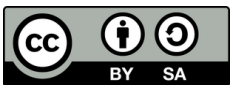
unesco

EVALUATION RESEARCH FRAMEWORK

Dr. Miloš Kankaraš

Published in 2026 by the United Nations Educational Scientific and Cultural Organization | Mahatma Gandhi Institute of Education for Peace and Sustainable Development, 35 Ferozeshah Road, New Delhi 110001, India

http://doi.org/10.56383/MNQB9884



Attribution
This publication can be cited as follows:
Kankaraš, M. (2026). Evaluation Research Framework of UNESCO MGIEP. New Delhi: UNESCO MGIEP.

Cover and design: Pallavi Shahi, UNESCO MGIEP
Illustrations: Adobe Stock and Noun Project

MGIEP/EVALUATIONRESEARCHFRAMEWORK/2026/1 Rev.

Contents

Acknowledgements

I am grateful to my colleagues at UNESCO MGIEP for their valuable feedback on the early drafts of this framework. Their dedication and critical contributions are vital for the overall success of the Institute's Programme Design and Evaluation mission. Special thanks to the Communication Team for their support in preparing this report for publication, ensuring its clarity and accessibility.

List of Abbreviations

- **AI**: Artificial Intelligence
- **ANOVA**: Analysis of Variance
- **DAC**: Development Assistance Committee
- **ICT**: Information and Communication Technology
- **IRT**: Item Response Theory
- **KPIs**: Key Performance Indicators
- **MGIEP**: Mahatma Gandhi Institute of Education for Peace and Sustainable Development
- **OECD**: Organisation for Economic Co-operation and Development
- **PD&E**: Programme Design and Evaluation
- **PISA**: Programme for International Student Assessment
- **PVE**: Prevention of Violent Extremism
- **R&D**: Research and Development
- **RCT**: Randomized Controlled Trials
- **SECs**: Social and Emotional Competencies
- **SDGs**: Sustainable Development Goals
- **SEL**: Social and Emotional Learning
- **SES**: Socio-economic status
- **UNESCO**: United Nations Educational, Scientific and Cultural Organization

Executive Summary

The United Nations Educational, Scientific and Cultural Organization's Mahatma Gandhi Institute of Education for Peace and Sustainable Development (UNESCO MGIEP) is committed to promoting peace and sustainable development through education. As part of this mission, UNESCO MGIEP is implementing a range of educational programmes in related areas, including online courses, in-person workshops, and hybrid programmes that combine elements of both online and in-person educational interventions.

Recognising that assessing the effectiveness of these programmes is important to successfully fulfil its mission, the Institute has committed to comprehensively evaluating its educational programmes. This paper presents the newly developed Evaluation Research Framework describing the procedures and guidelines which will be followed for all research aspects of its evaluation studies. The Framework also discusses key topics in evaluation research and outlines UNESCO MGIEP's approaches and standards in these areas. It has been designed to ensure scientific rigour and consistency of all applied procedures across UNESCO MGIEP's studies. This will help the Institute gather valid and reliable evaluation data on the effectiveness of its educational initiatives.

Purpose and Objectives

The Evaluation Research Framework serves as a strategic tool for systematically evaluating the design, implementation and impact of UNESCO MGIEP's educational programmes. It is designed to:

- Provide a structured approach to evaluating the effectiveness of UNESCO MGIEP's educational initiatives
- Ensure that all programmes include a built-in evaluation component
- Facilitate continuous improvement of programmes through ongoing monitoring and evaluation

Key Components of the Framework

1. **Evaluation objectives**: Clearly defined objectives need to be established for each evaluation study. These objectives should be based on the programme's Theory of Change and are designed to assess specific outcomes, such as improvements in knowledge about social and emotional competencies or the development of global citizenship skills (Rossi, Lipsey, & Freeman, 2004).

2. **Evaluation questions**: Based on the objectives, the framework outlines specific evaluation questions that the study seeks to answer. These questions should guide the choice of evaluation design, data collection methods, and analysis techniques (Bryman, 2016). For example, an evaluation question might ask, "To what extent does participation in UNESCO MGIEP's Media Literacy course improve learners' ability to identify and resist media disinformation campaigns?"

3. **Data collection and analysis**: The framework specifies the data collection methods to be used for each type of study, which may include surveys, knowledge and skill tests, qualitative questions, focus groups and/or interviews, and observations. The analytical techniques that will be used are also specified. t-Tests, regression analyses, and other statistical models could be used for analysis depending on the design and research questions (Creswell & Creswell, 2017).

4. **Reporting and utilization of findings**: Finally, the framework emphasizes the importance of reporting the results of evaluation studies in a clear and accessible manner. The findings are used not only to improve UNESCO MGIEP's programmes but also to inform stakeholders, including educators, policymakers, and the broader educational community, about the effectiveness of the Institute's education initiatives (Patton, 2002).

Strategic Goals and Outcomes

The implementation of the Evaluation Research Framework is aligned with UNESCO MGIEP's broader strategic goals of promoting innovative educational practices and scaling up successful programmes. The framework was developed in 2024 with the vision that eventually all UNESCO MGIEP programmes would have embedded evaluation components that enable real-time assessment and improvement. This would enable the Institute to:

- Demonstrate the impact of its educational initiatives on learners and communities
- Provide evidence-based recommendations for policy and programme development
- Ensure that its programmes are adaptable, scalable, and relevant to diverse cultural and educational contexts

Implication for Policymakers and Stakeholders

For policymakers and other stakeholders, the Evaluation Research Framework represents an important step in advancing UNESCO MGIEP's mission. The Framework not only supports the continuous improvement of educational programmes but also provides a model for evidence-based decision-making in educational policy and practice. By investing in rigorous and systematic evaluation, the Institute aims to position itself as a leader in evidence-based policy research, contributing to achieving the United Nations Sustainable Development Goals (SDGs), particularly SDG 4.7, which focuses on peace and sustainable development.

Chapter 1:
Introduction

1.1 About UNESCO MGIEP

The Mahatma Gandhi Institute of Education for Peace and Sustainable Development (MGIEP) is a specialised research and education institute under UNESCO. Established in New Delhi, India, in 2012, with the vision of transforming education for humanity, UNESCO MGIEP's mission is to promote peace and sustainable development through education by providing research insights that aid policymaking and foster innovative educational approaches.

1.2 UNESCO MGIEP's Program Design and Evaluation Framework

UNESCO MGIEP has established a new Programme Design and Evaluation (PDE) workstream to streamline operations and improve the efficacy and impact of its programmes. Design and evaluation are complementary components of the workstream and together constitute a holistic programme management structure, in line with UNESCO's Results-Based Management (RBM) framework.

As depicted in Figure 1, programme design and evaluation are integral parts of a broader programme design, evaluation and monitoring framework. They serve different but complementary functions and, when used in synchrony, can help create a virtuous circle of continuous improvement in programme management and enhance overall programme impact.

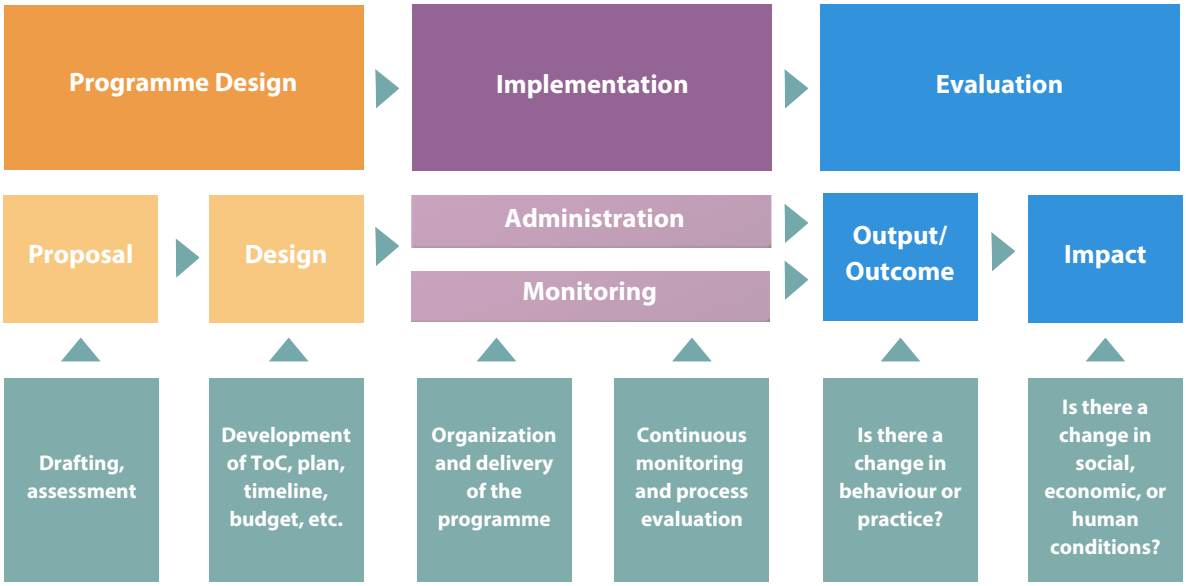


Figure 1: Evaluation and Monitoring Framework
(Developed in-house by the author)

1.2.1 Programme Design

The new PDE team at UNESCO MGIEP will be responsible for organising the development of new programmes, as well as the redesigning existing programmes, ensuring consistency and quality in the following procedures:

- Development of relevant frameworks, guidelines, templates, evaluation criteria, scoring sheets and other necessary documents
- Provision of required support to project leads in all stages of programme design
- Provision of timely and relevant feedback to project leads on all relevant aspects of the programme

The PDE team will aim to improve the overall quality and effectiveness of the newly designed and redesigned programmes and, thus, their impact. For this, it will rely on the approaches and procedures established in UNESCO's RBM system (UNESCO, 2013, 2024) and incorporate findings from evaluation studies of previous project iterations and the lessons learned from them.

1.2.2 Programme Evaluation

Evaluation research is an integral aspect of UNESCO MGIEP's broader approach to programme management. It helps ensure that the Institute's educational programmes achieve their intended objectives and contribute meaningfully to its mission. UNESCO MGIEP uses Monitoring, Evaluation, and Learning (MEL)–oriented evaluation research to inform decision-making, which has multiple advantages:

- Promotes accountability and informed decision-making
- Encourages institutional learning and continuous improvement
- Contributes to the field of education

1.3 Importance of evaluation research at UNESCO MGIEP

Evaluation research is a key integral aspect of UNESCO MGIEP's approach to its portfolio of educational programmes. As the Institute develops and implements various educational programmes – including online courses, in-person workshops, online workshops or some combination of these elements – it is critical to systematically measure the effectiveness of these interventions. Evaluation research involves the use of rigorous research methods to examine and assess the design, implementation, and results (activities, outputs, outcomes and impacts) of the programmes. This approach enables the Institute to ensure that its programmes are achieving their intended objectives, such as enhancing participants' knowledge about social and emotional competencies, fostering their global citizenship competencies or promoting sustainable development practices.

