate)

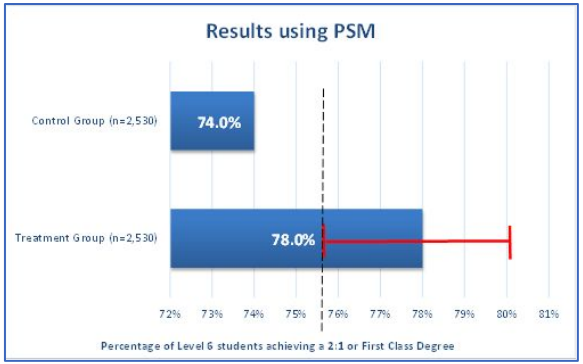
Some reflections (before results)

- Trade-off between sample size and good matches / balance
- Avoid ‘p-hacking’
- Are there known and/or unknown unobservable factors?
- Can we achieve Type 3 (causal) using matching methods?

Some tentative results

- In both cases treatment group had higher rates of 2:1 / First than control group ($p<0.05$)
- ‘Effect’ size seemingly higher in PSM (4 pps diff) than CCM… (2.5 pps diff)
- But these effect sizes were considerably lower than the aggregated analysis of participants v non-participants (13 pps)
- Similar trends found for Levels 4 and 5 GBA
- Confirms considerable selection bias that must be controlled for





Some tentative results

- Very useful learning exercise – try it!
- Institutional datasets are complex
- Type 3 causal methodologies (RCT, DiD, RDD) not always feasible
- But robust evidence of impact can still be gleaned using alternatives
- These methodologies provided strong Type 2 evidence of impact (Type 2.5?).

  As good evidence as we are going to get with retrospective data? A

  pragmatic solution?

Some tentative conclusions

- Recommended that the learnings are considered, and methodologies tweaked with evaluation built into the design of post-entry activities
- Explore outcomes data suited to DiD? E.g. module attainment?
- This will further develop opportunities to deliver causal evaluation

TASO

Q&A

#TasoCon24

TASO

QED webinar - sign up now!

#TasoCon24

TASO

Refreshment break

11:00–11:30

Next: Breakout sessions - Methods made easy

#TasoCon24

TASO

8 May

09:30	Opening and welcome remarks
09:40	In conversation: Supporting disabled students in HE
10:00	Impact evaluation: Using quasi-experimental designs in higher education
11:00	Break
11:30	Breakout session: Methods made easy – Assessing the quality of evidence
11:30	Breakout session: Methods made easy – Randomised controlled trials
13:00	Lunch
14:00	Breakout session: Unlocking the evaluation toolbox – Post-entry Mapping Outcomes and Activities Tool
14:00	Breakout session: Unlocking the evaluation toolbox – Access and Success Questionnaire
15:30	Break
16:00	Navigating ethics in HE evaluation
16:30	Close

TASO

Annual Conference:

How to Evaluate

Assessing the quality of evidence

Tatjana Damjanovic / Research Officer ,TASO

#TasoCon24

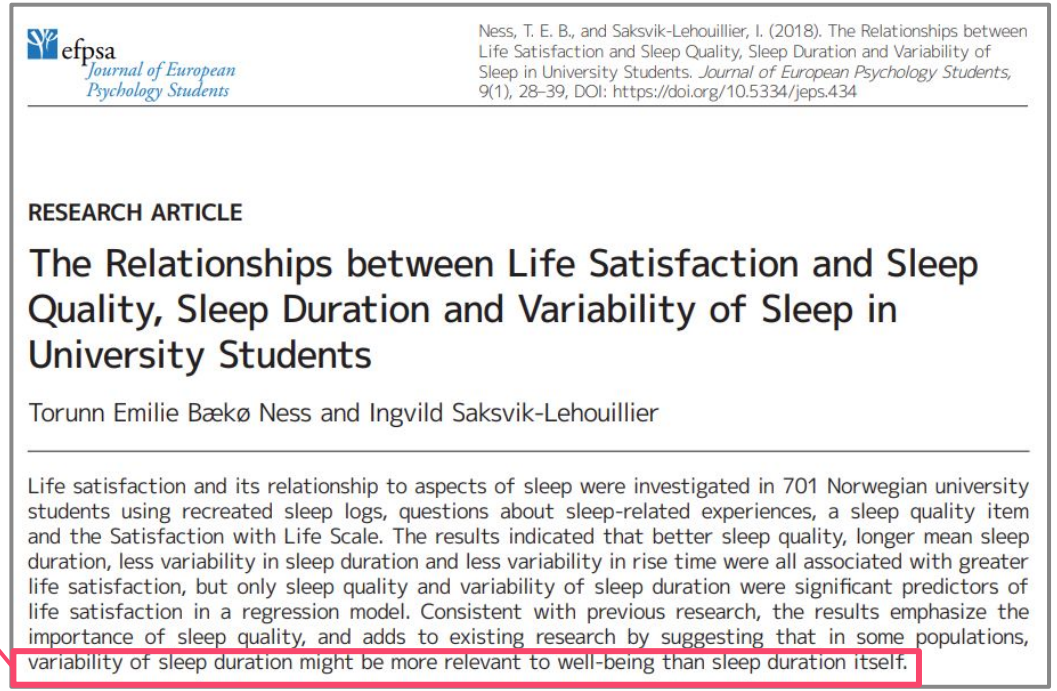
TASO

Overview of session

1. Types of evidence
2. Assessing Type 1 and 2 evidence
3. Assessing Type 3 evidence
4. Practical applications

Example 1

‘Variability of sleep duration might be more relevant to well-being than sleep duration itself’



efpsa Journal of European Psychology Students

Ness, T. E. B., and Saksvik-Lehouillier, I. (2018). The Relationships between Life Satisfaction and Sleep Quality, Sleep Duration and Variability of Sleep in University Students. *Journal of European Psychology Students*, 9(1), 28–39, DOI: https://doi.org/10.5334/jeps.434

RESEARCH ARTICLE

The Relationships between Life Satisfaction and Sleep Quality, Sleep Duration and Variability of Sleep in University Students

Torunn Emilie Bækø Ness and Ingvild Saksvik-Lehouillier

Life satisfaction and its relationship to aspects of sleep were investigated in 701 Norwegian university students using recreated sleep logs, questions about sleep-related experiences, a sleep quality item and the Satisfaction with Life Scale. The results indicated that better sleep quality, longer mean sleep duration, less variability in sleep duration and less variability in rise time were all associated with greater life satisfaction, but only sleep quality and variability of sleep duration were significant predictors of life satisfaction in a regression model. Consistent with previous research, the results emphasize the importance of sleep quality, and adds to existing research by suggesting that in some populations, variability of sleep duration might be more relevant to well-being than sleep duration itself.

Example 2

‘Among female students, both shortened and prolonged sleep durations were associated with increased likelihood of obesity. Among male students, there was no significant association’

Research Article | Open Access
Volume 2012 | Article ID 476914 | https://doi.org/10.1155/2012/476914

Show citation

Association of Sleep Duration with Obesity among US High School Students

Richard Lowry[1], Danice K. Eaton,[1] Kathryn Foti,[1] Lela McKnight-Eily,[2] Geraldine Perry,[2] and Deborah A. Galuska[3]

Abstract

Increasing attention is being focused on sleep duration as a potential modifiable risk factor associated with obesity in children and adolescents. We analyzed data from the national Youth Risk Behavior Survey to describe the association of obesity (self-report BMI ≥95th percentile) with self-reported sleep duration on an average school night, among a representative sample of US high school students. Using logistic regression to control for demographic and behavioral confounders, among female students, compared to 7 hours of sleep, both shortened (≤4 hours of sleep; adjusted odds ratio (95% confidence interval), AOR = 1.50 (1.05–2.15)) and prolonged (≥9 hours of sleep; AOR = 1.54 (1.13–2.10)) sleep durations were associated with increased likelihood of obesity. Among male students, there was no significant association between obesity and sleep duration. Better understanding of factors underlying the association between sleep duration and obesity is needed before recommending alteration of sleep time as a means of addressing the obesity epidemic among adolescents.

TASO

How do we know which study to trust?

Study 1

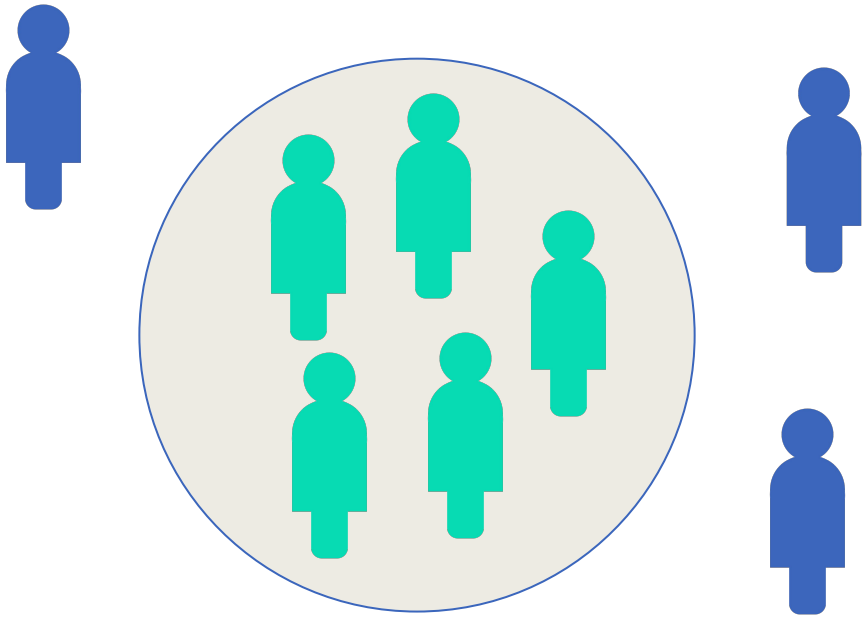
The Relationships between Life Satisfaction and Sleep Quality, Sleep Duration and Variability of Sleep in University Students

How do we know which study to trust?

Study 2

Association of Sleep Duration with Obesity among US High School Students

Why evidence quality matters



Evidence is central

Regulatory advice 6: The OfS also expects a provider to **access a number of useful sources of evidence that can be used to inform the intervention strategies** as well as the individual activities that sit within.

TASO

Not all studies are born equal

TASO

What do we want to know?

Is this study likely to be accurate?	Validity
Would this study have consistent results each time?	Reliability
Can we use its findings?	Generalisability

Types of evidence

Narrative evidence (Type 1)

Narrative or a coherent theory of change

“We have a coherent explanation of what we do and why”

“Our claims are research-based”

Types of evidence

Empirical evidence (Type 2)

Quantitative and/or qualitative evidence showing correlation

“We can demonstrate that our interventions are associated with beneficial results”

Narrative evidence (Type 1)

Narrative or a coherent theory of change

Types of evidence

“We believe our intervention causes improvement and can demonstrate the difference using a comparison group”

Causal evidence (Type 3)

Methodologies that show the causal effect of an intervention

Empirical evidence (Type 2)

Quantitative and/or qualitative evidence showing correlation

Narrative evidence (Type 1)

Narrative or a coherent theory of change

Narrative evidence (Type 1)

Description

The impact evaluation provides a narrative or a coherent theory of change to motivate its selection of activities in the context of a coherent strategy

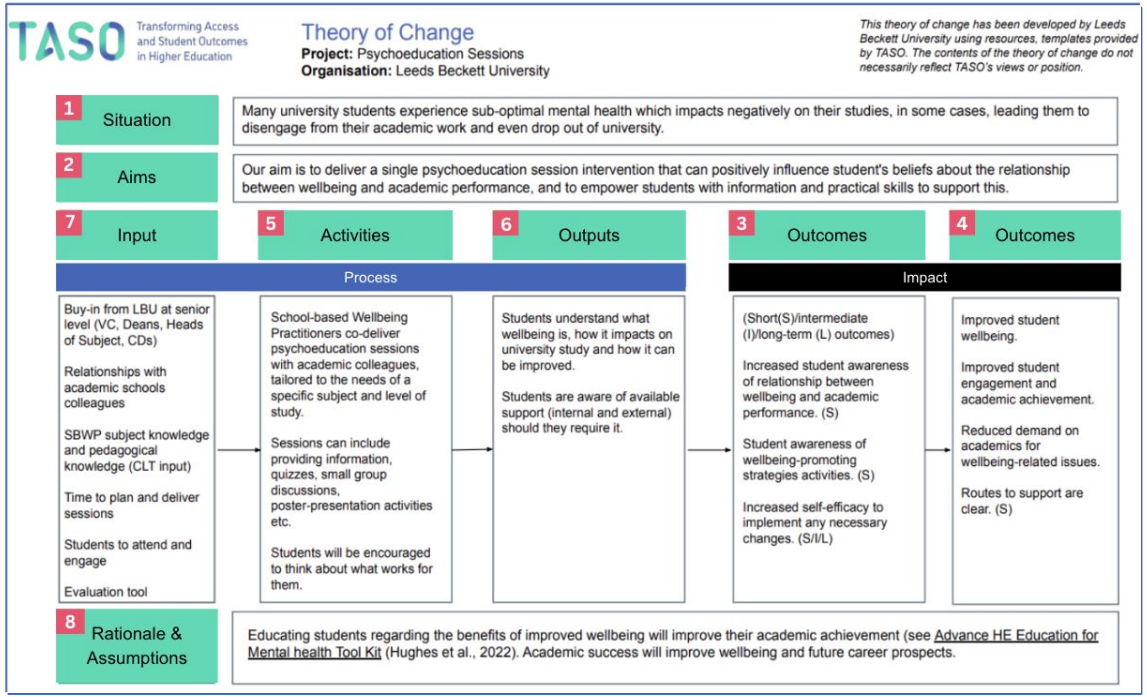
Evidence

Evidence of impact elsewhere and/or in the research literature or from your existing evaluation results

Claims you can make

- “We have a coherent explanation of what we do and why”
- “Our claims are research-based”

What counts as evidence?



