TASO Transforming Access and Student Outcomes in Higher Education

TASO Implementation and process evaluation (IPE) guidance

April 2024

NTU Nottingham Trent University Psychology

Acknowledgements

TASO would like to acknowledge:
Dr Emma J Vardy and Lydia F Smith, Nottingham Trent University
Rain Sherlock, TASO

Contents

Glossary of terms

Implementation and process evaluation (IPE)	IPE includes the generation and analysis of data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes, and the factors that influence these processes.
Impact evaluation	This type of evaluation asks, 'Did the intervention work?' Impact is the portion of an outcome change that can be attributed to the intervention or programme rather than other factors or influences. It is used to help decide whether an intervention or programme should be adopted, continued or modified for improvement.
Economic evaluation	The comparison of the value of the outcomes produced by an intervention with the costs of implementing the intervention.
Theory of change (ToC)	A ToC describes the underlying assumptions and mechanisms of how the planned activities will lead to the intended outcomes.
Effectiveness trial	This type of trial tests whether an intervention worked under real-life conditions. Effectiveness trials aim to evaluate performance in everyday practice.
Active ingredients	An intervention or programme contains multiple components. The active ingredients are those components of an intervention or programme that contribute to the outcomes reported.
Randomised controlled trial (RCT)	In an RCT, participants are randomly assigned to treatment (intervention) and control (business as usual) groups. Random assignment ensures a high degree of confidence that there are no systematic differences between the treatment and control groups except that the treatment group participated in the intervention.
Implementation fidelity	This refers to the degree to which an intervention or programme has been delivered as intended.

Introduction to IPE Guidance

There is at present no universal approach to conducting implementation and process evaluation (IPE) in the higher education (HE) sector. We use the Education Endowment Foundation's (EEF) definition of IPE – '*the generation and analysis of data to examine how an intervention is put into practice, how it operates to achieve its intended outcomes, and the factors that influence these processes*' (Humphrey et al., 2016) – to inform this guidance.

This guidance aims to support colleagues in the HE sector to plan, conduct and report high-quality IPE work. Its content has been prepared by consulting a wide range of existing IPE guidance and approaches, including literature outside the field of education. It contains an IPE framework, data collection tools and analytical tools to inform your IPE work.

For a comprehensive understanding of IPE and its importance to evaluation work, this guidance document should be read alongside other TASO evaluation guidance:

1. The Monitoring and Evaluation Framework: this guidance outlines a step-by-step guide for effective impact evaluation.
2. Theory of change (ToC) guidance: this provides the reader with an understanding of ToC, an important step before commencing an IPE.
3. Economic evaluation guidance: this focuses on comparing the value of the outcomes produced by an intervention with the costs of implementation.

This IPE guidance is to be used in conjunction with the TASO impact evaluation guidance to provide a detailed understanding of the impact of an intervention or programme. Interpreting the observed outcomes requires a knowledge of which aspects of the intervention were delivered and how well they were delivered.

This IPE guidance will help the HE sector to:

- Evaluate whether an intervention or programme was implemented as intended;
- Identify the elements of an intervention or programme that are necessary to produce the intended effects (outcomes);
- Establish whether the assumption(s) and mechanisms underpinning the intervention's ToC are correct.

Defining IPE

The term 'implementation and process evaluation' (IPE) is derived from the health-research term 'process evaluation' and adapted for the education field. IPE focuses on understanding what works for whom, in what context and why.

IPE provides researchers with a detailed knowledge of the implementation of an intervention to support the reporting of intervention or programme outcomes (Aarestrup et al., 2015). Understanding the processes of an intervention or programme gives evaluators, researchers and policymakers confidence in the outcomes reported (Bragstad et al., 2019).

IPE can help show, for example, that an intervention has not worked as intended and the reasons why this may be the case. This could include inappropriate intervention design, incomplete implementation, or insufficient accessibility of the target audience. If a problem is identified, the intervention can be adjusted for future delivery (Ishaak, de Vries & van der Wolf, 2014).

IPE can help distinguish whether an intervention has failed due to its design or its poor implementation. It can also help explain why an intervention worked and identify the active ingredients.

TASO IPE Framework

This section introduces the TASO IPE Framework, including the 11 IPE dimensions that you will use in your IPE work.

IPE is a multidimensional construct. Your IPE knowledge may focus on the term 'fidelity' – the degree to which the intervention or programme was delivered according to the agreed protocol. However, IPE is not only concerned with fidelity; it is also interested in understanding the process of delivery. It involves gathering data on intervention or programme experiences, long-term impact and changes. A focus solely on fidelity does not provide sufficient detail to understand the operation of an intervention or programme in practice, or what influences the intervention or programme outcomes, and thus limits the usefulness of the IPE. The 11 IPE dimensions to be discussed capture the fidelity of delivery, and factors that influence the implementation and the operation of the intervention or programme in practice, all of which are important for a quality IPE. We will first introduce each dimension in detail and then provide a summary of each dimension introduced in Table 1.

Control/comparison groups

Control or comparison groups are key in IPE work, and monitoring the conditions in these groups is important. For IPE work conducted within an RCT or quasi-experimental design, a proportion of participants will be assigned to the intervention (treatment) group and the remainder to the control/comparison group. It is assumed the control/comparison group will continue on a 'business as usual' basis, but it is important to determine – and collect information on – exactly what 'business as usual' looks like.

Within your IPE work, you need to monitor the control/comparison conditions, describe the nature of the group and report any treatment contamination or alternative services provided during the trial. Questionnaires can be used to collate information on what 'business as usual' looks like. This information is important for the impact evaluation, which works in tandem with the IPE. For example, if participants in the control group received an alternative intervention with similar aims, this may interfere with the perceived effects for the treatment group due to higher-than-expected gains in the control group where conditions were not truly business as usual. Conversely, if participants in the treatment group receive an unknown additional intervention, this may over-inflate the benefits of the original intervention under evaluation.

Adherence

If you are familiar with IPE, the term 'fidelity' is already in your vocabulary. Fidelity and adherence are similar dimensions. TASO has opted to use 'adherence' to broaden knowledge of IPE and move away from 'fidelity', which is grounded in health research. Adherence focuses on the extent to which the intervention or programme delivery corresponds with the intended delivery (Dane & Scheider,1998). A detailed understanding of the intervention or programme, alongside a completed ToC, is important for this dimension, as we will discuss later.

You may run programmes and interventions that are prescriptive, perhaps informed by a theoretical model or existing evidence base. In this case, to confirm adherence you would gather data to understand whether all the components of the intervention were delivered as designed. This data would indicate compliance with the intended delivery; for example, whether the facilitators were compliant and to what level, whether all components of the intervention were delivered as intended (high level of compliance) or only a number of the intended components were delivered as intended (low level of compliance).