The evaluation research conducted by UNESCO MGIEP has several distinct but related functions and objectives. Firstly, it supports accountability to various stakeholders and creates conditions for informed decision-making by providing transparent and evidence-based assessments of programme results. Secondly, it emphasizes the use of structured learning loops to inform decision-making. In particular, evaluation research provides critical feedback that the Institute's programme developers can use to continuously improve the quality and effectiveness of their educational programmes.

This feedback loop is essential for refining programme content, delivery methods, and pedagogical approaches, ensuring they remain relevant and effective in achieving the desired outcomes. Finally, evaluation research contributes to the broader field of educational research by generating insights that can be applied in other contexts, supporting global education initiatives in peace and sustainability.

1.4 Purpose of the Evaluation Research Framework

The Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) has established widely accepted evaluation criteria – namely, relevance, coherence, effectiveness, efficiency, impact, and sustainability (OECD, 2019). The Institute's Evaluation Research Framework aims to provide a comprehensive, consistent, scientifically sound and well-structured approach for designing evaluation research studies that measure effectiveness and efficiency of the Institute's educational programmes and generate data for measurement of impact and sustainability. This framework is designed to guide the development of research designs and implementation of empirical studies that evaluate whether UNESCO MGIEP's programmes are achieving their intended learning outcomes. Specifically, the framework addresses key aspects such as selecting a suitable research design, formulating relevant research questions and objectives, choosing appropriate measurement scales, identifying which aspects of educational programmes to assess, and how to analyse, interpret and apply the results.

The Framework integrates process and results evaluations to offer a comprehensive assessment of the educational programmes. Process evaluations focus on programme implementation, examining whether they are delivered as intended and identifying areas for improvement. Results evaluations, in contrast, measure programme outcomes and impact. **Outcome evaluations assess the extent to which changes** in participants' **knowledge**, **skills**, and **attitudes** can be **attributed** to the programme (i.e. outcomes), whereas impact evaluations focus on longer-term, higher-level effects. By combining these approaches and potentially expanding the framework to later include the "coherence" criteria, the Evaluation Research Framework aims to provide a holistic view of the effectiveness, relevance, impact and sustainability of UNESCO MGIEP's educational interventions.

1.5 The need for a robust evaluation framework

As UNESCO MGIEP continues to expand its educational programmes and reach a broader audience, the need for a robust and adaptable evaluation framework has become increasingly important. A well-structured framework will not only enhance the credibility and results of the Institute's programmes but also contribute to the global discourse on effective educational practices. By systematically evaluating its programmes, UNESCO MGIEP can ensure it remains at the forefront of evidence-based educational policy research and practice, leading the way in developing programmes that promote peace, sustainability, and global citizenship.

The Evaluation Research Framework described in this document serves as both an overview of UNESCO MGIEP's evaluation approach and a practical guide for designing and conducting empirical studies. It reflects UNESCO MGIEP's commitment to using robust scientific approaches for improving educational outcomes and set high standards of educational quality.

Chapter 2:
Evaluation Research in Education

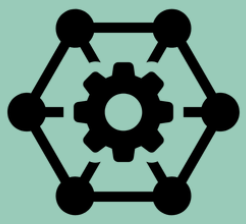
Evaluation research is a critical component of educational policy and research because it provides empirical evidence on the effectiveness, relevance, and sustainability of educational programmes, policies, and interventions. As education systems worldwide face increasing demands for accountability and evidence-based practices, evaluation research has become essential in determining whether educational initiatives achieve their intended outcomes (Fitzpatrick, Sanders, & Worthen, 2011). In addition, such studies provide the foundation for continually improving educational practices, ensuring that programmes are not only effective but also relevant to the needs of diverse learners.

The primary goal of evaluation research is to assess the effects of educational programmes on participants and to understand how they support broader educational goals. This involves a systematic examination of various aspects of the programme, including its design, implementation and outcomes (Rossi, Lipsey & Freeman, 2004). Through rigorous evaluation, educators and policymakers can identify areas of strength and those in need of improvement, thereby enhancing the overall quality of education.

2.1 Theoretical foundations of evaluation research

The theoretical foundations of evaluation research in education are based on several key concepts, including programme theory, logic models and validity.

Programme theory provides a conceptual framework for understanding how an educational programme is expected to work, linking its activities to its intended outcomes (Stufflebeam & Shinkfield, 2007). This framework serves as a guide for evaluators, helping them identify the key components of a programme that should be examined during the evaluation process.



Logic models, sometimes also called Results Frameworks or Theories of Change, help link inputs, activities, outputs, outcomes, and impact (UNESCO, 2013, 2024) . They visually illustrate the logical relationships between a programme's resources, activities, outputs, and outcomes, and the achievement of the intended programme goals (Bryman, 2016). Logic models are particularly useful in complex educational settings, where multiple factors may influence the programme's success. As they provide a visual representation of the relationships, evaluators can easily focus on the programme's most critical aspects and design evaluations that effectively measure its impact.

Validity concerns the accuracy and trustworthiness of evaluation findings. In evaluation research, it is typically classified as internal, external, and construct validity (Cohen, Manion & Morrison, 2018). Internal validity refers to the extent to which the evaluation correctly identifies causal relationships within the programme. External validity, in contrast, concerns the generalisability of the findings to other contexts, ensuring that the results are applicable beyond the study's specific setting. Construct validity concerns how accurately the evaluation measures the intended constructs, and whether they reflect the outcomes the programme aims to achieve (Shadish, Cook & Campbell, 2002).

These theoretical foundations are essential for guiding the development of evaluation questions, choosing appropriate methodologies, and interpreting the results of the evaluation. By adhering to these principles, evaluators can ensure that their research is rigorous, valid, and relevant to the broader field of education.

2.2 Methodological approaches in evaluation research

Evaluation research employs diverse methodological approaches, depending on the evaluation questions and research contexts. They are generally classified as quantitative, qualitative, and mixed methods approaches, each offering unique insights into the effectiveness of educational programmes.

2.2.1 Quantitative approaches

Quantitative evaluation methods involve collecting and analysing numerical data to assess programme processes and outcomes. They are particularly effective for measuring changes in the knowledge, skills, or attitudes of programme participants (Creswell & Creswell, 2017). The most used quantitative designs in evaluation research in education are pre- and post-tests, surveys and questionnaires, which assess participants before and after an intervention to measure changes resulting from the programme. For example, a pre- and post-test design may be used to evaluate the impact of a Social and Emotional Learning (SEL) programme by measuring students' knowledge and/or topic awareness of these domains before and after the programme.

Another commonly used quantitative approach is the experimental design, which involves randomly assigning participants to treatment and control groups. This approach allows evaluators to establish causality by comparing the outcomes of those who participated in the programme with the outcomes of those who did not (Shadish, Cook & Campbell, 2002). Quasi-experimental designs, which do not involve random assignment but still include comparison groups, are also widely used in educational evaluations where randomization is not feasible (Trochim & Donnelly, 2006). These designs help to control for potential confounding variables, thereby increasing the internal validity of the evaluation study.

2.2.2 Qualitative approaches

Qualitative methods can provide in-depth insights into educational programmes and their implementation processes and contexts. They include interviews, focus groups, observations and outcome harvesting, which allow evaluators to gather rich, detailed data on participants' experiences and perceptions as well as their behaviour (Bryman, 2016). For instance, interviews with teachers and students can shed light on how an educational programme is being implemented in the classroom and how participants perceive it (Patton, 2002).

Qualitative evaluation is particularly valuable for exploring the reasons behind a programme's success or failure. It can reveal the quality of instruction, participant engagement and the cultural relevance of the content, which are some of the factors that contribute to programme outcomes (Maxwell, 2013). By understanding these factors, evaluators can provide more detailed recommendations to improve the programme and adapt it to different contexts.

2.2.3 Mixed Methods approach

Mixed Methods evaluation combines quantitative and qualitative approaches to provide a more comprehensive understanding of programmes and their effectiveness. This approach is especially useful in complex educational settings, where multiple factors may influence outcomes (Creswell & Creswell, 2017). Mixed Methods evaluations can offer a more complete picture of how and why a programme works by leveraging the strengths of both quantitative and qualitative methods while avoiding their shortcomings.

For example, a mixed methods evaluation might involve administering pre-post tests to measure changes in participants' knowledge and skills, while also conducting focus group discussions to explore participants' experiences and perceptions of the programme (Greene, 2007). This approach enables evaluators to triangulate their findings, enhancing both the reliability and validity – and thus the overall usefulness – of the evaluation results (Johnson & Onwuegbuzie, 2004).

2.3 Challenges in evaluation research

Despite its importance, evaluation research in education faces several significant challenges. One of the most prominent challenges concerns validity, particularly internal and external validity. Internal validity can be compromised by factors such as selection bias, uncontrolled confounding variables and various types of measurement errors. This can lead to incorrect conclusions about the effectiveness of a programme (Shadish, Cook & Campbell, 2002). For instance, if a study does not adequately control for differences between the treatment and control groups, it may falsely attribute changes in outcomes to the programme when they could actually be due to pre-existing differences between the groups. Studies employing only treatment groups (pre-post designs) are more vulnerable to making type I errors (false positive), incorrectly concluding that a particular programme has a positive effect while the observed positive changes are due to other confounding factors that are not related or only indirectly related with the educational intervention.

External validity, or the generalizability of evaluation findings, is another concern. If the sample used in an evaluation is not representative of the wider population, or if the evaluation context differs significantly from real-world settings, the findings may not be applicable beyond the specific study (Cohen, Manion & Morrison, 2018). This challenge is particularly relevant in educational research, where programmes are often implemented across diverse settings and involve participants from varied backgrounds.

This means that the results of an evaluation study conducted, for example, in a school in New Delhi, will not necessarily be applicable to other contexts in India or, even less so, internationally. It also means that the evaluation results of a programme implemented in one format (e.g., a two-day in-person workshop combined with an online course) may not apply to other intervention formats (e.g. an asynchronous online version of the same programme). Factors such as delivery schedule, model of administration, duration, the involvement of lecturers/facilitators, and mode of delivery can impact programme outcomes even when the same materials are used.