TASO Transforming Access and Student Outcomes in Higher Education

Theory of Change

Project: Psychoeducation Sessions
Organisation: Leeds Beckett University

This theory of change has been developed by Leeds Beckett University using resources, templates provided by TASO. The contents of the theory of change do not necessarily reflect TASO's views or position.

1 Situation

Many university students experience sub-optimal mental health which impacts negatively on their studies, in some cases, leading them to disengage from their academic work and even drop out of university.

2 Aims

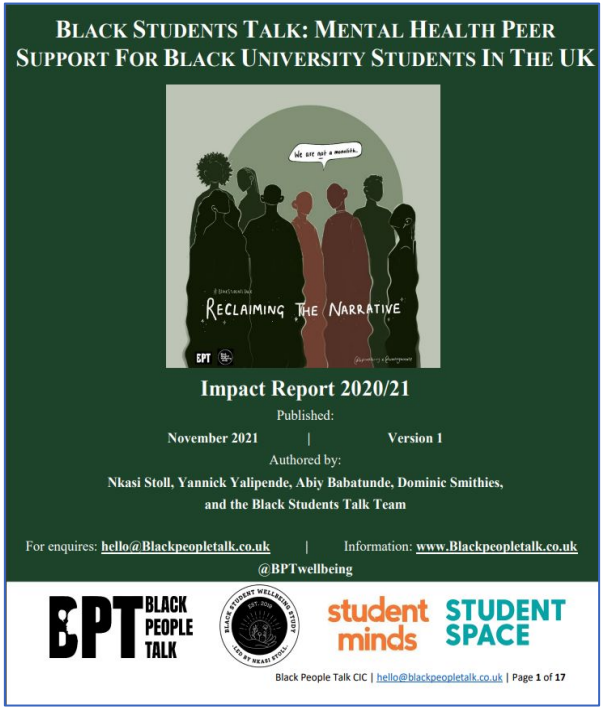
Our aim is to deliver a single psychoeducation session intervention that can positively influence student's beliefs about the relationship between wellbeing and academic performance, and to empower students with information and practical skills to support this.

7 Input	5 Activities	6 Outputs	3 Outcomes	4 Outcomes
	Process		Impact	
Buy-in from LBU at senior level (VC, Deans, Heads of Subject, CDs) Relationships with academic schools colleagues SBWP subject knowledge and pedagogical knowledge (CLT input) Time to plan and deliver sessions Students to attend and engage Evaluation tool	School-based Wellbeing Practitioners co-deliver psychoeducation sessions with academic colleagues, tailored to the needs of a specific subject and level of study. Sessions can include providing information, quizzes, small group discussions, poster-presentation activities etc. Students will be encouraged to think about what works for them.	Students understand what wellbeing is, how it impacts on university study and how it can be improved. Students are aware of available support (internal and external) should they require it.	(Short(S)/intermediate (I)/long-term (L) outcomes) Increased student awareness of relationship between wellbeing and academic performance. (S) Student awareness of wellbeing-promoting strategies activities. (S) Increased self-efficacy to implement any necessary changes. (S/I/L)	Improved student wellbeing. Improved student engagement and academic achievement. Reduced demand on academics for wellbeing-related issues. Routes to support are clear. (S)

8 Rationale & Assumptions

Educating students regarding the benefits of improved wellbeing will improve their academic achievement (see Advance HE Education for Mental health Tool Kit (Hughes et al., 2022). Academic success will improve wellbeing and future career prospects.

What counts as evidence?



BLACK STUDENTS TALK: MENTAL HEALTH PEER SUPPORT FOR BLACK UNIVERSITY STUDENTS IN THE UK

Impact Report 2020/21

Published:

November 2021 | **Version 1**

Authored by:

Nkasi Stoll, Yannick Yalipende, Abiy Babatunde, Dominic Smithies, and the Black Students Talk Team

For enquires: hello@Blackpeopletalk.co.uk | Information: www.Blackpeopletalk.co.uk

@BPTwellbeing

BPT BLACK PEOPLE TALK

student minds STUDENT SPACE

What counts as evidence?

Interpretative Phenomenological Analysis: A means of exploring aspiration and resilience amongst Widening Participation students

Authors: Gauntlett, Lizzie; Bickle, Ed; Thomas, Gail; Collins, Bethan; Heaslip, Vanessa; Eccles, Sue

Source: Widening Participation and Lifelong Learning, Volume 19, Number 2, May 2017, pp. 63-86(24)
Publisher: Open University
DOI: https://doi.org/10.5456/WPLL.19.2.63

Assessing the quality of narrative evidence

No theory of change. No engagement with literature or current debates. No clear link between intervention theory and outcomes.	Capturing qualitative data through interviews or focus groups with a small, targeted sample.	Capturing qualitative data through interviews or focus groups with a medium-sized sample and some thematic analysis of findings.	Capturing qualitative data through interviews with a medium-sized sample, conducting thematic analysis to extract latent themes and using methods to ensure the validity of findings (e.g. inter-rater testing; participant verification).

Weaker evidence → Stronger evidence

Assessing the quality of narrative evidence

No theory of change. No engagement with literature or current debates. No clear link between intervention theory and outcomes.	Capturing qualitative data through interviews or focus groups with a small, targeted sample	Capturing qualitative data through interviews or focus groups with a medium-sized sample and some thematic analysis of findings	Capturing qualitative data through interviews with a medium-sized sample, conducting thematic analysis to extract latent themes and using methods to ensure the validity of findings (e.g. inter-rater testing; participant verification)

Weaker evidence → Stronger evidence

Assessing the quality of narrative evidence

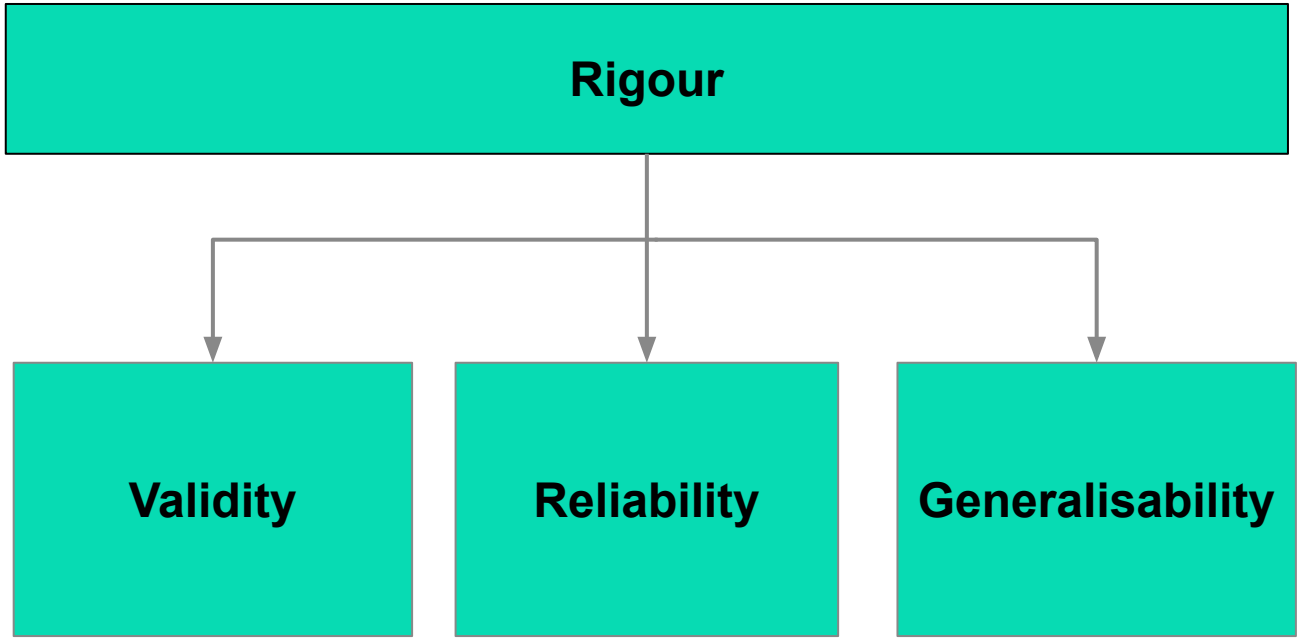
No theory of change. No engagement with literature or current debates. No clear link between intervention theory and outcomes	Capturing qualitative data through interviews or focus groups with a small, targeted sample	Capturing qualitative data through interviews or focus groups with a medium-sized sample and some thematic analysis of findings	Capturing qualitative data through interviews with a medium-sized sample, conducting thematic analysis to extract latent themes and using methods to ensure the validity of findings (e.g. inter-rater testing; participant verification)

Weaker evidence → Stronger evidence

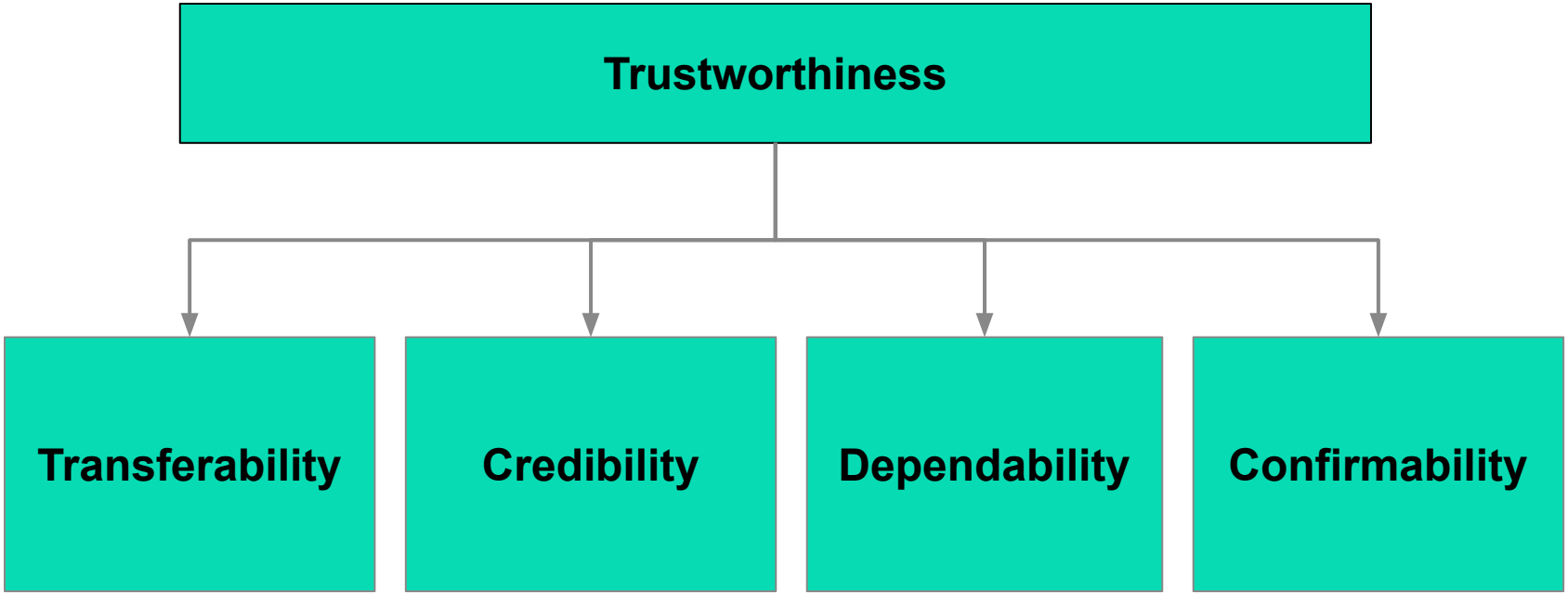
TASO

Rigour

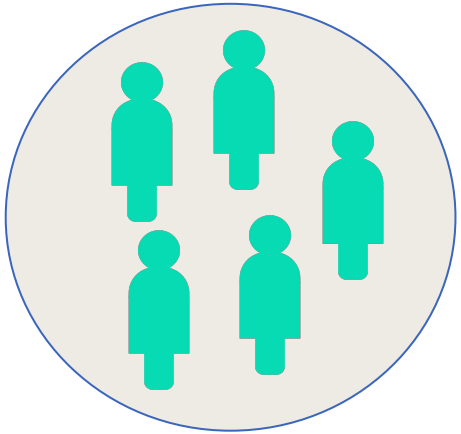
Rigour in quantitative research



Trustworthiness in qualitative research

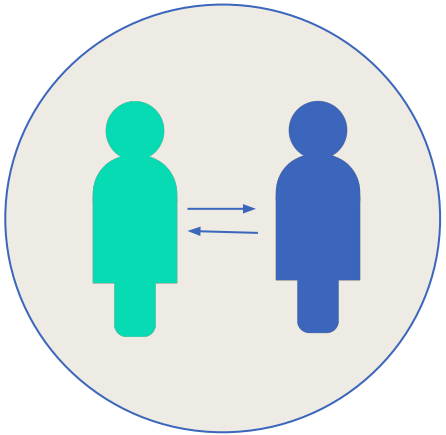


Assessing the quality of narrative evidence



Sampling

Appropriate analysis method



Inter-rater reliability

Example

FEEDBACK COLLECTION METHOD

To understand how Black students', describe their experiences of attending or facilitating Black Students Talk (BST), and designed an online qualitative survey on Google Forms to elicit the students' anonymous feedback. Attendees were asked by the facilitators after every session to fill out the survey. Facilitators were invited to fill out a survey specific to their role after every session by the project manager.