Alternatively, your programme or intervention may be non-prescriptive or have a degree of flexibility in delivery, such as a mentoring scheme where those involved decide what to cover in their sessions. In this case, there are no set practices; therefore, you would gather data on the delivery of the intervention or programme more generally and explore what happened across the interactions and whether there were any similarities to capture. Thus, the focus is on the key values and anticipated outcomes of the programme or intervention rather than adherence to a set structure as in a prescriptive intervention.

Exposure

Again, you may be familiar with the terms 'dosage' or 'dose received'. Exposure encapsulates these terms and focuses on how much of the intervention or programme was delivered and/or received by the intended audience. Avoiding medical language makes it more suitable for the HE sector. To measure exposure, you can capture data on:

1. Number of sessions/lessons/activities implemented (e.g. if the participants were to receive 60 sessions of an intervention how many of these sessions were completed/attended?)
2. The length of these sessions/lessons/activities (e.g. if the sessions/lessons/activities were designed to last one hour, were they delivered for an hour each time?)
3. Frequency of programme/intervention delivery (e.g. if the programme or intervention was designed to be delivered three times per week for 20 weeks, how many sessions were delivered?)
4. Did the control or comparison group experience any of the interventions or programmes? (contamination)

Data on exposure also supports an understanding of levels of compliance. Before the evaluation commences, you can decide on what compliance looks like for your intervention or programme; for example, this could be the minimum number of sessions, lessons or activities that need to have been delivered to be considered as being compliant with the intended delivery of the programme or intervention.

Quality

Quality focuses on the delivery of the programme and/or intervention and whether those delivering it have the skills and techniques to do so. For a prescribed intervention informed by psychological theory, you may need someone with skills and knowledge of these techniques to deliver the intervention or programme if you are not to impact the quality of delivery. For this dimension, you will collate data on how clearly and correctly the intervention or programme was delivered. Quality is strongly linked to the training and support offered to prepare the facilitators to deliver the intervention or programme as intended. If they do not feel they have the necessary knowledge or skills, this can impact their motivation and sense of self-efficacy, which in turn impacts the quality of delivery.

Stakeholder perspective

Within the literature, several IPE dimensions focus on gathering the experiences of the participants or those involved in the programme or intervention on elements such as participant responsiveness, acceptability and dose satisfaction. These elements have been combined in this framework to form the dimension of the stakeholder perspective. This perspective considers engagement, experience, satisfaction and the perceptions of all those involved in a programme or intervention – those responsible for developing and implementing the intervention as well as the intervention beneficiaries. This provides a holistic view rather than focusing on the perspectives of a single group. It enables the consideration of a diverse range of

viewpoints, allowing for more comprehensive and nuanced findings, while also being more inclusive and sensitive to the needs and concerns of the intended recipients. Stakeholder engagement and the relevance of the programme or intervention to stakeholders may impact the outcomes reported.

Reach

Reach is concerned with the rate and scope of participation. It can be measured by the number of people who receive an intervention compared to the intended sample of recipients, or the number of practitioners delivering the intervention compared to the number trained in or expected to deliver the intervention. Quantitative assessments of reach can be completed by collecting attendance data and then reporting the percentage of the intended participant sample receiving the programme or intervention. This data can provide an understanding of the reach of an intervention or programme in the intended target pool. For example, you may have a bursary scheme to which students can apply based on their household income. You can assess the uptake of the bursary scheme by comparing those that have accessed it to those that did not, within the intended sample. If reach is low, you can then explore why students did not access this bursary scheme in order to increase future uptake.

Recruitment

Recruitment within IPE work is not focused on student recruitment but on the recruitment strategies used to enrol and retain participants or facilitators to a programme or intervention. You may have a set criterion for recruitment and a range of mitigating circumstances which should be outlined at the start of the programme or intervention to consider the effectiveness of recruitment strategies. For example, if you have set up a programme for students of black heritage to address the attainment awarding gap, you would gather data on how students enrolled on the programme and how they were retained to reduce the risk of them withdrawing from the programme. By reviewing your recruitment methods, you will understand which were most effective (e.g. emails, or posters around campus). Information on recruitment strategies should be provided in the reporting of IPE work: it may be difficult to engage the intended audience and reviewing recruitment strategies can identify which methods are effective.

Context

Arguably one of the most crucial elements within IPE is context, which typically concerns the internal and external factors influencing a programme or intervention and its outcomes. These factors are often referred to as barriers and facilitators; however, a barrier in one context may be a facilitator in another, showing the fluidity of this dimension and its importance within IPE work.

Data gathered for context focuses on the environmental aspects of an intervention's setting to understand the effect of implementing the intervention and its potential outcomes (Bejerholm et al., 2022; Grimm et al., 2021). Roles, interactions and relationships will influence the intervention and its implementation (Grant, Bugge & Wells, 2020). The feasibility of the intervention relies on whether the local setting has the necessary financial, human and implementation resources to support delivery.

For example, student mentors who are recruited, trained and managed centrally are used by various academic departments in different ways. Mentors support a wide range of activities; therefore, the work completed depends on the extent to which a department has utilised and valued this resource. For example, in an intervention designed to increase the attendance of commuter students, the context could include ease of transport to campus, the number of contact hours per day, or the type of session for which attendance is required (e.g. lecture vs group work).

Adaptation

Adaptation concerns the changes made during the delivery of an intervention, such as elements changed or added to appeal to a specific audience or removed (Outhwaite, Gulliford & Pitchford, 2020). Adapting an intervention is a natural process but it is important to know how these adaptations impact intervention outcomes. Systematically tracking and reporting on any modifications made allows an understanding to be gained of how and why these changes happened and how they relate to the outcomes reported.

Adherence and adaptation are related and both are important in understanding the outcomes reported. There will always be a level of adaptation within the delivery of an intervention or programme; it is, therefore, a question of balancing the changes made with the intended delivery of the programme or intervention. Bragstad et al. (2019) highlight concerns when adaptations result in a different intervention from the one intended. When considering adaptation, it is important to understand the allocated flexibility for the intervention or programme under evaluation. Capturing data on this dimension can help improve an intervention or programme. For example, you can learn from practitioners which changes were made, and why, to improve the programme in the future. To support this dimension, explore whether the adaptations were made due to logistics (e.g. delivering a reduced intervention due to timetable constraints), timing (e.g. to fit the local context and issues arising during delivery) or impact (e.g. whether the adaptations were positive or negative or had no impact on the goals and theory of the programme or intervention).

Capturing data on adaptation for an intervention or programme is much easier when the programme is prescribed. However, not all interventions or programmes are prescribed or have a set plan of delivery and changes may be made to the core ideas of the programme or intervention. These may be because the intervention or programme was not realistic, or not fit for purpose, or due to engagement levels. These reasons may impact generalisability and transferability, so knowing the levels of allowable adaptation is important in both prescriptive and non-prescriptive interventions and programmes.