Developing reliable, valid and culturally appropriate measurement scales is a challenging process that often requires a high level of expertise. The scales also need to be tested in pilot studies before their implementation in an evaluation study.
As participants in educational programmes come from diverse backgrounds, developing scales that are suitable for use across socio-economic and cultural contexts can be challenging.

Another significant challenge in evaluation research is the resource-intensive nature of rigorous studies. High-quality evaluation studies (especially those assessing the programme's impacts) often require substantial time, expertise, and financial resources, which may not always be available. This can lead to compromises in the evaluation design, such as the use of smaller sample sizes or less rigorous methodologies, which can weaken the findings (Fitzpatrick, Sanders & Worthen, 2011).

Ethical considerations can also present challenges in evaluation research. Evaluators must ensure that participants' rights to privacy, confidentiality, and informed consent are protected (Yarbrough et al., 2010). In educational settings, this often involves balancing the need for rigorous evaluation with the ethical imperative to do no harm. For example, in experimental designs, withholding a potentially beneficial intervention from a control group can raise ethical concerns; such situations must be managed ethically (Sieber, 2009). Likewise, providing an intervention to an experimental group to prove a hypothesis can sometimes lead to unintended negative consequences (Evans, Scourfield & Murphy, 2015).

Chapter 3:
UNESCO MGIEP's Evaluation Research Framework

3.1 Purpose and characteristics of UNESCO MGIEP's evaluation studies

UNESCO MGIEP has developed a comprehensive Evaluation Research Framework to assess the effectiveness, relevance and sustainability of its educational programmes. The framework aims to provide a structured and systematic approach to evaluating UNESCO MGIEP's educational programmes, including its online courses and in-person workshops. The framework is essential for determining whether these programmes achieve their intended outcomes, such as enhancing SEL, promoting global citizenship, and fostering sustainable development.

UNESCO MGIEP's evaluation studies include both process and results evaluations. This dual approach provides a comprehensive understanding of a programme's effectiveness, appropriateness and impact, as well as the quality of its implementation, the relevance of the used modalities and its efficiency and timeliness. A key feature of UNESCO MGIEP's evaluation framework is its alignment with the Institute's broader mission of promoting peace and sustainable development through education. The evaluation studies are designed not only to assess the effectiveness of individual programmes but also to contribute to achieving the Institute's overarching goals. By consistently evaluating its programmes, UNESCO MGIEP can ensure that its initiatives are both impactful and aligned with its mission (Rossi, Lipsey & Freeman, 2004).

3.2 Core components of UNESCO MGIEP's evaluation studies

Till 2024, UNESCO MGIEP's evaluation studies used a single-group pre-post design as a core research component. This design involves measuring participants' competencies before and after they complete a programme, enabling the assessment of immediate effects. For example, in a biodiversity course, the participants might complete a pre-course assessment measuring their knowledge of biodiversity topics, followed by a post-course assessment. The difference in scores provides an indication of the programme's impact on participants' understanding of the subject.

Although the single-group pre-post design is relatively straightforward and easy to implement, it has several drawbacks. One of the primary limitations is the lack of a control or comparison group, which makes it challenging to attribute changes in outcomes directly to the programme. Without a control group, it becomes difficult to rule out other factors – such as participants' prior experiences or external influences – that might have caused the observed changes (Shadish, Cook & Campbell, 2002).

Despite these limitations, the single-group pre- and post-design provides valuable insights into the effectiveness of UNESCO MGIEP's programmes. For instance, evaluations of UNESCO MGIEP's online courses showed statistically significant improvements in participants' self-efficacy regarding the programme topics, suggesting that these programmes are achieving some of their goals. In the future, more advanced designs will be explored to not only provide stronger evidence of programme impact but also to enhance the rigour and reliability of the Institute's evaluations.

3.3 Extended design components of UNESCO MGIEP's evaluation studies

To address some of the limitations of the single-group pre- and post- design, UNESCO MGIEP is attempting to incorporate more complex and rigorous evaluation methods where possible. These additional design elements include the use of longer-term assessments, quasi-experimental and experimental designs that involve the use of control or comparison groups, observational studies to examine the behavioural impact of the programmes, focus groups, and other approaches that provide qualitative insights into the programme. By including these elements, evaluators can better isolate the effects of the programme and provide more robust evidence of its impact (Shadish, Cook & Campbell, 2002).

3.3.1 Assessment of longer-term effects

The standard pre- and post-test design assesses the short-term or immediate effects of an educational programme. However, all UNESCO MGIEP educational programmes aim for long-term effects or outcomes, that is, for changes in learners' knowledge, skills, and attitudes that will not only last but also influence other aspects of their skills, behaviours and, ultimately, their life choices (UNESCO, 2013, 2024). Consequently, UNESCO MGIEP evaluation studies aim to adopt research designs that capture such longer-term outcomes where possible. Currently, such designs are being implemented across all UNESCO MGIEP hybrid programmes, where participants are contacted six months after the programme completion and asked to complete a survey comprising a set of evaluation scales. These designs are also being used in several pilot studies for the evaluation of newly developing or revised online courses.

3.3.2 Quasi-experimental designs

Quasi-experimental designs are another evaluation strategy used by UNESCO MGIEP. These designs involve comparing outcomes between participants who have undergone the programme (intervention group) and those who have not (comparison group), but there is no random assignment. Although quasi-experimental designs do not offer the same level of control as true experiments,

they are often more feasible in educational settings where randomization may not be possible (Trochim & Donnelly, 2006). For example, in a quasi-experimental evaluation of UNESCO MGIEP's programme, SEL for Decision Makers, implemented in Bhutan, participants were divided into treatment and comparison groups. The UNESCO MGIEP team compared the results of the effectiveness analyses across the two groups, thus improving the internal validity of the research insights obtained.

3.3.3 Experimental designs

Experimental designs , particularly Randomized Controlled Trials (RCTs), represent the gold standard in evaluation research. In these kinds of studies, participants are randomly assigned to either the treatment/intervention group or the control group. This yields the highest level of internal validity as it is possible to control for confounding variables (Cohen, Manion, & Morrison, 2018). UNESCO MGIEP will implement experimental designs wherever feasible, especially in new or pilot programmes where rigorous testing is crucial to validly determine the programme's effectiveness. These designs can also be used to evaluate other aspects of educational programmes. For example, an experimental design can be used to compare how the order of administration of two educational components (e.g. an in-person workshop and an online course) affects programme effectiveness.

Incorporating these more rigorous designs into the Institute's evaluation framework will enhance its ability to draw causal inferences about the effectiveness of its programmes and identify the most effective interventional designs. It will also enable the Institute to refine its programmes using more reliable evidence, ensuring the programmes continue to meet learners' needs and align with the Institute's mission (Fitzpatrick, Sanders, & Worthen, 2011).

3.3.4 Observational studies

Observational designs are another important research component of evaluation studies that are and will continue to be used in UNESCO MGIEP's evaluation studies. While knowledge, skills and attitudes can be assessed by other methods, observational studies allow researchers to assess behavioural outcomes, evaluating the

effects of educational programmes on learners' behaviours in relevant real-life situations. These studies are especially important in UNESCO MGIEP projects involving teachers, because their students are likely to benefit from any positive changes. Therefore, it is not enough to only assess teachers' attitudes and knowledge. The degree to which any observed change in these translates into changes in their teaching practices must also be observed. Therefore, in its recently implemented Sri Lanka teacher training programme, the Institute conducted an observational study of participating lecturers that also incorporated a longitudinal research design, making the evaluation of this hybrid programme the most extensive evaluation study conducted by UNESCO MGIEP to date. The study involved observing the participating lecturers directly in the classroom before and after the administration of the educational intervention, which consisted of an in-person workshop and two online courses.

3.3.5 Qualitative approaches

Qualitative feedback can be obtained through open-ended survey questions, structured or semi-structured interviews, and focus groups. The latter is an especially efficient way of obtaining critical information on participants' views on different aspects of the programme and its overall quality and relevance. It is also well-suited for in-depth examinations of personal experiences and reflections which can often shed light on the key reasons driving the results obtained through quantitative approaches. Therefore, UNESCO MGIEP incorporates at least some qualitative elements into its evaluation designs so researchers can better make sense of quantitative data and arrive at valid conclusions about the meaning and significance of the empirical findings.

3.4 Key components of the evaluation framework

Several important components in the UNESCO MGIEP Evaluation Research Framework guide the design and implementation of evaluation studies. These components include:

1. **Evaluation objectives:** Clearly defined objectives need to be established for each evaluation study. These objectives should be based on the programme's Theory of Change and are designed to assess specific outcomes, such as improvements in knowledge about social and emotional competencies or the development of global citizenship skills (Rossi, Lipsey, & Freeman, 2004).

2. **Evaluation questions:** Based on the objectives, the framework outlines specific evaluation questions that the study seeks to answer. These questions should guide the choice of evaluation design, data collection methods, and analysis techniques (Bryman, 2016). For example, an evaluation question might ask, "To what extent does participation in UNESCO MGIEP's Media Literacy course improve learners' ability to identify and resist media disinformation campaigns?"

3. **Data collection and analysis:** The framework specifies the data collection methods to be used for each type of study, which may include surveys, knowledge and skill tests, qualitative questions, focus groups and/or interviews, and observations. The analytical techniques that will be used are also specified. t-Tests, regression analyses, and other statistical models could be used for analysis depending on the design and research questions (Creswell & Creswell, 2017).

4. **Reporting and utilization of findings:** Finally, the framework emphasizes the importance of reporting the results of evaluation studies in a clear and accessible manner. The findings are used not only to improve UNESCO MGIEP's programmes but also to inform stakeholders, including educators, policymakers, and the broader educational community, about the effectiveness of the Institute's education initiatives (Patton, 2002).

Chapter 4:
Application to UNESCO MGIEP programmes

In this chapter, we will discuss how the Evaluation Research Framework is used to evaluate the effectiveness of UNESCO MGIEP's programmes, detailing the practical implementation of evaluation studies, the types of data collected, and the methodologies used to analyse the impact of these educational interventions.

4.1 Research design

4.1.1 Online courses

UNESCO MGIEP's online courses are a key component of its educational programmes, aimed at providing scalable and accessible learning opportunities to a global audience. These courses cover a variety of topics aligned with UNESCO MGIEP's mission to promote the development of peaceful and sustainable societies through education, including SEL, prevention of violent extremism, climate change, biodiversity, global citizenship, and media and digital literacy, among others. Evaluating the effectiveness, relevance and sustainability of these online courses is critical to ensure that they meet their educational objectives and provide meaningful learning experiences for participants.