ANALYSIS

Thematic analysis (Braun and Clarke 2006) was conducted by and on the student's feedback which has been summarised below using illustrative quotes. conducted the write up of the feedback.

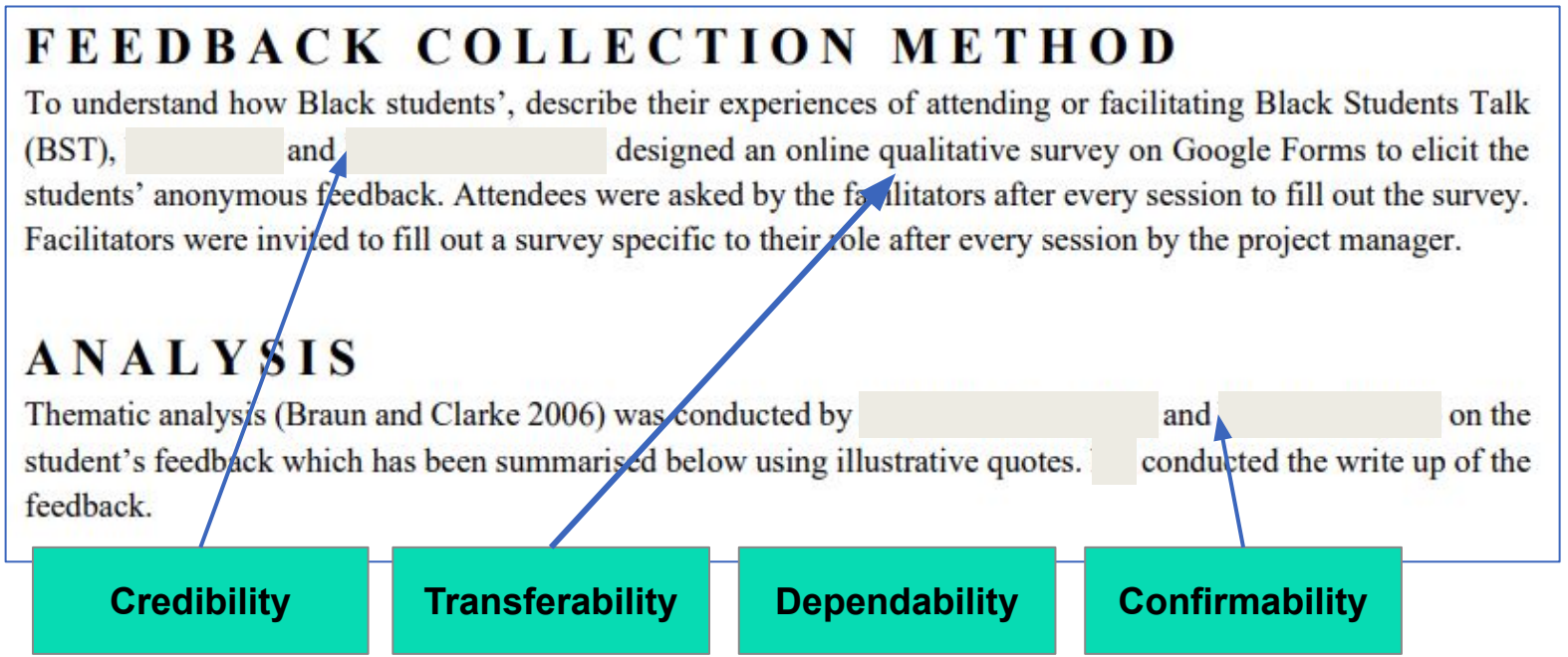
Example

FEEDBACK COLLECTION METHOD

To understand how Black students', describe their experiences of attending or facilitating Black Students Talk (BST), and designed an online qualitative survey on Google Forms to elicit the students' anonymous feedback. Attendees were asked by the facilitators after every session to fill out the survey. Facilitators were invited to fill out a survey specific to their role after every session by the project manager.

ANALYSIS

Thematic analysis (Braun and Clarke 2006) was conducted by and on the student's feedback which has been summarised below using illustrative quotes. conducted the write up of the feedback.



Assessing the quality of narrative evidence

No theory of change. No engagement with literature or current debates. No clear link between intervention theory and outcomes.	Capturing qualitative data through interviews or focus groups with a small, targeted sample	Capturing qualitative data through interviews or focus groups with a medium-sized sample and some thematic analysis of findings	Capturing qualitative data through interviews with a medium-sized sample, conducting thematic analysis to extract latent themes and using methods to ensure the validity of findings

Weaker evidence

Stronger evidence

Empirical Enquiry (Type 2)

Description

The impact evaluation collects data on impact and reports evidence that those receiving an intervention have better outcomes, though does not establish any direct causal effect

Evidence

Quantitative and/or qualitative evidence of a pre/post intervention change or a difference compared to what might otherwise have happened

Claims you can make

> We can demonstrate that our interventions are associated with beneficial results

Correlation versus causation

Eating more chocolate means more Nobel Prizes

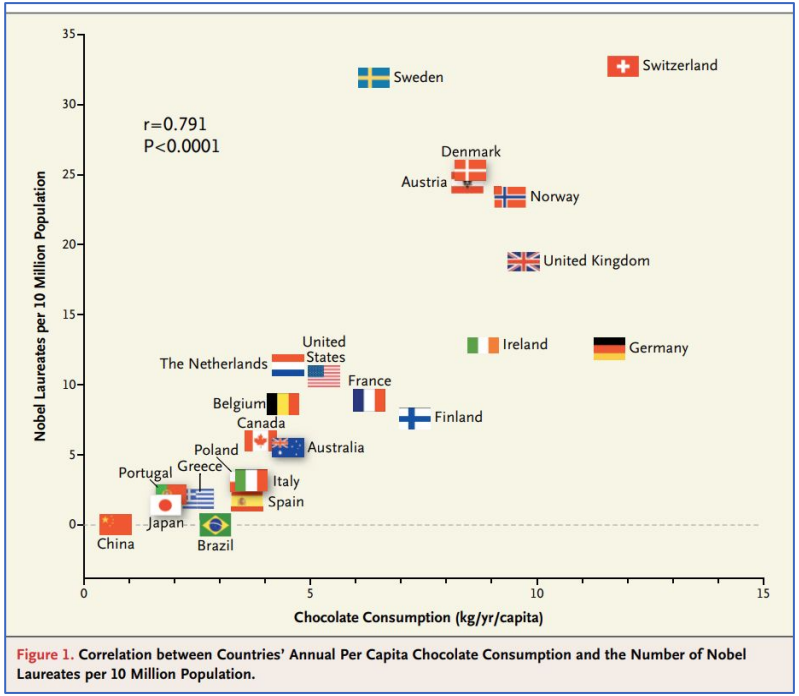


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Adapted from Messerli 2012

Correlation versus causation

National chocolate consumption correlates with the number of Nobel Laureates

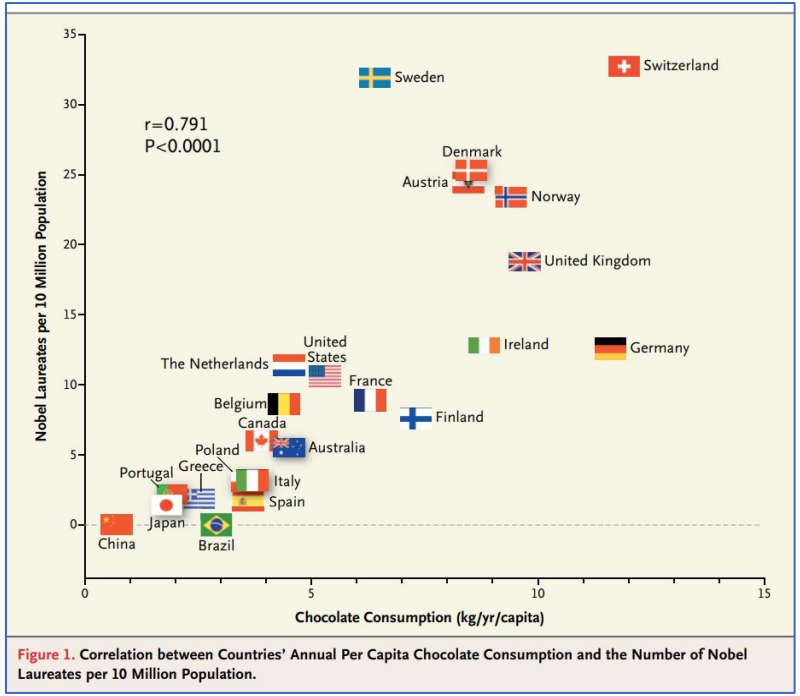


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Adapted from Messerli 2012

What counts as evidence?

RESEARCH DIRECTIONS

Managing your mind: how simple activities within the curriculum can improve undergraduate students' mental health and well-being

Eleanor R Edwards*, Heidrun Interthal, Heather A McQueen *

Institute of Cell Biology, School of Biological Sciences, The University of Edinburgh, King's Buildings, EH9 3FF

Assessing the quality of empirical evidence

Small sample. The data that is collected is not related to the aims of the intervention and data only collected at one time point.	Using quantitative data collection (e.g. surveys) to capture attitudes towards a programme.	Using quantitative data to capture attitudes or experiences before or after a programme, but without a control or comparison group.	Using pre- and post-intervention quantitative data to assess change in a validated instrument, but without the use of a comparison group.

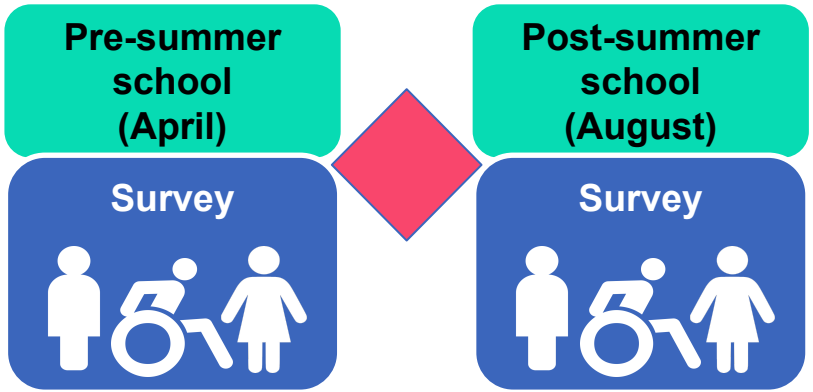
Weaker evidence → Stronger evidence

Assessing the quality of empirical evidence

Small sample. The data that is collected is not related to the aims of the intervention and data only collected at one time point.	Using quantitative data collection (e.g. surveys) to capture attitudes towards a programme.	Using quantitative data to capture attitudes or experiences before or after a programme, but without a control or comparison group.	Using pre- and post-intervention quantitative data to assess change in a validated instrument, but without the use of a comparison group.