Appropriateness

Appropriateness relates to the perceived fit and relevance of the intervention or programme to address a problem in a setting (Aldridge et al., 2016). This dimension is particularly relevant for new interventions or programmes. It has the potential to capture ‘pushback’ in implementation, if providers feel that a new programme does not align with the aims of their setting (Proctor et al., 2011). It therefore offers

additional analysis as to whether an intervention or programme is practical and achievable.

Programme differentiation

Programme differentiation is defined as the extent to which a programme's theory and practices can be distinguished from those of another programme or intervention (uniqueness) (Dane & Schneider, 1998). Data gathered under programme differentiation can help us to understand the most effective components of an intervention or programme and whether certain aspects were more effective than others and can, thus help us design future interventions or programmes (Griffin et al., 2014; see Dusenbury et al., 2003 for more information on programme differentiation). For example, this dimension could compare mentoring programmes across different HE providers to understand what they all have in common and what the differences are. Using this information, we would then understand the most effective parts of a mentoring scheme and be able to develop a new mentoring scheme combining all the components known to be effective.

Sustainability

The dimension of sustainability focuses on the extent to which an intervention or programme is sustained in a setting and becomes part of that setting's routine practice. Time needs to be allowed after first introducing a programme or intervention to explore sustainability: it is recommended that at least six months elapse before exploring the long-term impact of an intervention or programme. We understand that it is difficult to capture data for this dimension as it may be dependent on the time and funding available.

All dimensions of IPE have now been introduced in detail. Table 1 below summarises these dimensions, with examples of research questions and data collection tools.

Table 1: Summary of IPE dimensions

IPE dimension	Definition	Example research question(s)	Example data collection tools
Adherence	Adherence is defined as whether the intervention was delivered as intended. It is strongly linked to the training and support received to deliver the material, for example, formal training or support within the organisation, and buy-in.	Was the intervention or programme delivered as intended? Was the intervention or programme protocol followed when implementing the programme or intervention?	Quantitative ratings such as structured observations to look at what was delivered (adherence), and interviews to explore any difficulties in the delivery of the intervention or programme.
Exposure	Exposure is defined as the amount of intervention or programme received by participants and/or delivered to the participants.	How much of the intervention or programme was delivered? What was the total amount of time spent delivering the intervention or programme?	Attendance data can be collected on the number of sessions pupils/students attended. Logs can be completed each week to detail how many sessions were completed or not completed, and include the reasons why.
Quality	Quality considers how well the components of an intervention or programme are delivered. It is impacted by the skill and techniques needed, so is strongly linked to the training and	Did the training ensure that facilitators could implement the programme or intervention?	Training evaluation forms can be completed, and specific questions on training then included in interviews or focus groups. Alternatively, quantitative ratings such as structured observations could look at how the programme or intervention was delivered.

	support associated with intervention or programme.		
Stakeholder perspective	Stakeholder perspective captures the experiences and perceptions of those involved in the intervention or programme, including the recipients and facilitators.	How did individuals feel about the intervention or programme? What did the recipients and facilitators think about the content? How involved were the recipients in the activities and content? What impact do facilitators or participants think the programme or intervention had?	Focus groups or interviews are potentially useful methods to gather information on experiences and perceptions. Alternatively, a survey could be shared if an interview or focus group is difficult to arrange.
Reach	Reach focuses on whether the intervention or programme reached its intended audience.	Compared to the target sample for the programme or intervention, how many participants from the target audience were part of the final sample who participated in the intervention or programme itself?	Attendance data helps us understand who attended the programme or intervention and whether these participants were the target group.
Recruitment	Recruitment outlines the strategies used to recruit for an intervention or programme, in terms of both enrolment and	How were participants enrolled on to the programme or intervention? What strategies	Interviews or focus groups can explore how successful specific recruitment strategies were.

	retention.	were used to retain students on a programme or intervention?	
Context	Context captures the environmental factors outside the programme or intervention which impact delivery, including individual characteristics, organisational structure or resources available.	Did all settings have access to the same resources, for example, budget and time? Did the organisation value the programme or intervention being delivered? Did the programme or intervention align with the organisation's goals?	Field notes, interviews and focus groups may be useful here.
Adaptation	This dimension focuses on the changes made during the delivery of an intervention or programme.	Were any changes made to the programme or intervention and what were the reasons for any change?	Data can be collected during delivery in a log, for example, and then explored in detail via an interview or focus group.
Appropriateness	This element is important in the early stages of a programme or intervention to determine the perceived fit of the intervention to address an identified problem in a specific setting.	Are components of the intervention reaching the intended audience in a way that will elicit a response?	Focus groups or interviews can explore in detail whether the programme or intervention is fit for purpose.
Programme differentiation	This element focuses on the uniqueness of the programme or intervention. From this, features of the	Do all elements of the programme or intervention need to be implemented for it to be effective?	Data can be collected during delivery in a log and then explored in detail via an interview or focus group.

	programme or intervention can be identified that are essential to its success.	How does this intervention or programme differ from other, similar, schemes?	
Sustainability	Sustainability refers to the extent to which an intervention or programme is sustained in a setting after the initial delivery.	What elements of the programme or intervention have been maintained following the initial delivery? How has the intervention or programme led to changes in practice?	A survey can be completed sometime after the intervention or programme to capture the use of the material.

You now have a detailed knowledge of the foundations of IPE. Each dimension has been outlined with clear examples and research questions. To help you on your IPE journey, the guidance document will now take a step-by-step approach, following the steps outlined in the TASO Monitoring and Evaluation Framework (MEF).

Step One: Diagnose

The process of using a ToC to design an intervention based on desired outcomes is described in detail in the first step of the TASO MEF guidance. ToCs outline the intended relationships between the various aspects of an intervention, from inputs to impact, as well as the context, assumptions, mechanisms and underpinning rationale. By providing a map of how interventions are intended to work, a detailed ToC provides the evaluator with the tools to identify the cause of a negative or null result emerging from an impact evaluation. This may be due to theory failure (the intervention does not work as theorised to achieve the intended outcomes), implementation failure (the intervention was not implemented as intended), or methodology failure (the evaluation methodology was inadequate or conducted inadequately). An IPE can assess whether the implementation of the programme or intervention was aligned with the theoretical framework and/or ToC. TASO has published a range of resources to support the development of ToCs. In its guidance for producing an enhanced ToC, defining the intervention or programme is an important stage.

Define the programme or intervention

Whether the intervention is based on a core ToC or the evaluation is based on existing practice, it is important to take a step back to define your programme or intervention in detail. Imprecise descriptions of an intervention or programme will hinder the IPE work *(*Aaltio & Isokuortti, 2022). The Template for Intervention

Description and Replication (TIDieR) developed by Hoffman et al. (2014) has been adapted by TASO and is included in the guidance for developing an enhanced ToC. It is recommended that this template is completed as part of the IPE work to mitigate the risk of inadequate descriptions. For convenience, Table 2 provides an abridged version of the template adapted by TASO. This should be completed before proceeding to the production of an enhanced ToC Model.