Research questions and objectives

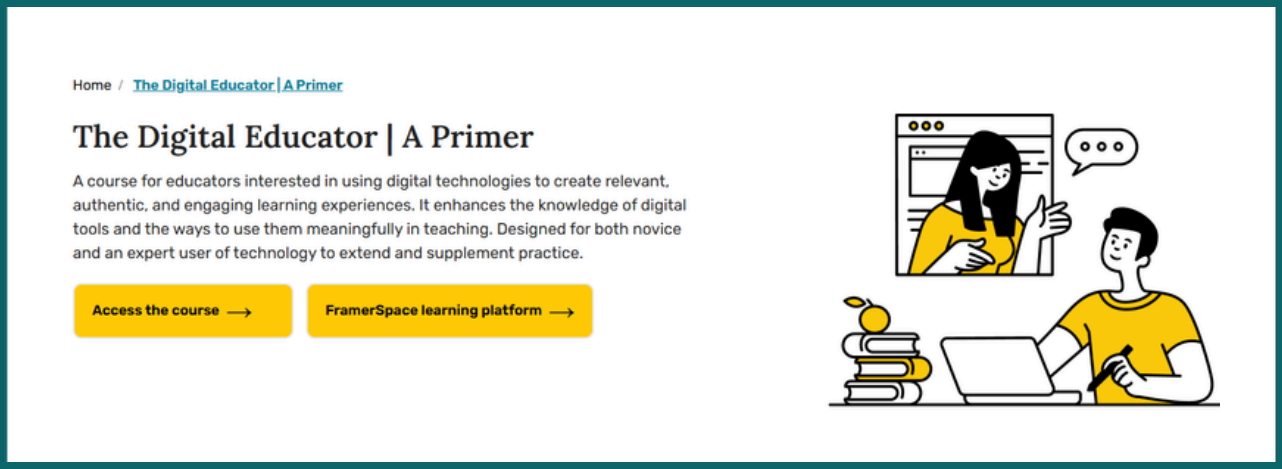
The evaluation of UNESCO MGIEP's online courses typically begins with the identification or determination of a clear list of intended learning objectives. These learning objectives are then used to design assessment scales that systematically assess all measurable learning objectives, including participants' acquisition of knowledge and skills, potential changes in attitudes and behaviours, and the overall quality of the learning experience. For example, an evaluation of a SEL online course might aim to answer the following questions: "To what extent do participants improve their knowledge of SEL principles after completing the course?" and "How do participants perceive the relevance and importance of the course topics after finishing the course?"

Research methods

UNESCO MGIEP uses a standard one-group pre- and post-test research design to evaluate its online courses. Pre- and post-tests are conducted through online surveys and are commonly used to measure changes in participants' course-related knowledge, skills, and attitudes. These evaluations often include a mix of questions rated on a Likert scale, open-ended questions, and knowledge assessment tests. For example, participants might be asked to complete an empathy self-efficacy scale before and after the course to assess changes in their confidence in their empathy competency.

Qualitative data are also collected through post-course feedback surveys. In some cases, post-course interviews and focus group discussions are also organized in which participants share their experiences, challenges, and suggestions for improvement. Additionally, UNESCO MGIEP will also start analysing course participation records including completion rates, engagement metrics (e.g., time spent on course modules), and course interaction data (e.g., discussion forum participation) to provide insights into participants' engagement with the course content and the overall learning process. Together these data will help UNESCO MGIEP evaluate not only the outcomes of its online courses but also the processes that contribute to these outcomes (Creswell & Creswell, 2017).

Importantly, wherever possible, an additional follow-up survey will be administered six months after the course completion to evaluate the potential longer-term effects of online courses, thereby enabling a more comprehensive assessment of lasting educational impact. Although such surveys have not been implemented yet, they will become a preferred option whenever feasible.



Case study: Digital Educator online course for educators

An example of UNESCO MGIEP's Evaluation Research Framework in action is the evaluation of Digital Educator: A Primer, the Institute's online course delivered as part of a teacher training programme in Sri Lanka. This course aims to improve teachers' digital pedagogical skills and knowledge. The teachers were encouraged to incorporate what they had learnt into their teaching practices to create more effective learning experiences. The evaluation study included pre- and post-course surveys to measure changes in the educators' attitudes, skills, and knowledge related to digital pedagogies.

The evaluation revealed a substantial increase in teachers' self-perceived competencies in digital pedagogies as well as increased awareness and appreciation of the importance of the course topic. Their knowledge of the course topics had also increased slightly, although not as much as expected. These findings were useful not only for course assessment but also for subsequent quality improvement, as they helped identify which aspects of the course's format and content needed improvement to make it more engaging and impactful.

4.1.2 In-person Workshops

In addition to online courses, UNESCO MGIEP offers in-person workshops designed to provide immersive learning experiences in areas such as SEL, digital pedagogies, media literacy and sustainable development. These workshops are typically conducted in collaboration with local partners and are customized to meet participants' specific needs whenever possible. Evaluating the effectiveness of these workshops is essential for understanding their impact and ensuring that they contribute to UNESCO MGIEP's diverse educational goals.

Research questions and objectives

In-person workshops are evaluated using research questions that focus on both the outcomes and the processes involved.
The questions are formulated to evaluate the most crucial aspects, ensuring that the findings are relevant and actionable (Stufflebeam & Shinkfield, 2007). Examples of such research questions are, "What effect does the workshop have on the participants' knowledge of sustainable development goals?" and "How effective is the workshop in facilitating changes in attitudes towards climate change?"

Research methods

To evaluate in-person workshops, UNESCO MGIEP uses a combined approach that includes surveys, observational methods, interviews and focus groups. Surveys are administered at the beginning and end of the workshops to track changes in participants' workshop-related knowledge, skills, and attitudes. These surveys often include self-assessments, where participants rate their own competencies or their self-confidence in areas such as emotional regulation, empathy, and digital pedagogies. Additionally, they incorporate performance tests that objectively assess respondents' knowledge or skills and, sometimes, behavioural intentions or typical actions. Observational data are collected during the workshops to assess participants' engagement with the activities and their interactions with facilitators and peer learners.

In addition, observational research studies in the real-life settings of the workshop participants have been pursued when possible.

These studies are especially important in the case of teacher training programmes, where teachers must apply the knowledge gained through workshops in their classrooms. For example, apart from the gains in participants' knowledge and changes in their attitudes, observational studies shed light on behavioural changes, thereby enabling a more comprehensive and policy-relevant assessment of the programme's effectiveness.

Post-workshop interviews and focus groups are also often conducted to obtain participants' feedback on the workshop experience, including what they found most valuable and areas for improvement. This qualitative data provides a deeper understanding of the workshop's impact, particularly regarding participants' personal growth and learning experiences (Patton, 2002).

4.1.3 Hybrid Education Programmes

UNESCO MGIEP often conducts hybrid programmes, which combine online courses and in-person workshops, in various subject areas and for diverse groups of learners. These programmes represent a unique and increasingly popular type of educational intervention, combining the flexibility and accessibility of online learning with the immersive, hands-on experience of in-person workshops. Evaluating the effectiveness of hybrid programmes is more challenging than evaluating individual online courses or in-person workshops because the impact of both the online and in-person components, as well as their interaction, must be assessed.

Research questions and objectives

The evaluation of hybrid programmes is guided by research questions that address both the individual and combined effects of the online and in-person components. Such questions help ensure that the evaluation captures the impact of both components of the programme (Creswell & Plano Clark, 2018). For example, key questions might include the following: "How do the online and in-person components of the programme complement each other in achieving the desired learning outcomes of the programme?," "What is the overall impact of the hybrid programme on participants' competencies?" and "How does the programme compare to stand-alone online courses or in-person workshops in terms of the overall impact?"

Research methods

Data collection for hybrid programmes involves a combination of methods used to evaluate online courses and in-person workshops, along with additional measures to capture interactions between the two components. Pre- and post-programme surveys are administered to assess changes in participants' knowledge, skills, and attitudes across the entire programme. In addition, an interim assessment is administered between the two educational components to track learning progress across the two steps and evaluate the effectiveness of individual programme components. These three surveys, include questions specific to both the online and in-person components, as well as items that address the overall learning experience. A typical schedule of assessment stages in the evaluation of UNESCO MGIEP's hybrid programmes is presented in Figure 2.

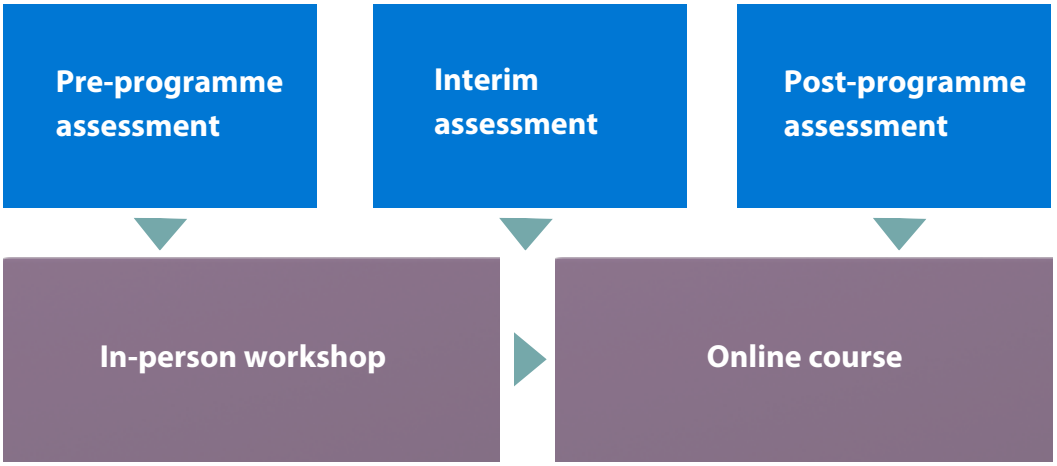


Figure 2: Assessment stages in the evaluation of UNESCO MGIEP's hybrid programmes (*Developed in-house by the author)*

In addition to the surveys, participant data from the online course platform, including engagement metrics and interaction data, as well as observational data from the in-person workshops are collected. This comprehensive data collection approach ensures that the evaluation captures both the direct effects of each component and the synergies between them. Qualitative data are collected through interviews and focus groups conducted both during and after the programme, providing insights into how participants perceive the integration of the online and in-person elements (Maxwell, 2013).