Weaker evidence — Stronger evidence

Assessing the quality of empirical evidence

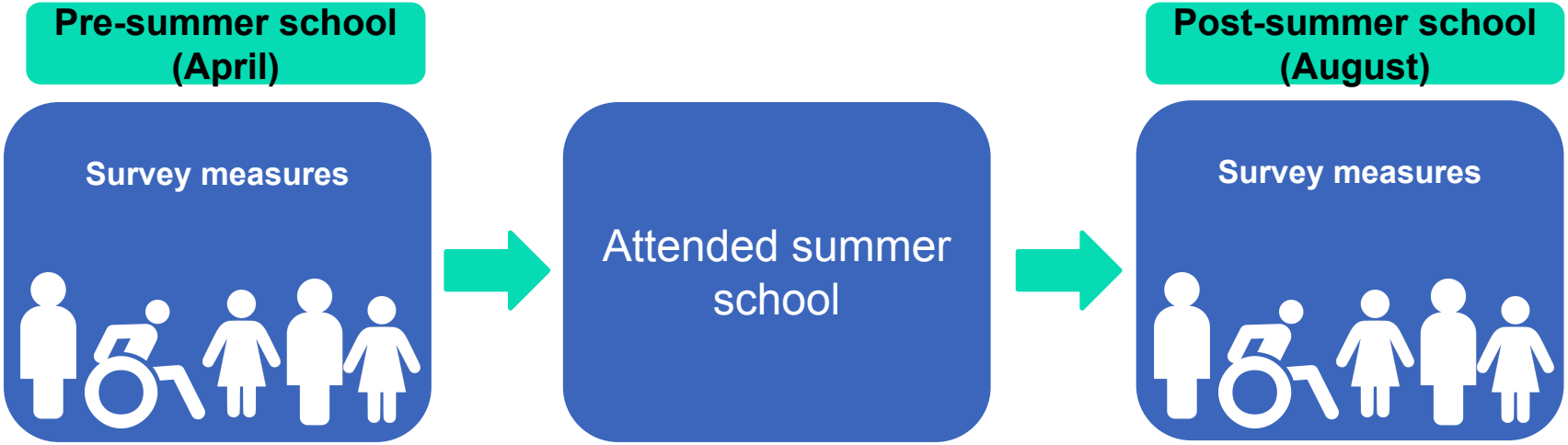


Pre and post-test measures

TASO Access and Success Questionnaire (ASQ)

Validated scales

TASO example: summer schools evaluation



TASO example: summer schools evaluation

Table 8: Estimated effects for the outcomes of interest

Outcome	Pre-summer-school mean	Post-summer-school mean	*t* statistic	Estimated effect (Cohen's *d*)
Likelihood of progressing to HE (7-point Likert scale)	6.42	6.65	2.16*	0.20
Self-efficacy relating to HE application (5-point Likert scale)	3.68	3.87	2.72**	0.24
Self-efficacy relating to post-entry success (5-point Likert scale)	3.94	4.01	1.18	0.10
Compatibility of HE with social identity (5-point Likert scale)	3.73	3.94	2.76**	0.23
Perception of financial barriers to HE (5-point Likert scale)	2.99	3.55	6.49***	0.58
Perception of knowledge barriers to applying to HE (5-point Likert scale)	3.13	3.94	9.06***	0.81
Notes: n = 142				
+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$				

TASO example: summer schools evaluation

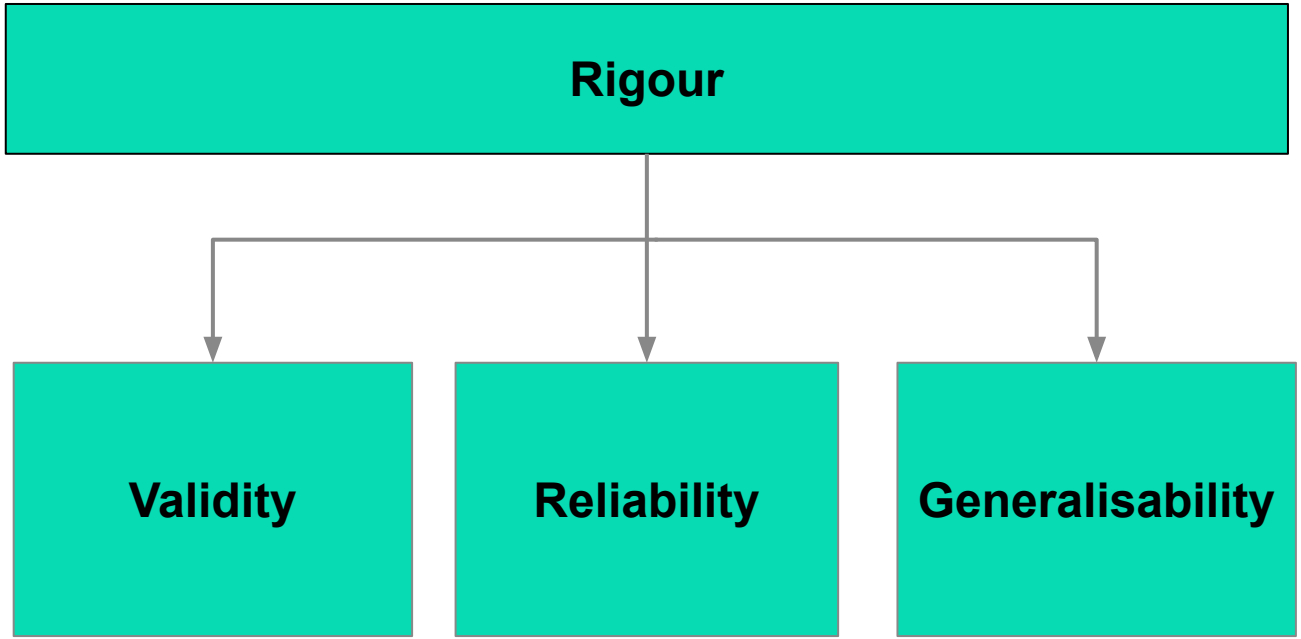
Table 8: Estimated effects for the outcomes of interest

Outcome	Pre-summer-school mean	Post-summer-school mean	*t* statistic	Estimated effect (Cohen's *d*)
Likelihood of progressing to HE (7-point Likert scale)	6.42	6.65	2.16*	0.20
Self-efficacy relating to HE application (5-point Likert scale)	3.68	3.87	2.72**	0.24
Self-efficacy relating to post-entry success (5-point Likert scale)	3.94	4.01	1.18	0.10
Compatibility of HE with social identity (5-point Likert scale)	3.73	3.94	2.76**	0.23
Perception of financial barriers to HE (5-point Likert scale)	2.99	3.55	6.49***	0.58
Perception of knowledge barriers to applying to HE (5-point Likert scale)	3.13	3.94	9.06***	0.81

Notes: n = 142

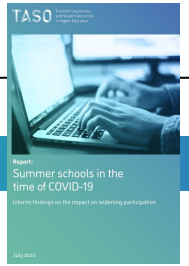
+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Rigour in quantitative research



Assessing the quality of empirical evidence

Small sample. The data that is collected is not related to the aims of the intervention and data only collected at one time point.	Using quantitative data collection (e.g. surveys) to capture attitudes towards a programme.	Using quantitative data to capture attitudes or experiences before or after a programme, but without a control or comparison group.	Using pre- and post-intervention quantitative data to assess change in a validated instrument, but without the use of a comparison group.



Weaker evidence → Stronger evidence

TASO

Activity

#TasoCon24

TASO

Attendee hub

https://taso.org.uk/taso-annual-conference-2024-attendee-guide/resources/

TASO

Over to you!

Please click on **link for task 1** of day two resources on the attendee hub: Assessing the quality of evidence

1. Which elements of this report are useful in deciding the strength of the evidence?

2. What can you find? What is missing?

TASO

Answers

There were four different versions of the signs and each drew on different principles from behavioural science - reciprocity bias, pro-social framing, endowment bias, time scarcity.

These **signs were installed on over one hundred fifty buses across Hammersmith, London (our pilot site)**, and the impact of the signs measured via on bus observations, interviews and surveys with drivers and interviews with passengers. Findings were compared with a similar set of 'control' buses that also operate in Hammersmith, but with no signs installed.

Answers

Weaker evidence

Stronger evidence

Causal evidence (Type 3)

Description

The impact evaluation methodology provides evidence of a *causal* effect of an intervention

Evidence

Quantitative and/or qualitative evidence of a pre/post treatment change on participants relative to an appropriate control or comparison group who did not take part in the intervention

Claims you can make

> "We believe our intervention causes improvement and can demonstrate the difference using a control or comparison group"

What counts as evidence?

A Mobile-Based Intervention to Increase Self-esteem in Students With Depressive Symptoms: Randomized Controlled Trial

Alina Bruhns [1]; Thies Lüdtke [1]; Steffen Moritz [1]; Lara Bücker [1]

Assessing the quality of causal evidence

Outcome measures aren't relevant to the activity, cross-contamination of treatment and comparator group.	A quasi-experimental study design with a small sample, quantitative pre- and post intervention data and a result that is only statistically significant after multiple corrections.	A randomised controlled trial design with a small sample, quantitative pre- and post intervention outcome data on a relevant construct and a statistically significant result with a small to medium effect size / conclusive zero effect.	A randomised controlled trial design with a large sample, quantitative pre- and post intervention outcome data captured for a relevant construct and a statistically significant result with a large effect size / conclusive zero effect.

Weaker evidence → Stronger evidence

Assessing the quality of causal evidence

Outcome measures aren't relevant to the activity, cross-contamination of treatment and comparator group.	A quasi-experimental study design with a small sample, quantitative pre- and post intervention data and a result that is only statistically significant after multiple corrections.	A randomised controlled trial design with a small sample, quantitative pre- and post intervention outcome data on a relevant construct and a statistically significant result with a small to medium effect size / conclusive zero effect.	A randomised controlled trial design with a large sample, quantitative pre- and post intervention outcome data captured for a relevant construct and a statistically significant result with a large effect size / conclusive zero effect.

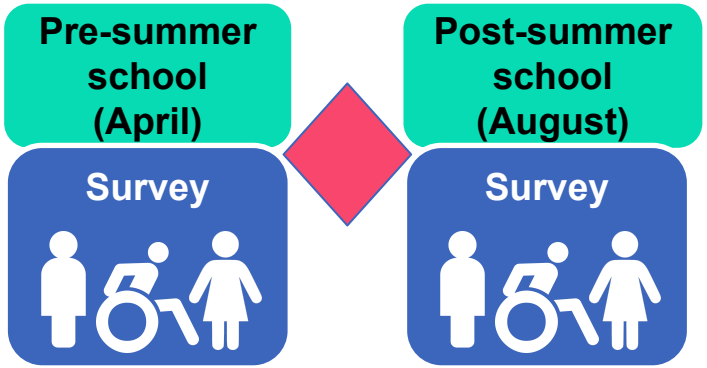
Weaker evidence → Stronger evidence

Assessing the quality of causal evidence

Outcome measures aren't relevant to the activity, cross-contamination of treatment and comparator group.	A quasi-experimental study design with a small sample, quantitative pre- and post intervention data and a result that is only statistically significant after multiple corrections.	A randomised controlled trial design with a small sample, quantitative pre- and post intervention outcome data on a relevant construct and a statistically significant result with a small to medium effect size / conclusive zero effect.	A randomised controlled trial design with a large sample, quantitative pre- and post intervention outcome data captured for a relevant construct and a statistically significant result with a large effect size / conclusive zero effect.