Table 2: Adapted TIDieR information

Section name	Information to include
Name	Name or phrase describing the intervention
Why is the intervention being run?	Rationale, theory and/or goal of the elements essential to the intervention
Who is the intervention for?	Participants or beneficiaries of the intervention
What is the intervention?	Materials: Physical or informative materials used in the intervention Procedures: Procedures, activities and/or processes used in the intervention
Who is delivering the intervention?	Expertise, background and any specific training
How is the intervention delivered?	Modes of delivery of intervention (e.g. face-to-face, internet) and whether it is provided individually or in a group format
Where is the intervention delivered?	Type(s) of location(s) where the intervention occurs
How many times will the intervention be delivered? Over how long?	Number of times the intervention will be delivered and over what timeframe
Will the intervention be optimised?	If the intervention will be personalised or adapted, describe what, why, when and how
How will implementation be optimised?	Strategies to maximise effective implementation

ToC and IPE

After clearly defining your intervention or programme, the next step of the IPE process is to complete the remainder of the enhanced ToC (eToC). The impact evaluation and the IPE should flow from a single agreed eToC and complement each other, based on the overarching aims of the study. A less than full understanding of

the intervention or programme change mechanisms will hinder the creation of a robust IPE. An eToC is then needed to ensure the quality of the IPE (Aaltio & Isokuortti, 2022*)*. At the initial stage of development, setting out the eToC will ensure that the intervention is defined, potential contributions to the outcomes are noted, the pathways to impact are described and, thus, the process of change is unpacked (Camacho Garland & Beach*,* 2023*)*. In preparing the eToC, clarify the key assumptions and change mechanisms (Khayyat, Nazar & Nazar, 2021). This will help to identify which documents to review as part of the IPE at a later stage (Paquette-Warren et al., 2014). For further guidance on how to develop a ToC, visit TASO's ToC resources and templates.

Step Two: Plan

The planning of an evaluation is described in detail across multiple subheadings in the second step of the TASO MEF guidance. The same subheadings are used below. These decisions will be guided by the eToC model and the research questions to be addressed by the IPE work.

It might be thought that an IPE is only useful when an intervention or programme has been tested and is to be evaluated using an RCT design. However, while interventions are still in development, an IPE can identify effective and less effective elements, enabling the intervention to be modified at an early stage. However, a comprehensive IPE may not be suitable for a pilot project; therefore, it is recommended to focus on the dimensions that will help shape the programme or intervention for future evaluations. If there are several activities or elements within an intervention, an IPE can explore how each activity impacts the outcomes reported, thus identifying the most and least effective activities (Ishaak, de Vries & van der Wolf, 2014). Appropriateness – the perceived fit of the intervention to address a problem in a particular setting – is a suitable dimension to examine in the early stages of intervention or programme development as it informs whether the intervention or programme addresses a need within the sector. If a programme or intervention is in the early stages of development, the IPE data collected during delivery could be used to shape the delivery, but this is not advisable outside the pilot stage.

IPE runs parallel to an impact evaluation. An IPE should collect evidence at several time points during the intervention or programme delivery. In some cases, data collection will take place at one time point only; for example, focus groups will typically take place at the end of an intervention, whereas observations should, where possible, take place at multiple points to examine different phases of the implementation. Gathering data at multiple time points enables the cross-checking of information; for example, the quality of delivery may decline over time, and this would be missed if only one session were observed during the duration of the programme or intervention.

Proportionality

In preparing the IPE, consider proportionality in its design. For instance, the IPE for a long-running programme or intervention will be more comprehensive than one for a

one-off event, unless it is part of a series of activities. It is important to consider the team's time, skills and budget when planning an IPE. Practitioners, evaluators and participants should not be overburdened by an IPE. An IPE should comprehensively address the why, whom and what within any constraints a team may have. These constraints can be addressed by accessing resources already completed and making use of data already held. A balance must be achieved between rigour and the available resources, but this can be addressed with careful planning to ensure the data collected addresses the research questions set.

Identifying research questions

This guidance document has introduced you to the TASO IPE Framework. Not all these dimensions will be relevant for every piece of IPE work. Therefore, you need to decide which dimensions are important, and this will differ across the various projects you are evaluating. The questions below have been designed to support you in determining which of the IPE dimensions outlined in the guidance are most relevant to your evaluation:

1. What are the research questions that emerged from the eToC model?
2. What data do you currently have access to that could be used (e.g. monitoring data on the reach of participants to report to Office for Students)?
3. What stage is the programme or intervention at in terms of the development and evaluation lifecycle, for example, the pilot stage or a randomised controlled design?

Let us take the third question as an example. If a programme or intervention is in the pilot stage, the IPE work may focus on the following:

1. Adherence – to understand whether the intervention or programme can be delivered as currently prescribed or whether adaptations are needed;
2. Exposure – to understand how much of the intervention was delivered;
3. Stakeholder Perspective – to understand the experiences of those delivering and receiving the intervention or programme;
4. Appropriateness – to understand whether the programme or intervention addresses an identified need that led to its development.

These dimensions outline the minimum IPE requirement for a pilot IPE. If an intervention or programme is at a later stage of the evaluation cycle, perhaps at an RCT stage, the IPE work should also consider the other dimensions outlined above. Sustainability, programme differentiation and adaptation are considered as optional as their inclusion will depend on the intervention or programme under evaluation. Adaptation would be of interest if the intervention or programme to be evaluated follows a prescribed set of materials or has an underlying theoretical principle to be delivered. Adaptation is important, as the information obtained may improve engagement or ensure that changes are made to benefit a particular cohort. Programme differentiation is relevant in interventions with multiple elements to determine the active ingredients, while sustainability is important if the evaluation aims to track the long-term impact of an intervention or programme.

Identifying outcome measures

Once you have established your research questions, you will need to define the outcome measures to answer them. TASO has published a table of Common Outcome Measures. We outline example data collection tools that can be used for IPE work.

These data collection tools fall into five categories:

1. Checklists or logbooks
2. Questionnaires
3. Interviews or focus groups
4. Behavioural observations, structured or unstructured
5. Use of administrative data, such as attendance data or field notes.

The choice of data collection tool will affect the type of data collected. Quantitative data is generally numerical. Of the approaches mentioned above, questionnaires and structured observations will typically result in quantitative data. Interviews, focus groups and field notes will typically generate qualitative data.

Checklists or logbooks

Logbooks enable the recording and tracking of interventions by collecting information on the IPE dimensions of adherence, exposure, context and stakeholder perspective. Using quantitative methods, this data can be gathered longitudinally across the intervention delivery period, eliminating the risk of recall bias. A logbook gives insight into the components that were delivered, the time allocated to the intervention or programme and the reaction to the material. To support the completion of a logbook, clear instructions are needed for the respondent, and closed checklist responses are preferred to increase completion rates. An example of a logbook is provided in Figure 1.