Case study: Hybrid SEL programme in Sri Lanka

A notable example of UNESCO MGIEP's hybrid programmes is the SEL and Digital Pedagogies programme, which combines two online courses designed specifically for teachers. The Social Emotional Educator course introduces SEL principles and their application in classroom settings. The Digital Educator course outlines key techniques and principles for using available digital tools and implementing them in classroom activities.
The programme also includes an in-person workshop that covers core topics from both courses, coupled with a range of practical exercises and activities designed to connect conceptual knowledge with real-life classroom situations. The evaluation of this hybrid programme aimed to assess both the individual impact of each component and the overall effectiveness.

The evaluation study included pre- and post-tests for each of the three components of the programme (an in-person workshop and two online courses). Data from the online platform and qualitative feedback from the questionnaire were also used. Importantly, this evaluation also included an additional observational study of the classroom behaviours of the participating teachers, which was administered both before and at the end of the programme.
The outline of the research design used in this evaluation study is presented in Figure 3.

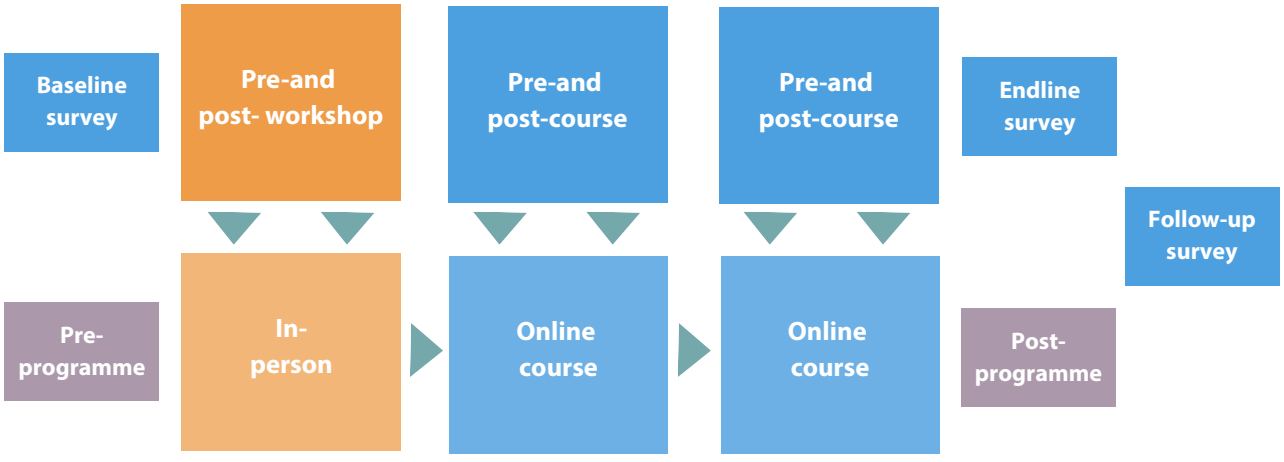


Figure 3: Research design used in the Sri Lanka hybrid SEL programme evaluation (*Developed in-house by the author)*

The results indicated a significant increase in participants' self-confidence in their pedagogical abilities and a greater appreciation for the topics covered by the two courses. However, the limited learning gains in knowledge warranted a deeper investigation into the actual type and scope of intervention impact.

4.2 Analysis techniques

Both quantitative and qualitative techniques are used to analyse the data from UNESCO MGIEP's educational programmes. Quantitative data from pre- and post-course assessments and background and feedback questionnaires are analysed using statistical methods. This includes calculating descriptive statistics, such as proportions, arithmetic means and standard deviations. Analysis to assess the psychometric properties of the scales use various methods such as identifying properties of individual items, checking scale dimensionality (principal component analysis and principal factor analysis) and scale reliability.

Programme effectiveness is calculated using various analyses, including paired t-Tests, Analysis of Variance (ANOVA), and regression analyses to determine the significance and magnitude of change in participants' competencies (Cohen, Manion & Morrison, 2018). These analyses provide quantitative estimates of the effectiveness of courses and help identify the aspects that led to positive changes as well as those that need to be improved.

Cohen's *d* is predominantly used to analyse effect size for the following reasons:

- **Magnitude indicator**: It measures the effect size of the difference between two means
- **Standardized metric**: It is expressed in standard deviation units
- **Versatile application**: It is used across various research domains
- **Interpretability**: Effects are categorized as small, medium or large
- **Dual metric**: It can also be expressed as percentiles

Importantly, instead of the usual focus on the statistical significance of the differences between the pre- and post-test values, UNESCO MGIEP's studies prioritize the examination of effect sizes. The main reasons for this are presented in Table 1. This is not to say that the Institute does not take statistical significance into account; rather, it goes a step further and asks how statistically significant the observed differences are.

Statistical Significance Testing *(p)*	Effect Size
Unreliable: Varies as a function of sample size	Not dependent on sample size
Uninformative: Forces continuous results into a dichotomy	Expresses the strength of the relationship in question
Arbitrary: Conventions (0.05/0.01/0.001) are based on rule of thumb, not on meaningful criteria	Continuous; can be compared or combined across studies

Table 1: Comparison of the characteristics of statistical significance and effect size testing

Published guidelines help determine what constitutes small, medium and large effect sizes empirically (Plonsky & Oswald, 2014). Importantly, these guidelines differentiate between within-group designs, which are almost exclusively used in UNESCO MGIEP's evaluation studies, and between-group designs, which are used in other research studies. However, it should be noted that in the past few years, effect-size interpretation in education has moved beyond generic thresholds and domain-specific L2 guidelines (Kraft, 2020; Simpson, 2021; Lee, Finn & Liu, 2019). For causal studies with standardized achievement outcomes, recent work proposes empirically grounded benchmarks calibrated to the distribution of field-based RCTs: small, <0.05 SD; medium, 0.05 to <0.20 SD; and large, ≥ 0.20 SD (Kraft, 2020). However, these should be adapted to the specific features of each study and consider costs and scalability.

These studies also caution against mechanical benchmarking and highlight sampling and sign issues. Recent analyses also show that a substantial share of interventions produce effect sizes <0.05 SD, underscoring the need to contextualize magnitudes using empirical distributions. Complementary approaches re-express effects in time (e.g., months of learning) to aid interpretation across grades and outcomes (Simpson , 2021; Greenberg, 2023; Kraft, 2020). This is why UNESCO MGIEP interprets effect sizes taking into consideration the context, design quality, and measurement reliability; we try to avoid rigid universal thresholds and instead triangulate based on practical significance and policy relevance.

Qualitative data from interviews and focus groups are analysed using qualitative analysis techniques, including thematic analysis, which involves identifying and coding patterns and themes within the data (Braun & Clarke, 2006). These analyses help to contextualize the quantitative findings with deeper insights into participants' experiences and perceptions. They can also indicate the factors responsible for the observed quantitative findings. The analysis of qualitative data is important because it can help inform the refinement of course content and delivery methods, ensuring that the courses remain relevant and engaging for diverse global audiences.

Chapter 5:
Challenges and limitations in implementing evaluation studies

Implementing UNESCO MGIEP's evaluation studies involves successfully managing a complex range of challenges and limitations. These arise from various factors, including the diverse contexts in which UNESCO MGIEP's educational programmes are implemented, the evolving nature of educational content, and the logistical constraints associated with conducting evaluations across multiple programmes, populations, and locations. Understanding and addressing these challenges is essential for ensuring that evaluation efforts are both effective and meaningful. This chapter explores the key challenges and limitations faced during UNESCO MGIEP's evaluation initiatives, including issues related to cultural diversity, logistical complexities, data collection and analysis, and the integration of evaluation findings into programme improvement. Strategies for mitigating these challenges are also discussed.

5.1 Cultural and inclusivity considerations (gender, disability, socio-economic status and other equity dimensions) and contextual sensitivity

UNESCO MGIEP's programmes are delivered to vastly different population groups including school children (12–18 years old), youth (18–30 years old), schoolteachers, and government officials across diverse national, cultural and socio-economic contexts, each with its own unique values, norms, and educational practices. This diversity creates challenges in designing and implementing evaluations that are culturally appropriate and that accurately reflect the experiences and outcomes of participants from different backgrounds (Kankaraš & Moors, 2010; Mertens, 2014). Thus, managing the cultural diversity and contextual sensitivity of its global educational programmes is a significant challenge to overcome in the implementation of future evaluation studies.

5.1.1 Challenges related to cultural sensitivity

Cultural sensitivity is crucial in the design of evaluation studies, particularly when developing evaluation instruments and methodologies that are relevant to and respectful of different cultural contexts. For example, concepts related to SEL may be understood and expressed differently in various cultures, which can complicate the development of standardized assessment tools (Kankaraš & Moors, 2010). Additionally, cultural norms around education, communication, and feedback can influence participants' responses to evaluation instruments, potentially leading to biased or skewed data (Cohen, Manion, & Morrison, 2018).

By incorporating cultural sensitivity into its evaluation studies, the Institute aims to produce findings that are both accurate and meaningful across diverse contexts (Fitzpatrick, Sanders, & Worthen, 2011). To address these challenges, the importance of culturally responsive evaluation practices is emphasized. This includes engaging with local stakeholders when designing evaluation tools to ensure that the instruments and methodologies are culturally

appropriate. In cases where tools that already exist are used, they are modified to ensure that they align with local cultural values and educational practices. When necessary, it also conducts pilot studies to test the suitability of these tools before full implementation.

5.1.2 Equity and inclusion considerations

Beyond cultural sensitivity, UNESCO MGIEP mainstreams equity across the evaluation cycle by explicitly considering gender, disability, socio-economic status (SES), language, rurality, migration status, and other intersecting identities with sampling plans powered to examine subgroup differences where feasible. Instruments and protocols are adapted to be accessible and fair (plain-language options, reasonable accommodations, and Universal Design –aligned administration), and indicators are designed to collect disaggregated data (e.g., by sex, disability, SES).

Mixed methods approaches are used to identify barriers and facilitators for under-represented groups, and analysis plans include checks for measurement bias and context-specific validity. Ethical procedures, such as locally appropriate informed consent, are used. During reporting, findings are interpreted through an equity lens, and limitations are documented transparently. This approach aligns with UNESCO's inclusion and gender priorities as well as its RBM framework on disaggregated data and equity-focused monitoring. It also conforms with the transformative evaluation tradition in education that foregrounds voice, participation, and social justice in study design and use (Ainscow, 2005; Mertens, 2009; UNESCO, 2017; UNDG, 2011).