Weaker evidence → Stronger evidence

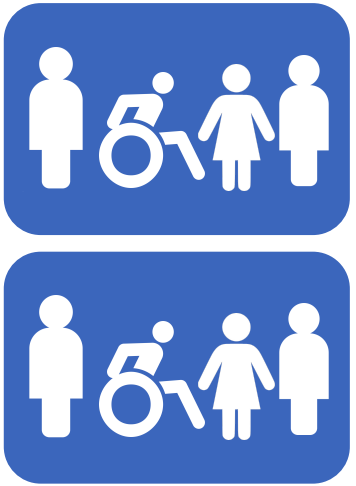
Assessing the quality of causal evidence



Pre-/ post-measures

TASO Access and Success Questionnaire (ASQ)

Validated scales



Comparison groups

Causal Evidence (Type 3)

Description

The impact evaluation methodology provides evidence of a *causal* effect of an intervention

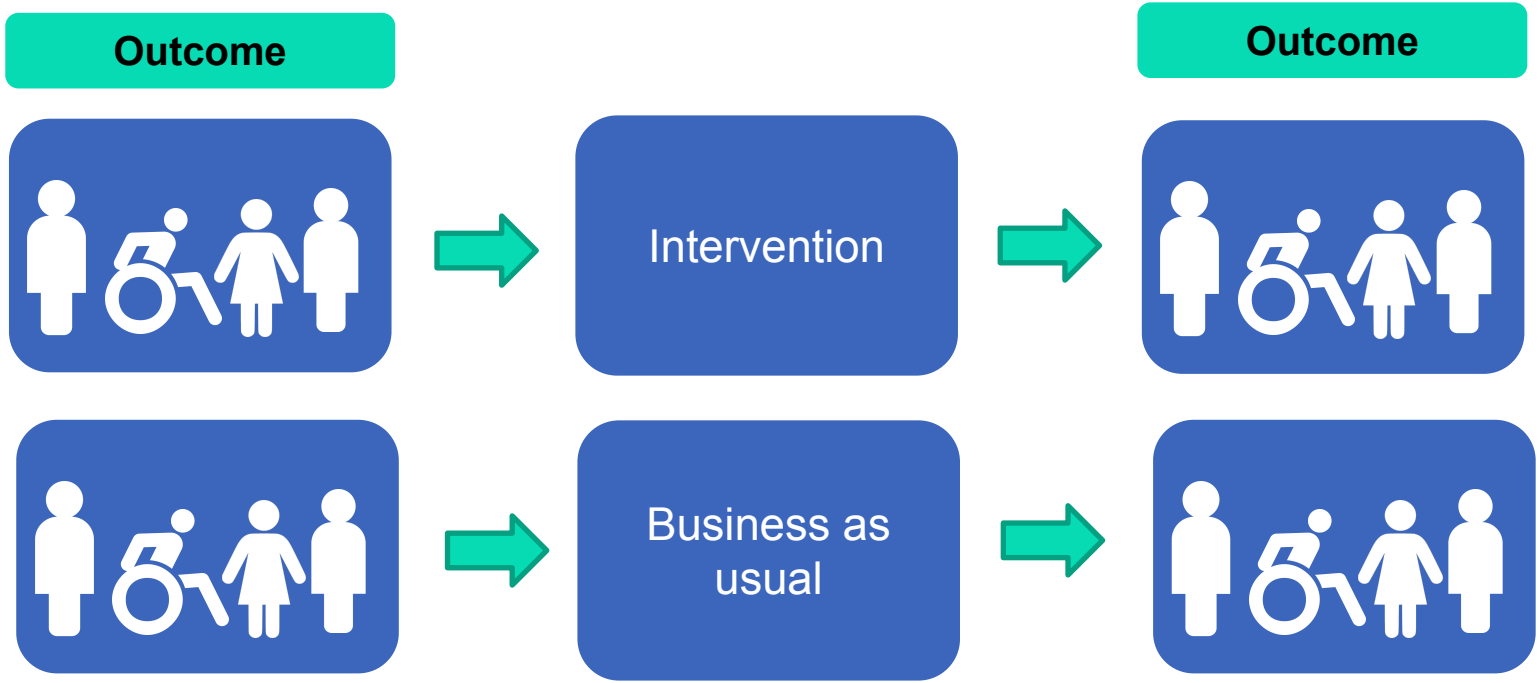
Evidence

Quantitative and/or qualitative evidence of a pre/post treatment change on participants relative to an appropriate control or comparison group who did not take part in the intervention

Claims you can make

“We believe our intervention causes improvement and can demonstrate the difference using a control or comparison group”

A matched comparison group

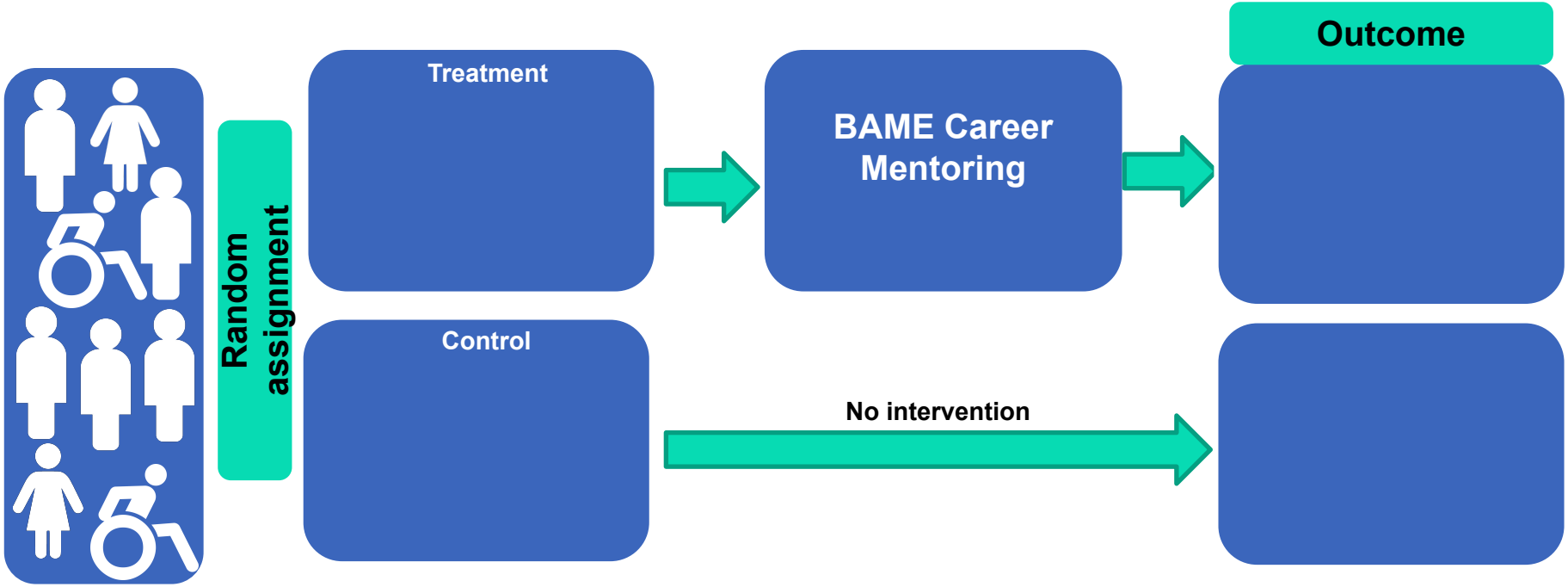


TASO

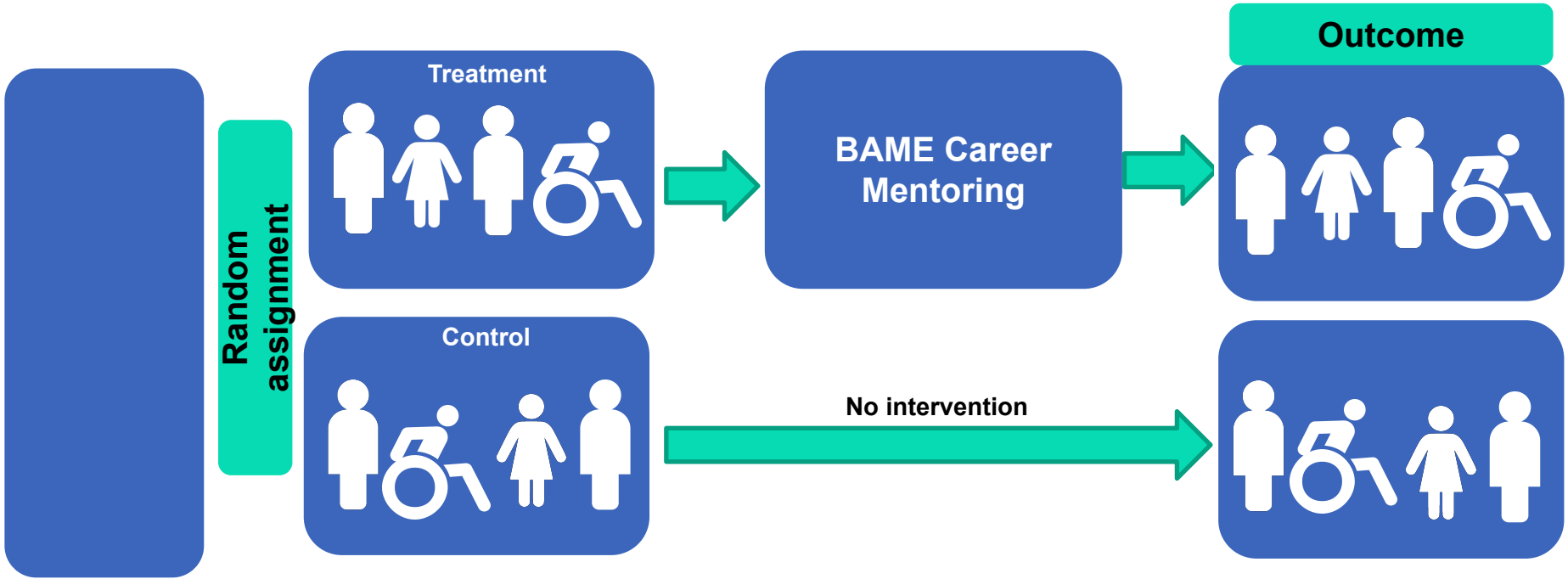
Comparing different groups

- **Demographic differences are matched**
- **What about unobservable differences?** (e.g. individual motivation)

Random comparison group



Random comparison group



Comparing different groups

- **Now it is valid to assume that the two groups are roughly similar on their observable and unobservable characteristics.**

- Therefore, observed differences in outcomes are **due to the intervention**, not pre-existing differences between groups.

Validity

Reliability

Generalisability

TASO

Activity

#TasoCon24

Over to you!

Please click on **link for Task 2** of day two resources on the attendee hub: Assessing the quality of evidence

1. Which elements of this report are useful in deciding the strength of the evidence?

Answers

- Large sample
- Validated scales use (PHQ-9, Rosenberg Self-esteem and WHO Quality of Life)
- Randomised comparison group with a waitlist control group
- Statistically significant results

Answers

A Mobile-Based Intervention to Increase Self-esteem in Students With Depressive Symptoms: Randomized Controlled Trial

Alina Bruhns [1]; Thies Lüdtke [1]; Steffen Moritz [1]; Lara Bücker [1]

Weaker evidence — Stronger evidence

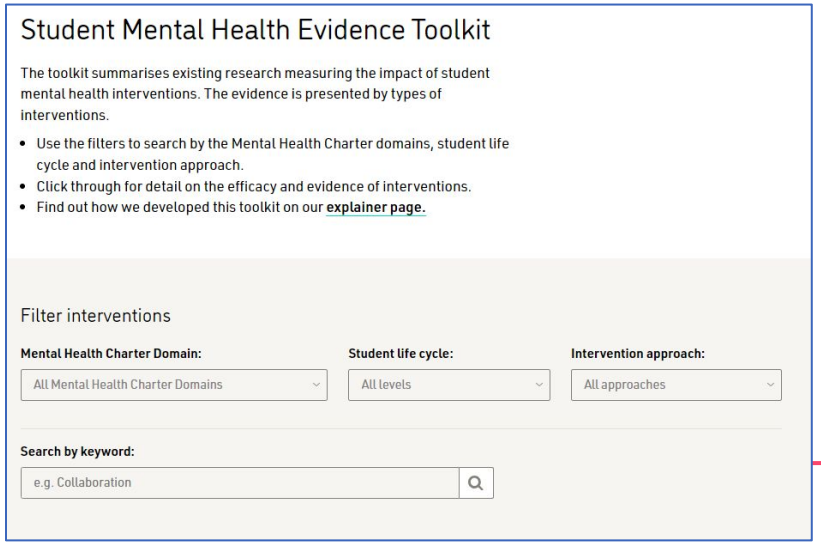
TASO

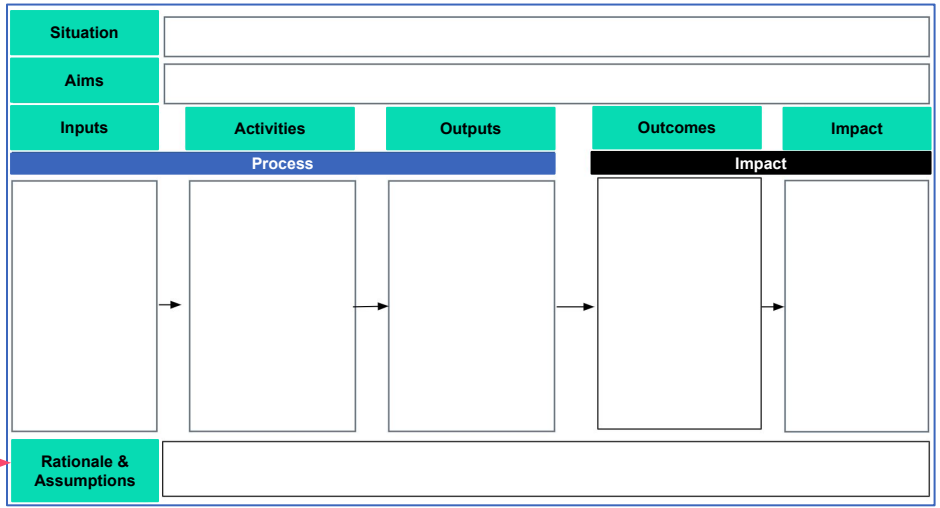
Evaluation is central

Annex B: Further information that sets out the rationale, assumptions and evidence base for each intervention strategy that is included in the access and participation plan.