Figure 1: Example logbook questions

1. Provide a short overview of your session (e.g. what went well, changes for next time)
2. Did the pupils engage in the session? Yes/No
3. Rate pupil engagement from 1 to 10 for engagement/understanding/ rapport
4. Is there anything else you would like to tell us?

Questionnaires

Questionnaires can be designed to include open or closed responses or both. Closed surveys typically use yes/no or tick boxes or a Likert scale. A Likert scale measures attitudes by offering possible answers on a scale from ‘strongly disagree’ to ‘strongly agree’. It is a common way of measuring responses in questionnaires. To

learn more about designing Likert scales, visit Designing Likert scales - TASO. Open responses allow participants to provide a written text response to a question.

Questionnaires are a quick and often efficient data collection tool. They can be used at the end of the intervention or programme to collect data on how participants felt about the duration and frequency of the intervention or programme components. However, questionnaires are self-reporting and may, therefore, be influenced by social desirability bias, in that respondents may be inclined to respond in a way they think the researchers want them to, such as giving a positive response.

For example, a course team puts on a series of three non-compulsory sessions designed to prepare students for writing their dissertation. Example questions to assess the accessibility of these sessions are provided below in Figure 2.

Figure 2: Example items for a questionnaire

Q1. How many sessions did you attend? [0,1,2,3]

Q2. Which, if any, of the following were barriers to you attending (select all that apply)

[Knowledge of the sessions, timing of the sessions, location of the sessions, content of the sessions, other (please specify), *none of the above*]

Q3. Which of the following best describes how you feel about the number of sessions provided? (select one)

[There were too few sessions, there were the right number of sessions, there were too many sessions]

Q4a. To what extent do you agree with the following statement:

The sessions were delivered in a way that was accessible to me.

[Strongly disagree, disagree, neither agree nor disagree, agree, strongly agree]

Q4b. Please explain your answer. [Free text box]

Interviews and focus groups

Interviews and/or focus groups are qualitative methods that can explore how an intervention or programme was implemented, identify the contextual factors influencing implementation that may have implications for the transferability of the intervention, and examine perceptions of impact. Interviews and focus groups can take place in person or online; you will need to select the most suitable approach for your participants. Both focus groups and interviews start with an opening to the topic area and are then structured with questions and prompts, ending with a summary and an opportunity for participants to add any additional comments on the topic area.

Interviews, as a data collection approach, allow multiple components of the IPE to be examined, such as adherence, exposure and stakeholder perspective. Interviews typically take place after the intervention or programme. The timing of the interviews

should be set to limit recall bias, a systematic error that occurs in remembering an event where detail is omitted or inaccurate information is given.

Focus groups are a form of interview; they collect data using group interactions to explore individual experiences. The IPE components that benefit from the use of focus groups are adherence, stakeholder perspective and adaptation. An advantage of focus groups is that they allow a more authentic discussion about an intervention or programme through a conversation between participants. However, their success depends heavily on the participants, who may be reluctant to mention barriers in front of others. Similarly, with young people, focus groups must be carefully constructed to result in rich data. Researchers need to consider factors such as the gender and school-year group of participants to limit conforming to group norms. For more information about designing and conducting interviews and focus groups, we recommend Newcomer, Hatry and Wholey (2015) as a resource.

You may want to conduct interviews with the facilitators and focus groups with the pupils at the end of a summer school. Figure 3 provides example questions you could ask the participants.

Figure 3: Example interview and focus group questions

Q1. How do you think the summer school went? (Prompts: attendance, engagement etc.)

Q2. What lessons have you learnt?

Q3. What do you remember from the summer school? (highlights/dislikes)

Q4. What did you think about the programme of activities?

Behavioural observations (structured and unstructured)

Observations provide objective information on adherence and exposure to an intervention. Observations can be structured or unstructured, in the form of field notes. Field notes are another qualitative data collection method. Structured observations are designed as a checklist of what researchers envisage should be observed in the delivery of the intervention or programme. The items on the checklist will be informed by the intervention protocol. Structured observations can be used to calculate a score for adherence. Figure 4 provides an example of an observation rating scale to assess whether the facilitators of a prescriptive intervention followed the delivery model.

Observational data is resource intensive, demanding a significant amount of effort and time. It is recommended that several observations are completed to capture fluctuations in delivery if the programme or intervention is to be delivered over a considerable length of time.

Figure 4: Example of structured observation checklist

For each session, assess the ambassadors on a scale from 0 to 6, and record the rating on the line next to the question. If you think the ambassadors fall between two descriptions, select the intervening odd number (1, 3 or 5). For example, if the ambassadors set a very good agenda but did not establish the priorities, assign a rating of 5, rather than 4 or 6.

0 Ambassadors did not set an agenda for the session.

2 Ambassadors set an agenda that was vague or incomplete.

4 Ambassadors worked with the young people to set a mutually agreed agenda that included specific targets (e.g. completing Activity 1).

6 Ambassadors worked with the young people to set an appropriate agenda with target problems, suitable for the time available. They established priorities and then followed the agenda.

If the descriptions for a given question do not seem to apply to the session you are rating, feel free to disregard them and use the more general scale below:

0	1	2	3	4	5	6
Poor	Barely Adequate	Adequate	Satisfactory	Good	Very Good	Excellent

Administrative data such as attendance data or field notes

If there are time constraints or budget concerns around an IPE, it is useful to consider what data is already collected that would help you to understand the implementation process. In this section, you would conduct a review of the information already collected. A documentation review can include reviewing a manual to understand the programme or intervention. It can also include reviewing notes made in a reflective log. If undergraduate students are being recruited to a programme, a document review of the process can be completed. An example is provided in Figure 5 below.

Figure 5: Example of a documentation review process

1. How many undergraduate students applied for the role?
2. What did the process of reducing numbers entail and was it effective?
3. At the assessment centre, collate notes from facilitators on the tasks and the assessment criteria to ensure students are suitable for the role.
4. Observations of students undertaken – collate this information to review the recruitment process.
5. Training sessions – attendance data and evaluation forms completed.

Selecting a research method

A mixed-method design should be utilised for IPE unless there is good reason not to do so. This approach collects and analyses quantitative and qualitative data within the same study, providing an in-depth and comprehensive understanding of the intervention or programme through triangulation. Triangulation highlights and compensates for the strengths and weaknesses of different methods, providing greater insight than just one method is likely to produce. It increases the credibility and validity of findings, with certain components of an IPE assessed in a variety of ways using mixed methods. For example, the IPE dimension of adherence could be measured by an observation, giving a quantitative rating of the delivery which could then be explored in more depth through an interview to understand the reasons for any challenges to adherence and the extent to which adherence is achievable. Alternatively, a logbook kept by the facilitators could be referred to during a semi-structured interview to explore experiences, barriers and facilitators to implementation. Gathering data from multiple sources can also help address missing data in cases where data collection may be problematic due to, for example, difficulties in accessing information from a school, or time constraints. What is important is to ensure that data sources are triangulated for a comprehensive IPE.