5.1.3 Contextual variability

Cultural diversity produces variability in the contexts in which UNESCO MGIEP's programmes are implemented. Factors such as local educational policies, institutional infrastructure, availability of resources, and socio-economic conditions can individually and collectively impact the delivery and effectiveness of educational programmes, as well as the feasibility of conducting evaluations. For example, a programme that is successful in a well-resourced urban setting may face significant challenges when implemented in a rural area with limited access to technology or educational resources. These contextual differences may warrant changes in programme designs, which if undocumented, can in turn complicate the interpretation of evaluation findings and make it difficult to draw generalizable conclusions (Bryk et al., 2015).

UNESCO MGIEP addresses this challenge by using a flexible evaluation research framework that can be adapted to different contexts. This framework includes the use of mixed methods approaches that combine quantitative and qualitative data analyses to capture the full complexity of the educational environment. The assessment tools used are also designed to be adaptable, to suit the characteristics of the participating learners and their socio-economic backgrounds. By tailoring evaluation tools and methods to the specific conditions of each programme setting, UNESCO MGIEP ensures that its evaluations remain relevant and useful for guiding programme improvement (Creswell & Plano Clark, 2018).

5.2 Logistical complexities

The logistical complexities involved in implementing evaluation studies across multiple programmes and locations poses another significant challenge for UNESCO MGIEP. These complexities include coordinating data collection efforts, managing resources, and ensuring consistency in the implementation of evaluation protocols across different sites.

5.2.1 Coordination and resource management

Coordinating evaluation studies across multiple locations requires careful planning and resource management. Different programme implementations have different timelines, capacities, and logistical constraints, such as the availability of resources like trained personnel, infrastructure, and technology which can vary significantly between sites, further complicating the implementation process (Fitzpatrick, Sanders, & Worthen, 2011).

To address these logistical challenges and ensure consistency in data collection, UNESCO MGIEP takes these constraints into account during the planning and design stages of each evaluation study. This includes engaging with local partners that can provide support or carry out data collection for some of the aspects of the evaluation studies. The Institute also uses digital tools and platforms to facilitate communication and coordination between sites, ensuring seamless implementation of the assessment instruments.

These approaches, apart from ensuring efficient use of the Institute's resources, offer several advantages, such as evaluation designs and materials that are better suited to the local context as well as real-time data collection and analysis which enables UNESCO MGIEP's evaluation team to provide ongoing guidance and support to its programmes (Patton, 2008).

5.2.2 Ensuring consistency in implementation

Ensuring consistency in the implementation of evaluation protocols across different sites is essential to generate reliable and comparable data. However, this can be challenging in a global organization like UNESCO MGIEP, where programmes are implemented in diverse settings. Differences in the interpretation of evaluation protocols, resource availability and the level of experience among local enumerators can all contribute to inconsistencies in data collection and analysis (Cohen, Manion, & Morrison, 2018).

To address this challenge, UNESCO MGIEP has developed standardized evaluation protocols that provide clear guidelines for data collection, analysis and reporting. Furthermore, all aspects of the evaluation research are centrally managed and implemented,

where possible. This includes a fully centralized set-up and implementation of all assessment scales, along with data gathering, storage, and management. Likewise, all data analyses are conducted by UNESCO MGIEP researchers, ensuring consistency in the choice of procedures and quality control. UNESCO MGIEP also plans to conduct regular training sessions and prepare detailed manuals, templates, and other resources to ensure that all evaluation teams are well-equipped to implement the protocols consistently across different sites (Fitzpatrick, Sanders, & Worthen, 2011).

5.3 Psychometric quality

The value of the evaluation data, especially in the context of UNESCO MGIEP's educational programmes largely depends on the psychometric quality of the assessment scales used in the studies. This can be gauged through two key characteristics: the reliability and validity of the scales.

5.3.1 Scale reliability

Scale reliability represents the degree to which results would be stable across time and when using parallel forms of measurement instruments. It depends on the type of items used in this scale, standardized application of study procedures, sample size and sample variability, etc.

Scale reliability can be increased in several ways. The simplest of all is to use existing, well-tested, reliable scales whenever possible. Another method is to increase the number of items, provided they are good measures of the test construct. Using objectively scored and consistent item scales, such as a Likert Scale, also improves scale reliability. Finally, making sure that the scale is delivered in line with the envisaged procedures and under standardized conditions further reduces measurement error and thus improves scale reliability.

UNESCO MGIEP takes great care to standardize assessment conditions and procedures, ensuring a high level of consistency not only across scales in an evaluation study but also across studies themselves. All assessment scales used in UNESCO MGIEP's evaluation studies are objectively scored and are based on standard, easy-to-understand, and consistently applied response scales.

5.3.2 Scale validity

Scale validity is another key aspect of the psychometric quality of assessment scales. It refers to the extent to which an instrument actually measures what it is intended to measure. In the context of evaluation studies of educational programmes, scale validity ensures that the conclusions drawn from the data are meaningful and measure intended learning outcomes.

There are many types of scale validity, but the three primary types are as follows:

- **Content validity**: It assesses whether the scale covers the domain of the construct being measured.
- **Construct validity**: It evaluates whether the scale accurately represents the theoretical construct and is often tested through convergent and discriminant validity.
- **Criterion-oriented validity**: It examines the relationship between the scale and external benchmarks and can be further divided into predictive (future outcomes) and concurrent (current measures) validity.

In educational evaluation research, scale validity can be adversely affected in several ways:

- **Construct ambiguity**: Poorly defined constructs often result in scales that do not align with theoretical frameworks.
- **Cultural bias**: Scales developed in one cultural (or wider socio-economic or linguistic) context may not be valid in another (Kankaraš & Moors, 2010).
- **Response bias**: Social desirability bias, acquiescence responding, and extreme response responding can distort data (Kankaraš, 2010).
- **Improper scaling (low discrimination)**: Inadequate alignment of item difficulty with participant abilities may skew results.
- **Lack of a good convergent scale**: The absence of an existing, well-validated scale for a given construct makes it more difficult to establish the convergent validity of a new scale.

These drawbacks will be addressed using the following methods:

- **Precise conceptualization**: Constructs are clearly defined, and well-established theoretical models are used to guide scale development.
- **Pilot testing**: Wherever possible, pilot tests are conducted to identify ambiguous items and improve the validity of the scale.
- **Cultural adaptation**: The assessment scales are translated and adapted using culturally sensitive methods and expert review.
- **Advanced statistical techniques**: Exploratory and confirmatory factor analyses, and Item Response Theory (IRT) models are used to test construct validity and item functioning and discrimination.
- **Minimizing response bias**: This is done by incorporating balanced scales (positive and negatively worded questions), anonymity assurances and the use of well-tested response scales.
- **Quality assessment**: Quality control questions and attention checks are used to reduce biases and detect inattentive respondents.
- **Existing instruments**: Wherever possible, existing instruments with robust psychometric properties are used – if not directly, then at least to evaluate the concurrent validity of newly developed scales.

5.3.3 Scale reliability versus comprehensive coverage of learning outcomes

It is important to note that the reliability of psychometric scales is directly proportional to the number of items/indicators used in the scale: that is, longer scales will have higher reliability than shorter scales, assuming that the items used in these scales are of the same quality.

However, when educational programmes with many learning outcomes need to be evaluated – as in the case of UNESCO MGIEP programmes – it is usually impractical to use long versions of individual assessment scales. This is because with longer scales, the assessment sets quickly become too lengthy for respondents, which adversely affects the quality of their responses and overall participation rates. There are two ways of addressing this problem. On the one hand, in programmes that consist of a particularly large set of learning objectives, we divide the overall battery of assessment scales into two or three subsets, which are then randomly administered to respondents. However, a larger sample size is needed for this approach, and, as such, it cannot be used in programmes with fewer participants.

Another approach to addressing the issue of respondent fatigue is shortening the assessment scales. In UNESCO MGIEP's evaluation studies, number of items in the original scales were reduced to between three and six items in almost all cases. The original scales were almost always designed to be applied in research focusing on only one (or a few) construct(s) and, therefore, include an extensive list of items (often more than 20). Applying them in this format would ensure higher scale reliability but would limit the ability to assess all the learning outcomes of a programme. However, this approach reduces the content validity of the evaluation scale. Even with much shorter assessment scales, low participation rates and overall respondent fatigue have been encountered. Furthermore, the main purpose of the evaluation studies was to obtain a comprehensive assessment of targeted learning outcomes, so scales with very high reliability coefficients are not necessary.

Chapter 6:
Future directions and continuous improvement

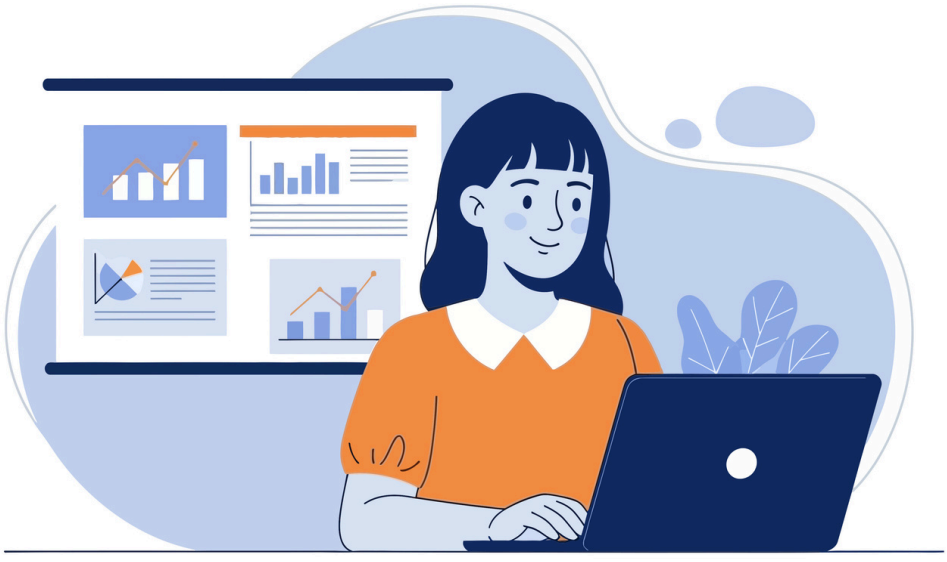
UNESCO MGIEP is committed to developing and refining its evaluation research practices and procedures to ensure the highest standards of educational quality and impact. As the Institute continues to expand its educational programmes and broaden its global reach, it aims to sharpen its focus on continuous improvement and innovation in evaluation methodologies and the integration of new technologies. This chapter outlines some of the possible future directions and recommendations for UNESCO MGIEP's evaluation research.