This section should set out further information about the evidence used to underpin each intervention strategy, and any rationale and assumptions related to the underpinning theory of change for each intervention strategy.

Evaluation is central





Further resources

TASO Resources

- TASO Evidence Toolkit
- Student Mental Health Evidence Toolkit
- Rapid evidence review protocol template
- TASO evidence ratings

Other Resources

- PRISMA guidelines (for synthesising evidence)
- CASP checklist

TASO

Conclusion

TASO

Q&A

#TasoCon24

TASO

Lunch

13:00–14:00

Next: Breakout sessions – Unlocking the evaluation toolbox

#TasoCon24

TASO

8 May

09:30	Opening and welcome remarks
09:40	In conversation: Supporting disabled students in HE
10:00	Impact evaluation: Using quasi-experimental designs in higher education
11:00	Break
11:30	Breakout session: Methods made easy – Assessing the quality of evidence
11:30	Breakout session: Methods made easy – Randomised controlled trials
13:00	Lunch
14:00	Breakout session: Unlocking the evaluation toolbox – Post-entry Mapping Outcomes and Activities Tool
14:00	Breakout session: Unlocking the evaluation toolbox – Access and Success Questionnaire
15:30	Break
16:00	Navigating ethics in HE evaluation
16:30	Close

TASO
Annual
Conference:
How to Evaluate
#TasoCon24

TASO

Unpacking your evaluation toolbox: Access and success questionnaire

#TasoCon24

Who we are

Dr Katie Jones
Research and Evaluation Manager
The Brilliant Club

Rain Sherlock
Head of Evaluation
TASO

Overview of session

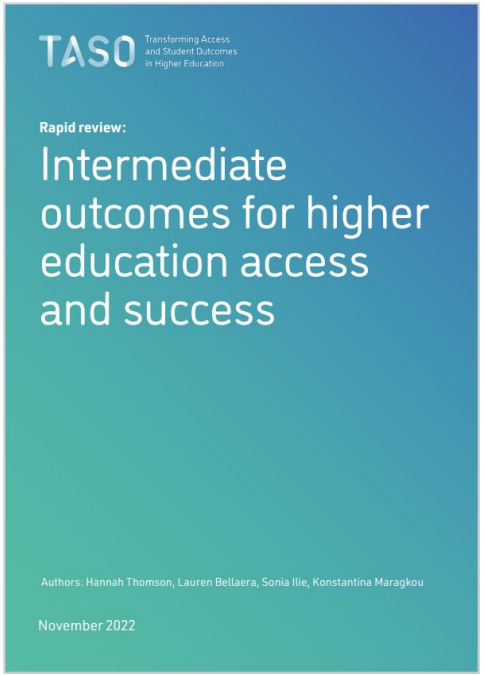
1. The role of intermediate outcomes in HE
2. Introducing the Access and Success Questionnaire
3. Using validated scales in practice
4. Group activity and Q&A

The role of intermediate outcomes in HE

#TasoCon24

What are long-term and intermediate outcomes?

- **Long-term outcomes** are measured using behavioural outcomes. In the case of WP, we often identify progression to university and success at university (e.g., retention and attainment) as the long-term outcomes.
- **Intermediate outcomes** refer to changes in behaviour, skills and attitudes that are associated with changes in long-term outcomes.



What types of intermediate outcomes matter?

Psychological constructs include:

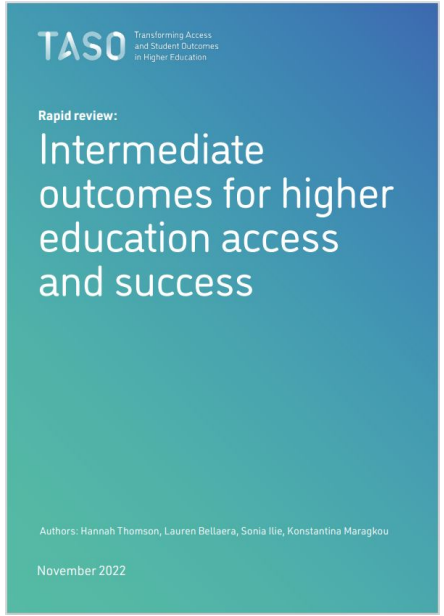
1. Cognitive and metacognitive outcomes: the mental processes that underpin learning (attention, memory, decision-making, self-regulation).

2. Motivational outcomes: the degree to which individuals are engaged in their learning, including in the face of setbacks.

3. Self-perceptions: individuals' feelings about their academic abilities and their levels of confidence.

4. Social outcomes: individuals' perceptions of how they belong in HE and the extent to which they belong at university.

What intermediate outcomes do you prioritise in your student access and success work?



How can we use scales to measure outcomes?

- To measure a psychological constructs, researchers often use questionnaires with multiple items.
- Items are added up to a score, and it is assumed that this score represents a person's position on the construct.
- It is important that the questionnaire has:
  - A good theory supporting the items you include in your scale.
  - A scale showing acceptable psychometric properties (e.g., reliability).
  - A scale related to other constructs in the ways hypothesised that captures group differences or causal processes expected to exist (e.g., in the case of WP, the outcome is associated with progression to or success at university).

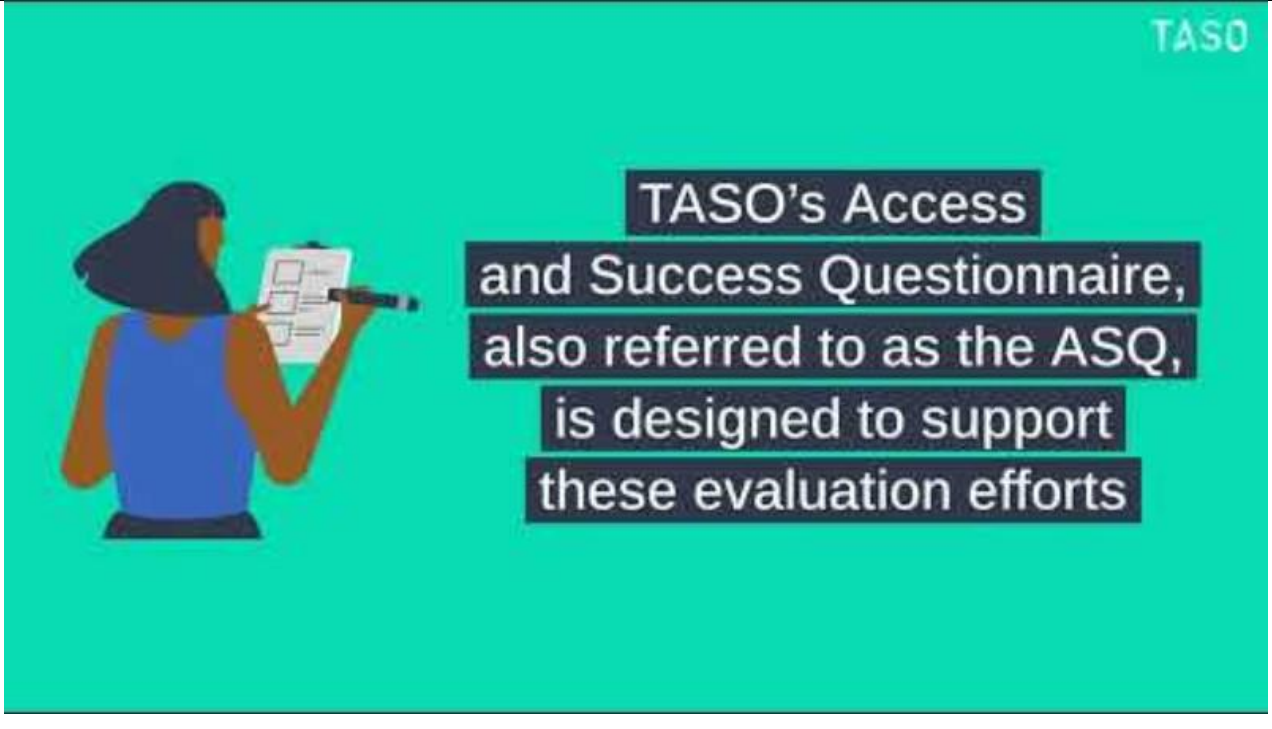
TASO

Introducing the Access and Success Questionnaire (ASQ)

#TasoCon24

TASO

Introducing the ASQ



Intermediate outcomes included in the ASQ

Outcome	Definition	Relevance
Academic self-efficacy	Individuals' confidence in their own ability to plan and execute the skills necessary to perform well academically in higher education.	Access (pre-entry)
Higher education expectations	The extent to which individuals expect to go to higher education.	
Knowledge of higher education	Individuals' knowledge about obtaining a place in higher education and what studying there might be like.	
Sense of belonging (pre-entry)	The extent to which individuals think they would feel connected to the higher education environment, peers, and others, if they were to progress to higher education.	
Cognitive strategies	The approaches individuals use to complete academic tasks and to prepare for and successfully learn.	Access (pre-entry) and Success (post-entry)
Metacognitive strategies	The approaches individuals use to monitor, plan and direct their own learning.	Success (post-entry)
Sense of belonging (post-entry)	The extent to which individuals feel connected to the higher education environment, peers, and others, and part of the community.	

Academic self-efficacy

Definition

Individuals' confidence in their own ability to plan and execute the skills necessary to perform well academically.

Measurement

The academic self-efficacy scale in the ASQ includes items such as “I have the academic ability to do well in higher education.”

Relevance to access and success

- In research, academic self-efficacy consistently emerges as one of the strongest correlates of attainment.
- There is a reciprocal relationship between academic self-efficacy and attainment, as students who attain highly are more likely to feel confident in their ability to do so again in future, which in turn boosts their attainment.
- Academic self-efficacy can be raised through interventions.

TASO

Higher education knowledge and expectations

Definition

Individuals' knowledge about how to apply to higher education, what higher education is like, and whether they expect to progress to higher education one day.

Measurement

The higher education knowledge scale in the ASQ includes items such as “I know what studying in higher education would be like."

Relevance to access and success

- It is important to ensure that all students have access to information about higher education, so they can make informed decisions about their future.
- However, this is unlikely to directly influence attainment, which is an important predictor of higher education progression.
- Additionally, students' expectations of whether they will progress to higher education is influenced by a complex set of factors that are not all within the scope of an intervention.

TASO

Sense of belonging

Definition

The extent to which individuals feel like they belong in higher education (post-entry), or would belong if they were to progress (pre-entry).

Measurement

The sense of belonging scale in the ASQ includes items such as “Higher education is for people like me."

Relevance to access and success

- University students with a higher sense of belonging are more likely to persist with their studies and have higher attainment.
- Studies from the US show that sense of belonging can be improved through interventions with university students from underrepresented backgrounds.
- Interim findings from TASO’s summer school evaluation indicate that access programmes can improve sense of belonging pre-entry

Cognitive strategies

Definition

The approaches individuals use to complete academic tasks and to prepare for and successfully learn.

Measurement

The cognitive strategies scale in the ASQ includes items such as “I can tell which information is most important when I study."

Relevance to access and success

- Using effective cognitive strategies has consistently been linked to higher attainment.
- However, studies have shown that students often do not use the most effective cognitive strategies, such as spreading studying out over time and practice testing.
- Teaching students how to use effective cognitive strategies can therefore contribute to raising attainment.

TASO

Metacognitive strategies

Definition

The approaches individuals use to monitor, plan and direct their own learning.

Measurement

The metacognitive strategies scale in the ASQ includes items such as “I can tell when I have understood a concept or idea."

Relevance to access and success

- At both school and university, metacognitive strategies contribute significantly to attainment.
- This link is found in both correlational and intervention studies across different subjects.
- Metacognition can be improved by teaching metacognitive strategies and showing students how to select an appropriate strategy for a specific task.

TASO

Scales included in ASQ

Scale*	No. of items	Example scale item	Age group
Academic self-efficacy	3	“I have the academic ability to do well in higher education.”	Y7-Y13 (age 11/12-17/18)
Higher education expectations	1	"I am thinking about going to higher education in the future."	
Knowledge of higher education	3	“I know what studying in higher education would be like."	
Sense of belonging (pre-entry)	3	“Higher education is for people like me."	
Cognitive strategies	4	“I can tell which information is most important when I study."	Y7-Y13 (age 11/12-17/18) and HE
Metacognitive strategies	5	“I can tell when I have understood a concept or idea."	HE
Sense of belonging (post-entry)	4	“I feel I belong in higher education”	

*Response options: Strongly disagree (1) – Disagree (2) – Neither agree nor disagree (3) – Agree (4) – Strongly agree (5)

TASO

The validation process

#TasoCon24

Why survey validation matters

- Intermediate outcomes are an important part of evaluating access and student success interventions – they are associated with positive changes in attainment, progression to and success in HE.
- If we don't have the tools to meaningfully evaluate these outcomes within our interventions with our target students, we will never truly know the impact of the work that we do.
- We have validated a multi-scale questionnaire for the sector that can be used to measure intermediate outcomes relevant to access and student success work.