There are four approaches to mixed-methods design, as outlined below, that can be used by the sector for IPE work. These frameworks are documented below to support the planning and subsequent reporting of a mixed-methods IPE. Additional information about these mixed-methods approaches can be obtained from DeCuir-Gunby and Schutz (2016), Mertens (2017) and Morgan (2013).

Convergent parallel design

A convergent parallel design collects both qualitative and quantitative data simultaneously. The data sources are then compared and analysed to interpret the results (Creswell & Pablo-Clark, 2011; Edmonds & Kennedy, 2016). This approach allows researchers to gather different but complementary data on the same topic, facilitating discussions on where the data converges and diverges, for example, interviewing undergraduate students who are part of a leadership programme whilst also conducting a survey of the same individuals to determine their satisfaction with the programme. By triangulating the findings from different methods, the design enhances the validity and reliability of the results. It offers a more comprehensive exploration of the research topic by blending multiple research techniques. One drawback of this mixed-method design is the potential conflict between quantitative and qualitative results. While these contradictions can offer new insights, resolving the differences may be challenging, possibly requiring additional data collection (Creswell & Pablo-Clark, 2011). However, when carefully planned and executed, a convergent parallel design can yield valuable insights and a deeper understanding of a phenomenon.

Explanatory sequential design

This design combines quantitative and qualitative methods to address their respective limitations. Quantitative data may not fully explain the relationship between variables (Guest & Fleming, 2015), so qualitative data can provide a deeper understanding. The qualitative phase is informed by earlier quantitative findings. An example of this might be surveying student mentors about their satisfaction with their jobs and then conducting interviews to gain additional insight into why they responded as they did. Although this method is easy to design, implement and report, a potential weakness is the subjectivity involved in determining which quantitative findings need further explanation. Additionally, the dual data collection phases can be time-consuming (Ivankova et al., 2006). Sequential explanatory designs are commonly used in RCT studies. It is important to note the distinction between this mixed-method design and sequential exploratory design, where qualitative data is collected first, followed later by quantitative data collection (Fetters et al., 2013).

Embedded design

An embedded design involves integrating both quantitative and qualitative data collection and analysis into a traditional research design. For instance, in an RCT, researchers may wish to explore the participants' experiences in, for example, the recruitment and retention process by conducting qualitative exit interviews (McBride et al., 2019). The qualitative study, in this case, is embedded within the primary quantitative study and does not directly address the main research objectives (i.e. determining the efficacy of the treatment). The benefit of this design is that it can enhance the data and conclusions drawn from the primary study, making it a suitable option when a researcher is constrained by limited time or resources. However, this approach can also introduce complexity and an additional burden for participants in the primary study, as well as potentially requiring expertise in handling the secondary data type.

Multiphase design

In a multiphase design, quantitative and qualitative studies are used together to answer a research question. The studies are conducted sequentially or concurrently, with each study building upon the previous one. This approach is particularly effective for large and longer-term evaluation projects, as it helps to achieve a comprehensive understanding of an issue. For example, a convergent design can be used to evaluate the effectiveness of an intervention, an explanatory design can investigate the factors influencing the outcomes, and an embedded design can monitor the implementation and adaptation of the intervention. The strength of this multiphase approach includes the flexibility of mixed-methods research design, which allows the triangulation of results and obtains a deeper understanding (Sharma et al., 2023). Although this design is difficult to implement, due to the time and resources needed, the interpretation of results must be the ongoing focus of the project (Tashakkori & Creswell, 2007).

Developing an analysis strategy
Qualitative data

The most common data analysis approaches used are thematic analysis and content analysis. In this guidance document, we have therefore detailed these below; other analysis approaches can, however, be used.

Thematic analysis

Thematic analysis allows researchers to explore patterns of meaning across their data, with a researcher's subjective experience often at the core of making sense of the data. This approach could be used for data collected via focus groups and interviews. The process involved comprises six steps: familiarisation with data, generating codes, generating codes into themes, reviewing themes, defining and naming themes, and reporting findings. Familiarisation with the data includes repeated cycles of reading to create further insight, then producing codes that represent the meanings and patterns seen in the data. These codes can then be arranged into potential themes, with these themes reviewed to consider whether the data supports them and whether any themes need to be merged or removed. Each theme is then refined by developing a detailed analysis and informative label. The final step in producing the findings includes using quotes and comments from the transcribed data in the analysis.

Within the literature, datasets that were thematically analysed often used NVivo software, which assists researchers to analyse and systematically visualise qualitative data. Particularly with coding, users can create categories of data from one source or multiple cases, and use mapping tools to establish relationships between chunks of data. This additional organisation allows the researcher to question the data, draw conclusions and verify findings across units of analysis.

For more information on thematic analysis, it is recommended that researchers consult Braun and Clarke (2021); Dhakal's (2022) resource review of NVivo may also be helpful.

Content analysis

Content analysis is used to determine the presence of certain words, themes or concepts within a data set. It is based on deductive coding, using a set of predetermined codes and finding extracts that fit those codes. It can be used as a qualitative data analysis approach, or a quantitative approach by quantifying instances of coded concepts within the data. It may be used for field notes, documents and summarised notes from meetings and workshops.

Qualitative content analysis can be used when wanting to apply an interpretative level of analysis to the data; researchers may find Neuendorf's (2017) guidebook useful in achieving this, alongside Hsieh and Shannon (2005).

Quantitative data

Descriptive statistics are used to summarise the characteristics of the data collected. They give the central tendency (a single value representing the middle or centre of a data set), the mean (the sum of all values divided by the number of values in a data set), the median (the middle score of a data set when arranged in order) and the mode (the most frequent value in a data set).

Descriptive statistics can also provide information on measures of dispersion (the spread of data), such as the range (the difference between the lowest and highest values in a data set), variance (how far each value in the set is from the mean) and standard deviation (how dispersed the data is to the mean – a high standard deviation indicates that the data is more spread out). They also indicate the distribution of data through, for example, skewness – a measure of the asymmetry of a distribution, where data points cluster towards one side of a scale, creating a curve that is not symmetrical, thus enabling a more comprehensive understanding of an intervention and its outcomes. In an education programme, descriptive statistics might describe the distribution of student achievement and attendance rates. From this data, areas of improvement in the curriculum, teaching methods or resource allocation to aid student learning outcomes can be identified. Within mixed-methods analysis, the analysis of quantitative data to produce descriptive statistics will occur concurrently with the generation of initial codes of the qualitative data.

Creating a research protocol

A research protocol is a written document that describes the overall approach to be used throughout the intervention or programme. At this stage, you will want to download the IPE research protocol to complete. In the protocol clearly outline your IPE plan, including the IPE dimensions you wish to select for the evaluation.

Step Three: Measure

The third step of the TASO evaluation framework involves the active processes of collecting and analysing the data and keeping a record of your evaluation. Table 3 summarises the data collection tools for an IPE alongside the advantages and limitations of each so that you can make an informed decision on the most suitable methods for your IPE.