6.1 Expanding the evaluation framework

One of the primary proposed future directions for UNESCO MGIEP will be the expansion of its Evaluation Research Framework to include a broader range of educational programmes and outcomes. Although the current framework is robust and well-suited to the existing programmes, the evolving nature of education and the increasing complexity of global challenges necessitate a more comprehensive approach to evaluation. This expansion can involve developing new evaluation models that capture the multifaceted impacts of UNESCO MGIEP's programmes, including their long-term effects on participants and communities.

6.1.1 Incorporating new programme areas

As UNESCO MGIEP continues to develop new educational initiatives, particularly in emerging areas such as digital citizenship and sustainable development, the Evaluation Research Framework will evolve to incorporate these new programme areas. This will involve identifying relevant outcomes for each programme area, developing appropriate evaluation instruments, and ensuring that evaluation methodologies align with programme-specific goals and contexts (Patton, 2008). For example, in evaluating a new programme on sustainable development, the framework might include measures of participants' environmental knowledge, attitudes, and behaviours, as well as assessments of their engagement in sustainability practices.



6.1.2 Longitudinal evaluation studies

To better understand the long-term impact of its educational programmes, UNESCO MGIEP aspires to increase the use of longitudinal evaluation studies. Longitudinal studies are particularly useful for evaluating outcomes and, when appropriately framed, can assess longer-term impact (impact evaluations) in programmes aimed at fostering deep, lasting change, such as SEL or initiatives focused on global citizenship. These studies that track participants over extended periods, enabling the assessment of sustained changes not only in their knowledge, skills and attitudes but, even more importantly, in their behaviours and, eventually, in their life outcomes and achievements can help evaluators gain insights into how programmes influence participants' development and contribute to broader societal change (Fitzpatrick, Sanders, & Worthen, 2011).

6.1.3 Broadening the set of assessed learning outcomes

To avoid missing any potential learning outcomes, the evaluation scales should continue to be re-examined for their alignment with programme learning outcomes and expanded by including indicators for previously non-assessed characteristics.
This will enable a comprehensive evaluation of all potential learning outcomes, thus ensuring the validity of the conclusions regarding the programme's overall effectiveness.

6.2 Integration of new types of assessments

A key area for future development in evaluation research is the integration of new types of assessments that capture a wider range of learning outcomes. Traditional assessments, such as standardized tests and surveys, are valuable tools, but they do not always fully capture the complex, multifaceted nature of learning and development. To address this limitation, UNESCO MGIEP will continue exploring innovative assessment methods that provide a more holistic view of participants' progress.

6.2.1 Performance-based assessments

Performance-based assessments require participants to demonstrate their skills and knowledge in real-world or simulated scenarios. These assessments evaluate participants' ability to apply what they have learned in practical situations, which is particularly relevant for programmes focused on SEL, sustainable practices, and global citizenship. For example, a performance-based assessment might involve participants in a role-playing exercise where they must navigate a complex social situation, demonstrating their emotional regulation, empathy, and problem-solving skills (Wiggins, 1998).

6.2.2 Digital and adaptive assessments

Adaptive testing techniques have been widely implemented in education testing over the years. But they are also undergoing changes due to technological advances in the field of digital technologies and artificial intelligence tools. This opens new avenues for improving assessment approaches. If implemented in a scientifically sound and informed way, these technologies can not only improve the efficiency of tools and their psychometric properties but also provide real-time feedback to participants and educators in identifying areas of additional support (Shute & Ventura, 2013). Historically, UNESCO MGIEP has embraced digital technologies and houses an advanced digital learning platform through which its online courses are delivered, placing it in an advantageous position to deploy these technologies and accompanying data analytics in the coming years.

6.2.3 Social and Emotional Learning assessments

Given UNESCO MGIEP's emphasis on SEL, the Institute will aim to support the development of more nuanced assessments of SECs. These assessments can go beyond traditional self-report surveys to include tools such as peer assessments, parent and teacher reports, performance tests, behavioural observations, and digital SEL platforms that track participants' interactions and behaviours over time. By using a combination of these methods, a comprehensive picture of participants' SEC development can be obtained, leading to enhanced validity of the evaluation.

6.3 Adoption of advanced data analytics

The adoption of advanced data analytics can play a crucial role in enhancing the depth and accuracy of data-driven decisions, enabling the expansion of evaluation efforts. This could include deployment of powerful tools such as predictive analytics, machine learning, and data visualization to identify hidden patterns.

6.3.1 Predictive analytics

Predictive analytics involves the use of statistical models and machine learning algorithms to identify patterns based on historical data. UNESCO MGIEP will aim to explore the use of predictive analytics to identify factors that contribute to the success of its educational programmes and to predict participants' future performance in similar programmes. For example, by analysing data from past programmes, UNESCO MGIEP can identify key predictors of successful programme completion and tailor its interventions to support participants who may be at risk of dropping out (Siegel, 2016).

6.3.2 Machine learning applications

Machine learning offers additional opportunities to enhance UNESCO MGIEP's evaluation research. By applying machine learning algorithms to large datasets, the Institute can uncover complex relationships between variables that might be missed in traditional analysis methods. For example, machine learning can be used to analyse patterns of engagement in online courses, helping to identify which types of content or activities are most effective in promoting learning and retention. These insights can then be used to optimize programme design and delivery (Baker & Inventado, 2014).

6.3.3 Data visualization and reporting

Data visualization tools are becoming increasingly important for making complex data more accessible and actionable. Tools such as interactive dashboards and infographics can greatly enhance the outreach of research findings by making them more succinct, user-friendly, and decision oriented. Use of briefs and evidence notes generated through these tools to communicate key insights to UNESCO MGIEP's stakeholders can facilitate a deeper understanding of the evaluation results and their implications for programme improvement (Few, 2012).

6.4 Capacity building and professional development

Expansion and refinement of evaluation research require expertise in alignment with prevalent UNESCO norms and evaluation practices to maintain quality and responsiveness in emerging challenges and opportunities. This requires access to learning opportunities in advanced evaluation methodologies, data analytics, and emerging technologies, as well as encouraging knowledge sharing and innovation-oriented practices. Professional development in these areas will be made available for staff through workshops, online and offline resources, and collaborations with other institutions in the field.

Annex: List of UNESCO MGIEP online courses

Online courses for K-12 learners

Climate Change: Understand, Reflect, Empathize and Act

This self-paced interactive course is designed for middle school students and teachers. It helps them acquire a range of SECs to address climate change issues and work towards building a more sustainable future.

Biodiversity and Human Well-Being

This is a self-paced interactive course on biodiversity and human well-being. It consists of six modules that enable learners to understand the concepts of natural capital, inclusive wealth and ecosystem services and their connection to human well-being.

Global Citizenship

The course aims to build the knowledge, skills, and attitudes necessary for contributing to a more just, peaceful, inclusive, and sustainable world. The aim of the course is for learners to acquire knowledge regarding contemporary global challenges and develop the SECs needed to understand and address them.

Media Literacy: Fostering a Cognitive-Emotional Approach

This self-directed and interactive course aims to help learners gain a deep understanding of media literacy and recognize its significance both in their daily lives and for society as a whole. Employing the Libre pedagogical framework, which encourages various teaching methods, including reflection, discussion, inquiry, storytelling, simulations and games, this course can be seamlessly integrated into regular teaching and learning practices.

World Rescue

This game-based course comprises fast-paced gameplay set in Kenya, Norway, Brazil, India, and China. Participants will meet five

young heroes, help them solve global problems at the community level and gain knowledge and skills fundamental to making better decisions for the health of the planet.

Sky: Becoming Us

This is a free, interactive, self-contained, online game-based course for high school students to explore interpersonal interactions and cultivate prosocial skills and attitudes. Paired with the game Sky: Children of the Light, Becoming Us guides the students as they deliberate on concepts of society, humanity, trust, and cooperation.

Bury Me, My Love: Identity in Crisis

In this game-based course, participants follow the journey of a Syrian refugee, Nour, who undertakes a perilous journey to safety in Europe. This course helps in understanding a refugee's life and exploring concepts such as migration, home, belonging and identity.

Courses for Youth

Self-directed Emotional Learning for Empathy and Kindness (SEEK)

SEEK is a resiliency-informed course that helps in cultivating human values as skills, so we can thrive as individuals and as a society. This three-part online learning course includes practices such as reflective writing, contemplative journaling exercises and guided audio practices, which help participants master the required skills.

SEL for Youth Waging Peace

Aimed at youth peacebuilders, this course helps them cultivate critical SECs, develop systems-informed interventions to prevent violent extremism, and work towards building sustainable peace in their immediate communities.

Media Literacy: Fostering a Cognitive-Emotional Approach

This course helps learners gain a deep understanding of media literacy and recognize its significance both in their daily lives and for society. The course which employs the Libre pedagogical framework can be seamlessly integrated into regular teaching and learning practices.

Courses for Teachers

SEL for Schools

SEL for Schools is a certified two-part course for teachers: (1) SEL for Teachers helps build their own SECs (2) SEL for Classrooms provides teachers with a rich pool of hands-on learning activities to enhance their students' SECs, well known to also improve academic performance.

The Digital Teacher

This is for teachers interested in using digital technologies to create relevant, authentic, and engaging learning experiences. Participants enhance their knowledge of digital tools and learn how to effectively integrate them into teaching. It is designed for both novice and expert users of technology to extend and supplement practice.

The Social Emotional Educator: Primer

This course builds knowledge and awareness of SECs for educators and supports the application of pedagogical practices in classrooms that build these competencies in learners. It is rooted in the importance of positive relationships between educators, learners and the larger community.

The Digital Educator: A Primer

This course empowers educators in digital pedagogy. It focuses on leveraging technology as an ecosystem for improved learner agency across virtual and hybrid learning scenarios. Using Universal Design for Learning (UDL) principles, educators learn to incorporate digital tools in their teaching practices to create impactful learning experiences. The course integrates social-emotional learning competencies, promoting whole-brain learning.

Pilot Courses

SEL 4 Decisionmakers (SEL4DM)

This course is designed to bring SECs to the forefront of public service delivery by focusing on applying SECs to intrapersonal and interpersonal relationships and decision-making.