TASO

Validation process: an overview

1. Identify and define the outcomes.
2. Assemble initial long-list of scales, with prompts and response options.
3. Test the scales with population of interest by asking them how they interpret them (cognitive testing).
4. Collect responses to scales with population of interest (pilot the scales) and analyse resulting data.
5. Interpret the analysis results and plan next steps.

Step 1 & 2: Identify and define the outcomes

IDENTIFYING OUTCOMES

Rapid Evidence Review and Sector Consultation

Survey & Focus Groups

Scales:

1. Higher education aspirations and expectations
2. Academic self-efficacy
3. Sense of belonging
4. Motivation
5. Growth mindset
6. Study skills
7. Metacognition

ANALYSING EXISTING DATA

Data from The Brilliant Club for sample of 23,000+ learners

Scales:

1. Academic self-efficacy
2. Sense of belonging, pre-entry
3. Cognitive strategies
4. Metacognitive strategies
5. Knowledge of higher

SHORT-LIST OF SCALES

Scale items derived from combined evidence from previous two steps and a review of existing scales.

Step 3: Test the scales (cognitive testing)

COGNITIVE TESTING (Round 1)

Qualitative approach with twelve participants, aged 14 to 22

Scales:

1. Academic self-efficacy
2. Sense of belonging, pre-entry
3. Sense of belonging, post-entry version 2
4. Cognitive study strategies
5. Metacognitive strategies
6. Critical engagement with information
7. Knowledge of higher education
8. Higher education intentions and expectations

REFINED WORDING OF SCALE ITEMS

Scale items refined based on insights from cognitive testing and earlier analysis.

Step 4: Collect responses and analyse results

TESTING SCALES WITH NEW SURVEY DATA
FIRST SURVEY
Sample of 386 young people in & out of education aged 16-22

SECOND SURVEY
Sample of 121 sixth-form learners

THIRD SURVEY
Sample of 52 higher education learners
Scales across the three surveys:
1. Academic self-efficacy
2. Sense of belonging, pre-entry
3. Sense of belonging, post-entry version 1
4. Sense of belonging, post-entry version 2
5. Cognitive study strategies
6. Metacognitive strategies
7. Critical engagement with information
8. Knowledge of higher education
9. Higher education intentions and expectations

PARTIALLY-VALIDATED SCALES

Scales and constituent items derived from combined evidence across all previous steps

Shared for trial use with Higher Education Providers.

Repeat steps 3 & 4 and complete step 5

TESTING SCALES WITH NEW HEP DATA

Scales trialled by Higher Education Providers with over 3,300 learners. Data provided directly or via the Higher Education Access Tracker (HEAT).

Scales:

1. Academic self-efficacy
2. Sense of belonging, prospective
3. Sense of belonging, post-entry
4. Cognitive study strategies
5. Metacognitive strategies
6. Critical engagement with information
1. Higher education knowledge and expectations

ANALYSING EXISTING DATA (Sub-group and younger age with data from The Brilliant Club)

Further analysis of existing data, focused on specific sub-groups (including defined by gender, first-in-family in higher education) and specifically looking at learners aged 11 to 12 only.

Scales: as per HEP data testing

COGNITIVE TESTING (Round 2)

Qualitative approach with six participants, aged 11 to 13.

Further minor refinements to a small number of items were made following this final round of cognitive testing.

Scales: as per HEP data testing

FINAL SCALES AND ITEMS The ASQ

Seven scales included in the ASQ

- Academic self-efficacy
- Knowledge of higher education
- Cognitive strategies
- Sense of belong (pre-entry)
- Higher education expectations
- Metacognitive strategies
- Sense of belonging (post-entry)

TASO

Using validated scales in practice

#TasoCon24

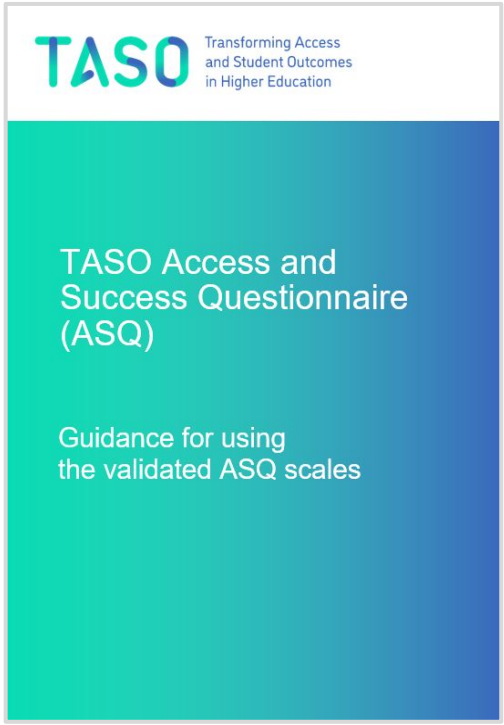
How to use the ASQ

STEP 1: IDENTIFY THE PROGRAMME AND ITS OUTCOMES

- For guidance about how to identify the best outcome measure for a programme, check the TASO Monitoring and Evaluation Framework and consider using the Mapping Outcomes and Activities Tool (MOAT).

STEP 2: DESIGN THE EVALUATION

- Decide what kind of evidence the evaluation will generate. Validated scales are most useful for generating Type 2 (empirical) or Type 3 evidence (Causal). For example, you may use the ASQ as part of a pre-post survey design to generate Type 2 evidence.



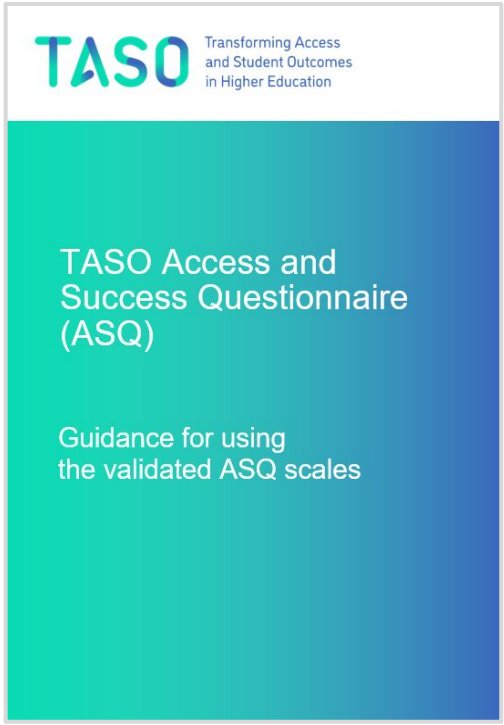
How to use the ASQ

STEP 3: IDENTIFY RELEVANT VALIDATED SCALES

- Specify in your evaluation plan precisely which validated scale you will use, and how you will collect data with it. This includes outlining when the scale will be administered, to whom, and whether on paper or in an online format.

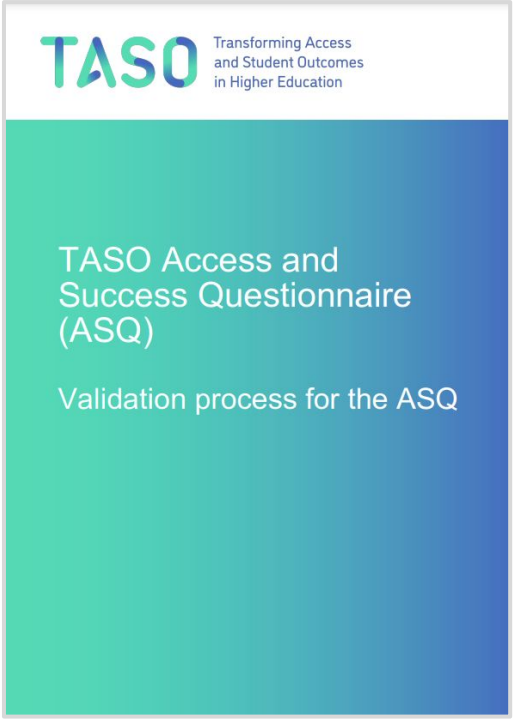
STEP 4: CARRY OUT THE EVALUATION AND RECORD THE DATA

- Record the data you have collected using the ASQ validated scales. You may use the ASQ spreadsheet, that allows for input of any data you may have collected using ASQ validated scales.
- HEAT now includes a facility to collect and upload data from any of the ASQ validated scales.



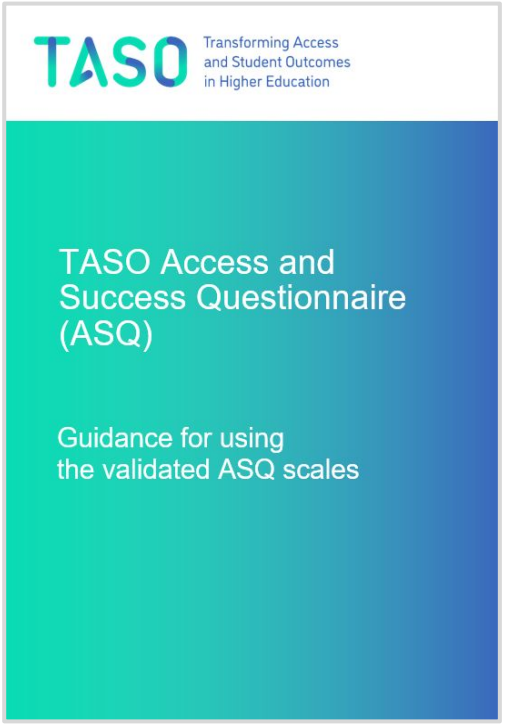
What to do if we cannot use a validated scale?

- Make modest changes to the wording and structure of existing scales to improve their usability in your context.
- Consider developing bespoke, validated, scales specifically for the constructs you are interested in.
- Use TASO guidance on how to develop validated scales.



How is the sector using the ASQ?

- In Access and Participation Plans to measure intermediate outcomes for access and student success intervention strategies
- Via the Higher Education Access Tracker (HEAT) survey tool
- Via HEP administered surveys as part of their evaluation activity
- As the basis for developing additional bespoke validated scales



Project collaborators

Researchers from the University of Cambridge and The Brilliant Club collaborated with TASO on the Survey Validation Project

Sonia Ilie

Konstantina Maragkou

Lauren Bellaera

Hannah Thomson

TASO

Group activity

#TasoCon24

TASO

Attendee hub

https://taso.org.uk/taso-annual-conference-2024-attendee-guide/resources/

Using the ASQ in your work

1

Start by identifying the programme or activity you'd like to evaluate

2

Match your desired outcome with the relevant validated questionnaire scale

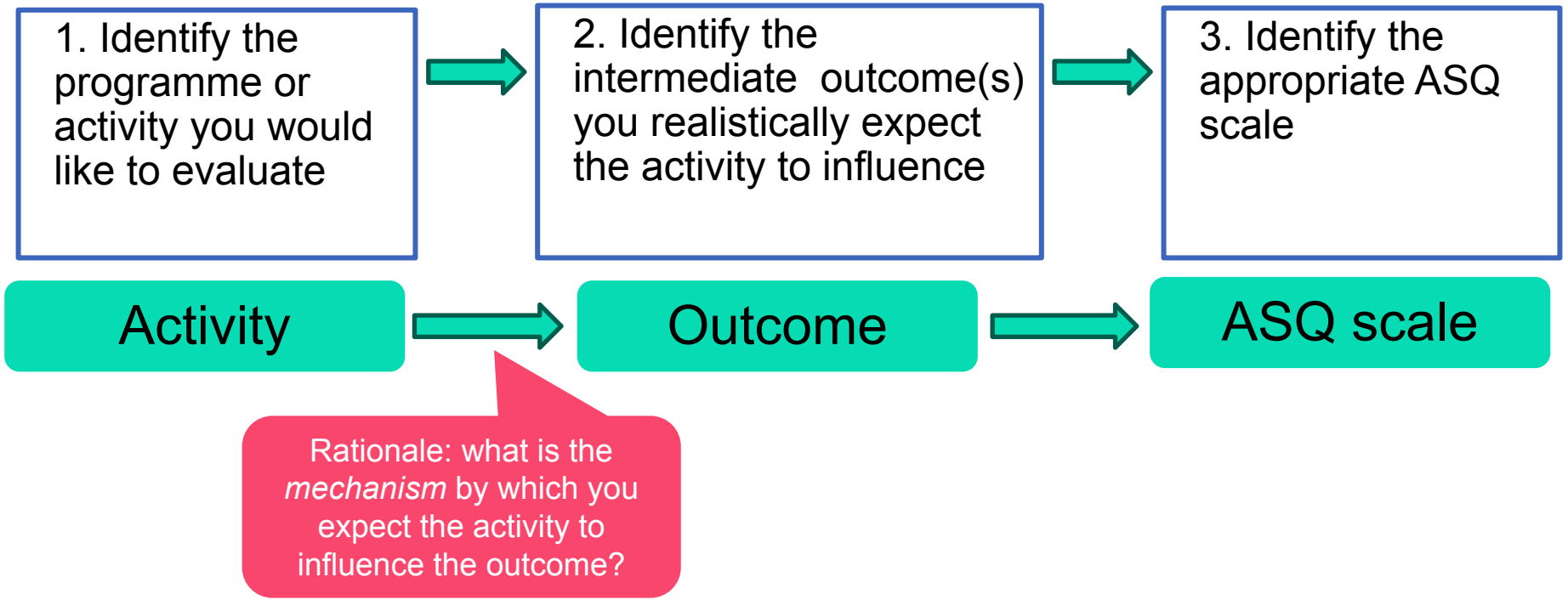
3

Use the scale in a survey of participants (and any comparison group if applicable)