Table 3: Table of data collection approaches

Data collection approach	IPE Dimension (s)	Advantages	Limitations
Questionnaires	Adherence, stakeholder perspective, context, sustainability, appropriateness	Less resource intensive than face-to-face interviews or focus groups	Limited ability to fully explore participants' or implementors' views of the programme

		Easy to administer, collecting views from a wide range of respondents to allow the researcher to make robust inferences	Self-reported process evaluation data may be influenced by social desirability bias. Likert scales can be prone to central tendency bias, where individuals avoid the most extreme responses (strongly agree or disagree). May lead to a point of oversaturation where further data will not provide value-added insights, making the data collected redundant
Logbooks	Adherence, exposure, stakeholder perspective, reach, context, adaptation, programme differentiation, appropriateness	Enables recording and tracking of the delivery of the intervention or programme Data can be gathered immediately, eliminating the risk of recall bias	Difficult to calculate with certainty the quantity of intervention delivered from data reported in logbooks, due to missing data
Administration data, such as attendance data	Adherence, exposure, recruitment, reach, context	Increases understanding of the status of an intervention within setting	Can be time-consuming to access, collate and analyse
Interviews	Adherence, exposure, stakeholder engagement, context, sustainability, adaptation, programme differentiation, appropriateness	Offer detailed explanations of how contextual factors affect implementation Can generate insights into explicit and implicit beliefs about the intervention,	May be time-consuming and costly to conduct and analyse May be subject to bias in the form of impression management, demand effects, and/or differences in

		perceptions of impact, etc. Allows a focus on unanticipated issues and insights	understanding of implementation requirements Concerns over post-intervention interviews and potential recall bias, leading to loss of detail, influencing the richness of interviews in cases with a longer time interval
Focus groups	Adherence, exposure, stakeholder engagement, context, sustainability, adaptation, programme differentiation, appropriateness	Offer detailed explanations of responsiveness to all components of an intervention Can be an important source of triangulation data, with other sources Provide insights into factors affecting implementation	Focus groups that are heterogenous in gender and school years (for example) may inhibit the sharing of contrasting views and encourage the conforming to predominant norms
Observations	Adherence, exposure, stakeholder perspective, context, adaptation, programme differentiation appropriateness	Can reveal changes in implementation process over time Enable increased understanding of what the intervention looks like in practice	Resource intensive, demands significant time and effort A one-off observation does not capture fluctuations in the delivery of an intervention, but multiple observations are burdensome, may interact with the implementation and increase attrition

Step Four: Reflect

The final step of the TASO MEF describes how to report your findings and put your evidence to work. You can find the IPE reporting template here.

Implementation influences programme/intervention outcomes; therefore, your impact evaluation and IPE need to be considered holistically. To understand outcomes, you need to understand how the programme or intervention was implemented.

Implementation data can be used in two ways to understand outcomes. The first is by creating two groups which differ in their compliance in implementing the intervention or programme – for example, a compliant group with high levels of adherence and a group with low compliance. This assumes there will be variation in the delivery of an intervention or programme, but this may not always be the case. A lack of variability will make it impossible to make useful comparisons between groups. The second method is an ongoing analysis of the implementation, for example, reporting the percentage of participants reached or the level of exposure that participants received. To understand this in more detail, we recommend reading Durlak and DuPre (2008).

Assessment of the implementation is essential; evaluations that do not carefully consider implementation by collecting IPE data are incomplete. If you do not know what was delivered, you will not be able to interpret the outcome data comprehensively. Including IPE in your evaluation work will help you to understand who the programme or intervention works for, so that you can understand whether the intervention or programme was effective for females, males, students from particular groups, etc.

IPE is the gathering of data on the process of change that the intervention attempts to make; the evaluation explains how these outcomes were reached. At present, IPE inclusion is limited in the HE sector. This guidance demonstrates TASO's commitment to supporting colleagues to complete holistic, robust evaluations of programmes and interventions. It aims to help researchers undertake IPE with confidence in order to report what works, for whom, and how – all of which are important questions in evaluation.

References

Aaltio, E. and Isokuortti, N., 2022. Developing a programme theory for the Systemic Practice Model in children's social care: Key informants' perspectives. Child & Family Social Work, 27(3), pp.444-453.

Aarestrup, A.K., Jørgensen, T.S., Jørgensen, S.E., Hoelscher, D.M., Due, P. and Krølner, R., 2015. Implementation of strategies to increase adolescents' access to fruit and vegetables at school: process evaluation findings from the Boost study. BMC Public Health, 15(1), pp.1-16.

Aldridge, W.A., Boothroyd, R.I., Veazey, C.A., Powell, B.J., Murray, D.W. and Prinz, R.J., 2016. Ensuring Active Implementation Support for North Carolina Counties Scaling the Triple P System of Interventions. FPG Child Development Institute.

Allison, C., Patrick, L., Steeves, E.A., Hellwinckel, C., Zhou, W. and Colby, S., 2020. P155 Process Evaluation of a Sustainable Food Systems Course for First-Year College Students. Journal of Nutrition Education and Behavior, 52(7), pp.S89-S90.

Basch, C.E., Sliepcevich, E.M., Gold, R.S., Duncan, D.F. and Kolbe, L.J., 1985. Avoiding type III errors in health education program evaluations: a case study. Health Education Quarterly, 12(3), pp.315-331.

Bejerholm, U., Allaskog, C., Andersson, J., Nordström, L. and Roe, D., 2022. Implementation of the Recovery Guide in inpatient mental health services in Sweden—A process evaluation study. Health Expectations, 25(4), pp.1405-1417.

Bragstad, L.K., Bronken, B.A., Sveen, U., Hjelle, E.G., Kitzmüller, G., Martinsen, R., Kvigne, K.J., Mangset, M. and Kirkevold, M., 2019. Implementation fidelity in a complex intervention promoting psychosocial well-being following stroke: an explanatory sequential mixed methods study. BMC Medical Research Methodology, 19, pp.1-18.

Braun, V. and Clarke, V., 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology*, *18*(3), pp.328-352.

Camacho Garland, G. and Beach, D., 2023. Theorizing how interventions work in evaluation: Process-tracing methods and theorizing process theories of change. Evaluation, 29(4), pp.390-409.

Campbell, R., Rawlins, E., Wells, S., Kipping, R.R., Chittleborough, C.R., Peters, T.J., Lawlor, D.A. and Jago, R., 2015. Intervention fidelity in a school-based diet and physical activity intervention in the UK: Active for Life Year 5. International Journal of Behavioral Nutrition and Physical Activity, 12(1), pp.1-14.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. and Balain, S., 2007. A conceptual framework for implementation fidelity. Implementation science, 2, pp.1-9.

Clarke, V. and Braun, V., 2021. Thematic analysis: a practical guide. Thematic Analysis.