References

Ainscow, M. (2005) 'Developing inclusive education systems: *What are the levers for change?', Journal of Educational Change*, 6(2), pp. 109-124. https://doi.org/10.1007/s10833-005-1298-4

Baker, R. and Inventado, P. (2014) '*Educational data mining and learning analytics', Learning, Analytics, and Knowledge*, 3(2), pp. 61-75. https://doi.org/10.1007/978-1-4614-3305-7_4

Braun, V. and Clarke, V. (2006) '*Using thematic analysis in psychology', Qualitative Research in Psychology*, 3(2), pp. 77-101. https://doi.org/10.1191/1478088706qp063oa

Bryk, A.S., Gomez, L.M., Grunow, A. and LeMahieu, P.G. (2015) *Learning to improve: how America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Bryman, A. (2016) *Social research methods*. 5th edn. Oxford: Oxford University Press.

Cohen, L., Manion, L. and Morrison, K. (2018) *Research methods in education*. 8th edn. London: Routledge.

Creswell, J.W. and Creswell, J.D. (2017) *Research design: qualitative, quantitative, and mixed methods approaches*. 5th edn. Thousand Oaks, CA: Sage Publications.

Creswell, J.W. and Plano Clark, V.L. (2018) *Designing and conducting mixed methods research*. 3rd edn. Thousand Oaks, CA: Sage Publications.

Evans, R., Scourfield, J. and Murphy, S. (2015) '*The unintended consequences of targeting: young people's lived experiences of social and emotional learning interventions*', *British Educational Research Journal*, 41(3), 381-397. https://doi.org/10.1002/berj.3155

Few, S. (2012) *Show me the numbers: designing tables and graphs to enlighten*. Burlingame, CA: Analytics Press.

Fitzpatrick, J.L., Sanders, J.R. and Worthen, B.R. (2011) *Program evaluation: alternative approaches and practical guidelines*. 4th edn. Boston, MA: Pearson.

Greenberg, M.T. (2023) *Evidence for social and emotional learning in schools*. Palo Alto, CA: Learning Policy Institute.

Greene, J.C. (2007) *Mixed methods in social inquiry*. San Francisco, CA: Wiley.

Johnson, R.B. and Onwuegbuzie, A.J. (2004) 'Mixed methods research: a research paradigm whose time has come', *Educational Researcher*, 33(7), pp. 14-26. https://doi.org/10.3102/0013189X033007014

Kankaraš, M. (2010) 'Essays on measurement equivalence in cross-cultural survey research', Ridderkerk, Netherlands: Ridderprint.

Kankaraš, M. and Moors, G. (2010) 'Researching measurement equivalence in cross-cultural studies', *Psihologija*, 43(2), pp. 121-136. https://doi.org/10.2298/PSI1002121K

Kraft, M.A. (2020) 'Interpreting effect sizes of education interventions', *Educational Researcher*, 49(4), 241-253. https://doi.org/10.3102/0013189X20912798

Lee, J., Finn, J.D. and Liu, X. (2019) 'Time-indexed effect size for educational research and evaluation: reinterpreting program effects and achievement gaps in K–12 reading and math', *Journal of Experimental Education*, 87(2), 193-213. https://doi.org/10.1080/00220973.2017.1409183

Maxwell, J.A. (2013) *Qualitative research design: an interactive approac*h. 3rd edn. Thousand Oaks, CA: Sage Publications.

Mertens, D.M. (2009) *Transformative research and evaluation*. Thousand Oaks, CA: Sage Publications.

Mertens, D.M. (2014) *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods*. 4th edn. Thousand Oaks, CA: Sage Publications.

OECD (2019) *Better criteria for better evaluation: revised evaluation criteria definitions and principles for use*. Paris: OECD Publishing. https://doi.org/10.1787/15a9c26b-en

Patton, M.Q. (2002) *Qualitative research and evaluation methods*. 3rd edn. Thousand Oaks, CA: Sage Publications.

Patton, M.Q. (2008) *Utilization-focused evaluation*. 4th edn. Thousand Oaks, CA: Sage Publications.

Plonsky, L. and Oswald, F.L. (2014) 'How big is "big"? Interpreting effect sizes in L2 research', *Language Learnin*g, 64(4), pp. 878-912. https://doi.org/10.1111/lang.12079

Rossi, P.H., Lipsey, M.W. and Freeman, H.E. (2004) *Evaluation: a systematic approach*. 7th edn. Thousand Oaks, CA: Sage Publications.

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shute, V.J. and Ventura, M. (2013) *Stealth assessment: measuring and supporting learning in video games*. Cambridge, MA: MIT Press.

Sieber, J.E. (2009) *Planning ethically responsible research*. Thousand Oaks, CA: Sage Publications.

Siegel, E. (2016) *Predictive analytics: the power to predict who will click, buy, lie, or die*. 2nd edn. Hoboken, NJ: Wiley.

Simpson, A. (2021) 'Benchmarking a misnomer: a note on "Interpreting effect sizes in education interventions"', Educational Researcher, 52(3), 180-182. https://doi.org/10.3102/0013189X20985448

Stufflebeam, D.L. and Shinkfield, A.J. (2007) *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass.

Trochim, W.M. and Donnelly, J.P. (2006) *The research methods knowledge base*. 3rd edn. Boston, MA: Cengage Learning.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2017) *A guide for ensuring inclusion and equity in educatio*n. Paris: United Nations Educational, Scientific and Cultural Organization.

UNESCO, Bureau of Strategic Planning (BSP). (2013) *RBM monitoring and reporting guidelines* (BSP/RBM/2012/4 REV.2). Paris: United Nations Educational, Scientific and Cultural Organization.

UNESCO, Bureau of Strategic Planning. (2024) *Results-based management (RBM) approach as applied at UNESCO: guiding principles* (BSP/RBM/2008/1.REV.12). Paris: United Nations Educational, Scientific and Cultural Organization.

United Nations Development Group (UNDG). (2011) *Results-based management handbook: harmonizing RBM concepts and approaches for improved development results at country level* [Online]. Available at: https://unsdg.un.org/sites/default/files/UNDG-RBM-Handbook-2012.pdf (Accessed: 8 November 2025).

Wiggins, G. (1998) *Educative assessment: designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Yarbrough, D.B., Shulha, L.M., Hopson, R.K. and Caruthers, F.A. (2010) *The program evaluation standards: a guide for evaluators and evaluation users.* 3rd edn. Thousand Oaks, CA: Sage Publications.

Key Terms

- **Accountability**: The obligation to report, explain and be answerable for outcomes, particularly in educational evaluations.
- **Assessment**: The systematic collection and analysis of information to evaluate learning progress, programme outcomes or participant performance.
- **Behavioural Observation**: A method of systematically watching and recording participants' actions and behaviours in natural or controlled settings.
- **Benchmarking**: Comparing programmes or outcomes against best practices to identify areas for improvement.
- **Causal Inference**: Determining whether a programme or intervention directly causes observed changes in outcomes.
- **Cohen's *d***: A standardized measure of effect size used to express the magnitude of differences between two means.
- **Construct Validity**: The degree to which a test or scale accurately measures the theoretical construct it is intended to assess.
- **Control Group**: A group in experimental designs that does not receive the intervention, serving as a baseline for comparison.
- **Cultural Adaptation**: Modifying evaluation tools and methods to align with the cultural context and norms of participants.
- **Data Collection**: Gathering information through surveys, interviews or observations to support research or evaluation.
- **Digital Literacy**: The ability to effectively use digital tools and technologies for communication, information access and problem-solving.
- **Effectiveness**: The extent to which a programme or intervention achieves its intended outcomes under real-world conditions.
- **Effect Size**: A measure of the strength of a relationship or the magnitude of change produced by an intervention.
- **Empirical Evidence**: Information obtained through observation, experimentation or measurement, used to support or refute claims.

- **Evaluation Design**: The blueprint for conducting an evaluation, providing details of the methods, tools and processes to be used.
- **Evaluation Research**: A systematic approach to assessing the design, implementation and outcomes of programmes or interventions.
- **External Validity**: The extent to which evaluation findings can be generalized to other contexts or populations.
- **Feedback loop**: A process in which information about programme performance informs and improves subsequent implementation or outcomes.
- **Focus group**: A facilitated group discussion method to gather in-depth insights from participants.
- **Formative evaluation**: Evaluation conducted during programme development to provide feedback for real-time improvements.
- **Framework**: A structured model outlining the components, processes and goals of a programme or evaluation.
- **Goal setting**: Defining specific, measurable, achievable, relevant and time-bound objectives for a programme or intervention.
- **Holistic evaluation**: An approach that considers all aspects of a programme, including processes, outcomes and contextual factors.
- **Hybrid learning programmes**: Educational programmes combining online and in-person components for enriched learning experiences.
- **Impact assessment**: Evaluation of the long-term effects or outcomes of a programme.
- **Indicators**: Quantifiable measures used to assess progress toward achieving specific objectives.
- **Inferential statistics**: Statistical methods used to make generalizations or predictions about a population based on sample data.
- **Informed consent**: Ensuring participants agree to take part in research with full understanding of its purpose and procedures.
- **Internal validity**: The extent to which an evaluation accurately demonstrates that observed outcomes are caused by the intervention.

- **Knowledge transfer**: Applying knowledge gained from a programme to real-world settings, such as workplaces or communities.
- **Learning outcomes**: Specific knowledge, skills, attitudes or behaviours expected as a result of educational programmes.
- **Likert scale**: A survey format where participants rate their level of agreement with statements, often on a numerical scale.
- **Longitudinal stud**y: Research involving repeated observations of the same variables over an extended period.
- **Mixed-methods evaluation**: Combining quantitative and qualitative approaches for a comprehensive evaluation.
- **Outcome evaluation**: Assessment focusing on the impact and results of a programme or intervention.
- **Performance-based assessment**s: Assessments requiring participants to demonstrate their skills and knowledge in real-world scenarios.
- **Pre-post test design**: Evaluation design measuring participants' knowledge or skills before and after a programme.
- **Psychometric quality**: The reliability and validity of assessment tools used in evaluation studies.
- **Quasi-Experimental Design**: Evaluation design comparing outcomes between intervention and non-intervention groups, without random assignment.
- **Stakeholders**: Individuals or organizations with a vested interest in the outcomes of a programme or evaluation.
- **Validity**: The degree to which an evaluation accurately measures what it is intended to measure.

unesco

ICSSR Building, First Floor
35 Ferozshah Road,
New Delhi 110001

www.mgiep.unesco.org
+91 11 23072356-60