4

Collect and analyse the data

Using the ASQ in your work



TASO

Q&A

#TasoCon24

Refreshment break

15:30–16:00

Next: Navigating ethics in HE evaluation

#TasoCon24

TASO

8 May

09:30	Opening and welcome remarks
09:40	In conversation: Supporting disabled students in HE
10:00	Impact evaluation: Using quasi-experimental designs in higher education
11:00	Break
11:30	Breakout session: Methods made easy – Assessing the quality of evidence
11:30	Breakout session: Methods made easy – Randomised controlled trials
13:00	Lunch
14:00	Breakout session: Unlocking the evaluation toolbox – Post-entry Mapping Outcomes and Activities Tool
14:00	Breakout session: Unlocking the evaluation toolbox – Access and Success Questionnaire
15:30	Break
16:00	Navigating ethics in HE evaluation
16:30	Close

TASO
Annual
Conference:
How to Evaluate
#TasoCon24

Navigating ethics in HE evaluation

#TasoCon24

Speakers

Dr Richard Davies
HE Research and Development Lead
University of Central Lancashire

Rain Sherlock
Head of Evaluation
TASO

Why should we be thinking about ethics?

OfS Regulatory advice 6

As a provider further develops its evaluation strategies, the OfS expect it to consider how it intends **recording, publishing and sharing its evaluation activity and findings**.

TASO's Research Ethics Guidance

With thanks to

Dr Richard Davies
Dr Cherry Canovan
Dr Peter Lucas
Prof Andrea Manfrin
Dr Terigele

Access the guidance

https://taso.org.uk/evidence/research-ethics-guidance

TASO

Overview

1. Ethics, Evaluation and Research
2. Couple of case studies
3. Important considerations
4. Some thoughts on process

TASO

Evaluation and Research

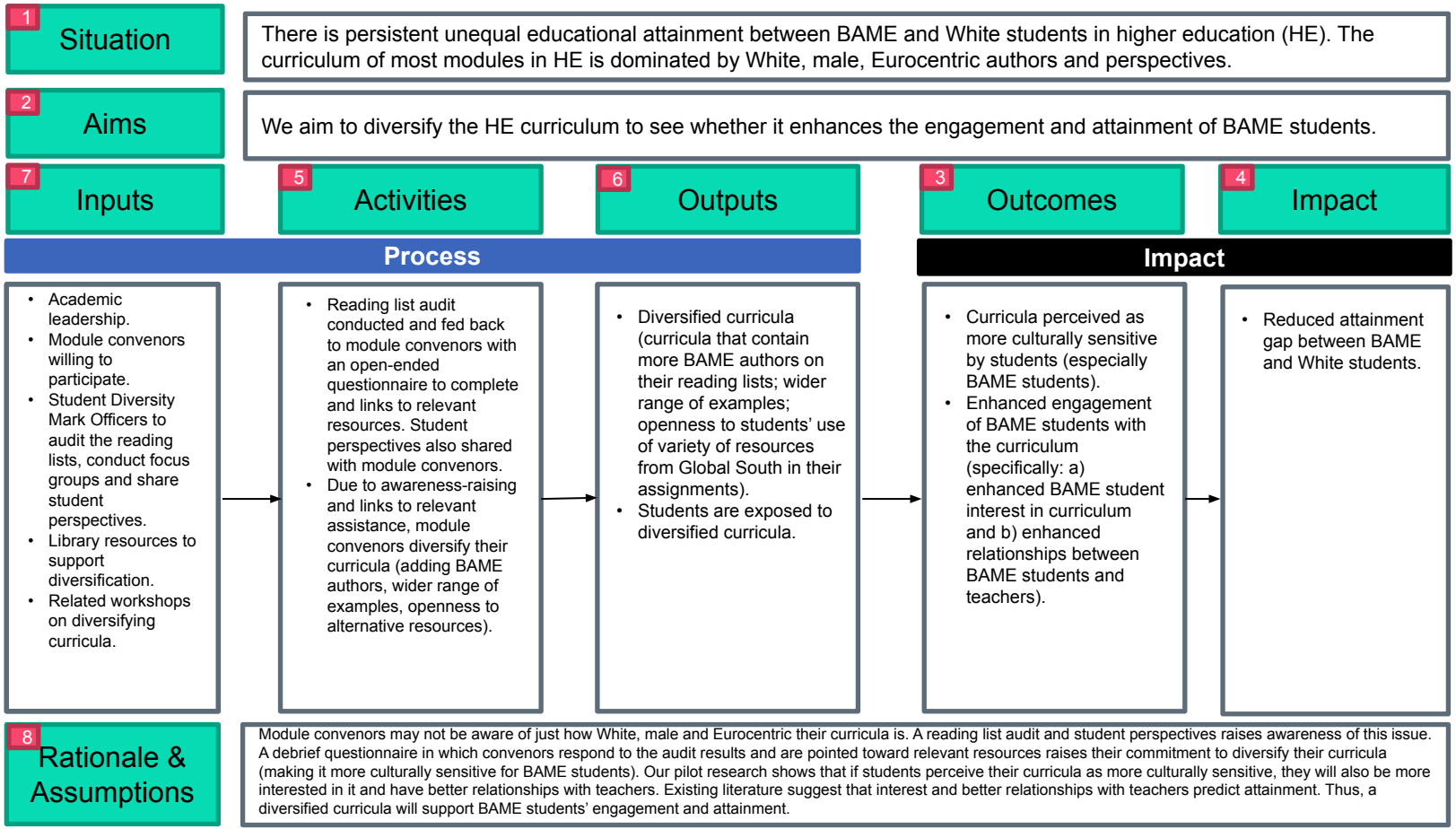
#TasoCon24

Audit	Research
Never involves experiments[1] on human subjects	May involve experiments on human subjects
Is a systematic approach to the peer review of educational support to identify opportunities for improvement and to provide a mechanism for bringing them about	Is a systematic investigation that aims to increase the sum of knowledge
Never involves allocating people randomly to different treatment groups	May involve allocating people randomly to different treatment groups
Never involves a completely new intervention	May involve a completely new intervention
Never involves disturbance to the participants beyond that required for normal educational activity	May involve extra disturbance or work beyond that required for normal educational activity
May involve students/pupils with the same educational need being given different interventions, but only after full discussion of the known advantages and disadvantages of each approach. The student/pupils can choose freely which intervention they receive	May involve the application of strict selection criteria to students/pupils with the same educational need before they are entered into the research study
Measures against a standard set of expectations to address educational needs	Usually involves an attempt to test a hypothesis
Is used to inform internal policy and practice with limited circulation to those involved in the activities	Is intended to be communicated to a broad audience beyond the individuals and institutions directly involved

TASO

Case studies

#TasoCon24


1
Situation
There is persistent unequal educational attainment between BAME and White students in higher education (HE). The curriculum of most modules in HE is dominated by White, male, Eurocentric authors and perspectives.
2
Aims
We aim to diversify the HE curriculum to see whether it enhances the engagement and attainment of BAME students.
7
Inputs
5
Activities
6
Outputs
3
Outcomes
4
Impact
Process
Impact
• Academic leadership.
• Module convenors willing to participate.
• Student Diversity Mark Officers to audit the reading lists, conduct focus groups and share student perspectives.
• Library resources to support diversification.
• Related workshops on diversifying curricula.
• Reading list audit conducted and fed back to module convenors with an open-ended questionnaire to complete and links to relevant resources. Student perspectives also shared with module convenors.
• Due to awareness-raising and links to relevant assistance, module convenors diversify their curricula (adding BAME authors, wider range of examples, openness to alternative resources).
• Diversified curricula (curricula that contain more BAME authors on their reading lists; wider range of examples; openness to students' use of variety of resources from Global South in their assignments).
• Students are exposed to diversified curricula.
• Curricula perceived as more culturally sensitive by students (especially BAME students).
• Enhanced engagement of BAME students with the curriculum (specifically: a) enhanced BAME student interest in curriculum and b) enhanced relationships between BAME students and teachers).
• Reduced attainment gap between BAME and White students.
8
Rationale & Assumptions
Module convenors may not be aware of just how White, male and Eurocentric their curricula is. A reading list audit and student perspectives raises awareness of this issue. A debrief questionnaire in which convenors respond to the audit results and are pointed toward relevant resources raises their commitment to diversify their curricula (making it more culturally sensitive for BAME students). Our pilot research shows that if students perceive their curricula as more culturally sensitive, they will also be more interested in it and have better relationships with teachers. Existing literature suggest that interest and better relationships with teachers predict attainment. Thus, a diversified curricula will support BAME students' engagement and attainment.

Example one

Outcomes
• Curricula perceived as more culturally sensitive by students (especially BAME students). • Enhanced engagement of BAME students with the curriculum (specifically: a) enhanced BAME student interest in curriculum and b) enhanced relationships between BAME students and teachers).

TASO

1 Situation

Student mental health and wellbeing is in decline, exacerbated by the COVID-19 pandemic. Students desire someone to talk to above any other form of support. Students may be low- or non-engaging for a variety of reasons, including poor mental health and well-being. The Student Engagement Dashboard already effectively identifies low- and non-engagers, and the Contact and Engagement Service (CES) can then initiate a conversation with them about their lack of engagement.

2 Aims

We aim to coach low- and non-engaging students to develop self-efficacy, and to signpost these students to relevant support services if necessary. This will motivate and empower them, leading to an increase in academic engagement, with a resulting increase in attainment and progression rates for these students.

Process

Impact

7 Inputs

- Student participants
- Calling team
- CES Coordinators
- Academic Tutors
- Student support staff
- Budget
- Student Engagement Dashboard
- Administrative Data

5 Activities

- Engagement alert generated
- Email sent to student
- Coaching telephone call with CES
- Follow-up via email and dashboard annotation

6 Outputs

- Student self-efficacy
- Raised student motivation
- Raised sense of student empowerment
- Improved student knowledge of university

3 Outcomes

- Sense of belonging at the university
- Being an active member of the campus community
- Accessing support when needed

4 Impact

- Increased academic engagement
- Raised student attainment
- Increased student progression

8 Rationale & Assumptions

According to research, learning analytics provides an effective platform from which early alert systems for low engagement can be implemented. Moreover, coaching approaches have been seen to increase student progression. We assume the following: students engage with the telephone call; the telephone call leads students to change behaviour over both short and long term; and changed behaviour patterns (inc. engagement levels) result in higher levels of progression and attainment.

TASO

Example two

- The evaluation is more difficult to operationalise
- It most likely involves quantitative data analysis
  - ***These have ethical as well as quality implications***

- Deals with topics students might find disturbing
  - ***Requires sensitive handling and signposting to well-being support***
- Students need to know what is required and why it is worth doing
  - ***Requires clear messaging to ensure informed consent***

TASO

Important considerations

#TasoCon24

TASO

Important considerations

- Level of risk
- Lack of experience
- Institution level effects

TASO

Some thoughts on process

#TasoCon24

TASO

Some thoughts on process

- Clearly separate *the activities* from *the evaluation*
- Clearly limit expedited review for low-risk evaluations only
- Be clear about what is data collection and secondary use of existing data
- Potential risks are considered, recorded, and mitigated
- Data ought to be securely managed
- If in doubt – go for a full review

TASO

Q&A

#TasoCon24

Contact

Dr Richard Davies

Centre for Collaborative Learning
University of Central Lancashire

rdavies15@uclan.ac.uk

@RichardTaff

TASO

Feedback survey

#TasoCon24

TASO

Thank you for joining us!

#TasoCon24