Creswell, J.W. and Clark, V.P., 2011. Mixed methods research. SAGE Publications.

DeCuir-Gunby, J.T. and Schutz, P.A., 2016. *Developing a mixed methods proposal: A practical guide for beginning researchers* (Vol. 5). Sage Publications.

Dhakal, K., 2022. NVivo. *Journal of the Medical Library Association: JMLA*, *110*(2), p.270.

Dusenbury, L., Brannigan, R., Falco, M. and Hansen, W.B., 2003. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health education research*, *18*(2), pp.237-256. https://doi.org/10.1093/her/18.2.237

Edmonds, W.A. and Kennedy, T.D., 2016. An applied guide to research designs: Quantitative, qualitative, and mixed methods. Sage Publications.

Evans, R., Brockman, R., Grey, J., Bell, S., Harding, S., Gunnell, D., Campbell, R., Murphy, S., Ford, T., Hollingworth, W. and Tilling, K., 2018. A cluster randomised controlled trial of the Wellbeing in Secondary Education (WISE) Project–an intervention to improve the mental health support and training available to secondary school teachers: protocol for an integrated process evaluation. Trials, 19(1), pp.1-13.

Fetters, M.D., Curry, L.A. and Creswell, J.W., 2013. Achieving integration in mixed methods designs—principles and practices. Health services research, 48(6pt2), pp.2134-2156.

Gaglio, B., Shoup, J.A. and Glasgow, R.E., 2013. The RE-AIM framework: a systematic review of use over time. American Journal of Public Health, 103(6), pp.e38-e46.

Grant, A., Bugge, C. and Wells, M., 2020. Designing process evaluations using case study to explore the context of complex interventions evaluated in trials. Trials, 21(1), pp.1-10.

Griffin, T. L., Pallan, M. J., Clarke, J. L., Lancashire, E. R., Lyon, A., Parry, J. M., Adab, P., & WAVES study trial investigators (2014). Process evaluation design in a cluster randomised controlled childhood obesity prevention trial: the WAVES study. The international journal of behavioral nutrition and physical activity, 11, 112. https://doi.org/10.1186/s12966-014-0112-1

Grimm, S.E., Pouwels, X., Ramaekers, B.L., van Ravesteyn, N.T., Sankatsing, V.D., Grutters, J. and Joore, M.A., 2021. Implementation barriers to value of information analysis in health technology decision making: results from a process evaluation. Value in health, 24(8), pp.1126-1136.

Guest, G. and Fleming, P., 2015. Mixed methods research. Public Health Research Methods, pp.581-610.

Hoffmann, T.C., Glasziou, P.P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.G., Barbour, V., Macdonald, H., Johnston, M. and Lamb, S.E., 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Bmj, 348.

Hsieh, H.F. and Shannon, S.E., 2005. Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), pp.1277-1288.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K., 2016. Implementation and process evaluation (IPE) for interventions in educational settings: A synthesis of the literature. London: EEF.

Ishaak, F., de Vries, N.K. and van der Wolf, K, 2014. 'Test implementation of a school-oriented drug prevention program "Study without Drugs": pre- and post-testing for effectiveness.', BMC Public Health, 14(1), pp. 503–529.

Ivankova, N.V., Creswell, J.W. and Stick, S.L., 2006. Using mixed-methods sequential explanatory design: From theory to practice. Field methods, 18(1), pp.3-20.

Khayyat, S.M., Nazar, Z. and Nazar, H., 2021. A study to investigate the implementation process and fidelity of a hospital to community pharmacy transfer of care intervention. PLoS One, 16(12), p.e0260951.

McBride, K.A., MacMillan, F., George, E.S. and Steiner, G.Z., 2019. The use of mixed methods in research.

Mertens, D.M., 2017. *Mixed methods design in evaluation* (Vol. 1). SAGE publications.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D. and Baird, J., 2015. Process evaluation of complex interventions: Medical Research Council guidance. BMJ, 350.

Morgan, D.L., 2013. *Integrating qualitative and quantitative methods: A pragmatic approach*. Sage publications.

Morgan-Trimmer, S., 2015. Improving process evaluations of health behavior interventions: learning from the social sciences. Evaluation & the Health Professions, 38(3), pp.295-314.

Morrison, J.Q., Newman, D.S. and Erickson, A.G., 2021. Process evaluation of literacy practices within a multi-tiered system of supports framework. Journal of Applied School Psychology, 37(2), pp.140-164.

Mühlfelder, M., Konermann, T. and Borchard, L.M., 2015. Design, implementation, and evaluation of a tutor training for problem-based learning in undergraduate psychology courses. Journal of Problem-based Learning in Higher Education, 3(2).

National Assessment Governing Board, 2010. 'Design Document for 12th Grade NAEP Preparedness Research Judgmental Standard Setting Studies: Setting Standards on the National Assessment of Educational Progress in Reading and Mathematics for 12th Grade Preparedness'.

Newcomer, K.E., Hatry, H.P. and Wholey, J.S. eds., 2015. *Handbook of practical program evaluation* (pp. 1-864). San Francisco, CA: Jossey-Bass & Pfeiffer Imprints, Wiley.

Neuendorf, K.A., 2017. The content analysis guidebook. sage.

Outhwaite, L.A., Gulliford, A. and Pitchford, N.J., 2020. A new methodological approach for evaluating the impact of educational intervention implementation on learning outcomes. International Journal of Research & Method in Education, 43(3), pp.225-242.

Paquette-Warren, J., Roberts, S.E., Fournie, M., Tyler, M., Brown, J. and Harris, S., 2014. Improving chronic care through continuing education of interprofessional primary healthcare teams: a process evaluation. Journal of Interprofessional Care, 28(3), pp.232-238.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R. and Hensley, M., 2011. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Administration and policy in mental health and mental health services research, 38, pp.65-76.

Sharma, L.R., Bidari, S., Bidari, D., Neupane, S. and Sapkota, R., 2023. Exploring the Mixed Methods Research Design: Types, Purposes, Strengths, Challenges, and Criticisms. Glob Acad J Linguist Lit, 5.

Shelton, R.C., Cooper, B.R. and Stirman, S.W., 2018. The sustainability of evidence-based interventions and practices in public health and health care. Annual review of public health, 39, pp.55-76.

Tashakkori, A. and Creswell, J.W., 2007. The new era of mixed methods. Journal of mixed methods research, 1(1), pp.3-7.

Walton, T.R. and Carrillo-Higueras, F., 2019. Evaluating the effectiveness of university widening participation activities in rural Australia. Assessment & Evaluation in Higher Education, 44(5), pp.799-819.

Wilson, D.K., Griffin, S., Saunders, R.P., Kitzman-Ulrich, H., Meyers, D.C. and Mansard, L., 2009. Using process evaluation for program improvement in dose, fidelity and reach: the ACT trial experience. International Journal of Behavioral Nutrition and Physical Activity, 6, pp.1-10.