A Practical Guide for MLE Practitioners

DevInsights
accentuating insights for social development

Understanding Evaluation Designs
From Theory to Practice

www.devinsights.co.in

A -1, LGF 1, Sector 59, Noida, Gautam Budh Nagar, Uttar Pradesh – 201301
Telephone: +91 120-4217386
Website: www.devinsights.co.in

Printed in India
First Printing, 2024

Citation:
Dev Insights Private Limited, 2026, “A Practical Guide for MLE Practitioners” February 2026; New Delhi.

A Practical Guide for MLE Practitioners

Understanding Evaluation Designs
From Theory to Practice

CONTENTS

- Introduction - 1
- Programme Context: Adolescent Girls' Education Programme (AGEP) - 2
- Why Do We Need Evaluation - 2–3
- Types of Evaluation - 3–4
- Logic Model: From Inputs to Impact - 4
- Experimental Evaluation Designs (Randomised Controlled Trials) - 5
- Challenges in Implementing RCTs in Real-World Programmes - 6
- Quasi-Experimental Evaluation Designs - 6
- Difference-in-Differences (DiD) Approach - 7
- Matching Techniques in Evaluation - 8
- Cross-sectional and Longitudinal Designs - 9
- Evaluation Approaches Beyond Attribution - 9–10
- Key Takeaways for MLE Practitioners - 10
- Conclusion - 10

Understanding Evaluation Designs – Learning with AGEP (Barabanki)

This manual explains key evaluation concepts using practical examples, which is meant for new Monitoring, Learning, and Evaluation (MLE) professionals. The running example used throughout is an imaginary Adolescent Girls' Education Programme (AGEP) in Barabanki, Uttar Pradesh, India.

Case Scenario – AGEP in Barabanki

The Adolescent Girls' Education Programme (AGEP) supports adolescent girls who are at risk of dropping out from school. The programme focuses on girls in Classes 6 to 8 studying in government schools in rural Barabanki district of Uttar Pradesh, India.

The main objective of AGEP is to improve school attendance among adolescent girls and reduce dropout. It tries to do this by creating awareness, providing academic support, and increasing family and community support for girls' schooling.

Key activities of AGEP include:

- Weekly life-skills sessions for confidence building and awareness
- Subject support classes for weak subjects
- Parent meetings to encourage families to support girls' education
- Home visits for girls who are frequently absent

Girls in rural areas often miss school due to long travel distance, household responsibilities, safety concerns, and lack of study support. These issues make them more likely to drop out after primary school.

For this scenario, we assume that AGEP was implemented for one year across three blocks of Barabanki district. The programme team works with schools, teachers, parents, and community members to support girls' continued education.

Why Do We Need Evaluation?

At the end of a programme, we usually expect to see some positive changes. For example, attendance may increase or dropout may reduce. However, these changes may not be caused only by the programme. Many other things can happen at the same time.

Suppose AGEP completed one year of implementation and school data shows that attendance among girls has increased. Although this looks good, during the same year other events or schemes may have influenced the results.

Some examples include:

- Government scholarships for girls
- Distribution of free textbooks or uniforms
- Mid-day meal scheme improving attendance
- Better road connectivity increasing safety
- Post-pandemic return-to-school efforts

If we only do a 'before–after' comparison (attendance before AGEP vs attendance after AGEP), we may wrongly believe that AGEP caused the change. This is why evaluation is required.

Evaluation introduces the counterfactual question: "What would have happened if the programme did not exist?" This question helps us judge whether the programme contributed to the observed change.

The table below shows how one observation can have many explanations:

Observed Change	Possible Reasons (Not Only AGEP)
Attendance increased	AGEP + scholarships + mid-day meal + better roads
Dropout reduced	AGEP + family decisions + no early marriage + teacher follow-up
Marks improved	AGEP + tuition + parent support + teacher extra classes

Therefore, the purpose of evaluation is not only to show improvement but to understand the programme's true attribution. This makes decision-making more accurate and fair.

Types of Evaluation

There are different types of evaluation. Each type answers a different question. The three most common types are Process Evaluation, Outcome Evaluation, and Impact Evaluation.

Below is a simple table explaining the difference:

Type of Evaluation	Main Question	AGEP Example
Process Evaluation	Is the programme being implemented as planned?	Are life-skills sessions and home visits happening as scheduled, and • Are life-skills sessions being delivered with quality? • Are girls participating and understanding? • Etc.

Outcome Evaluation	What has changed for the beneficiaries?	Has girls' attendance improved after one year?
Impact Evaluation	Did the programme cause the observed change?	Did attendance improve because of AGEP, or would it have improved anyway?

Important points:

- Process Evaluation focuses on delivery quality, not results.
- Outcome Evaluation shows change but does not prove causation.
- Impact Evaluation tries to show causation by comparing with a counterfactual.

Many researchers get confused between outcome and impact. The difference is:

Outcome	Impact
Shows what changed	Shows if the programme caused the change
Does not require counterfactual	Requires a counterfactual or comparison
Short to medium-term results	Medium to long-term causal effects

For example, if AGEP girls' attendance increased from 60% to 72%, that is an outcome. But if we want to know whether AGEP is the reason for the increase, that becomes an impact question.

Logic Model – Inputs to Impact

To make evaluation easier, programmes often use a Logic Model. A Logic Model shows how a programme uses resources and activities to create change over time.

The common Logic Model sequence is:

Level	Simple Meaning	AGEP Example
Inputs	Resources used	Money, staff, training material
Activities	What the programme does	Sessions, meetings, home visits
Outputs	Direct products of activities	Number of sessions held, number of home visits
Outcomes	Changes in behavior or status	Improved attendance, confidence
Impact	Long-term change	Reduced dropout, higher completion rates

Explanation:

- Inputs are the resources required to run the programme.

- Activities are the actions performed using the inputs.
- Outputs are the immediate and direct results of those activities.
- Outcomes are the changes that happen because of the outputs.
- Impact refers to long-term and sustained changes in the system or society.

In evaluation, programmes usually achieve outcomes and contribute to impact. This means impact depends on many factors beyond the programme itself.

Experimental Design (RCT) – The Ideal Case

An Experimental Design, also called a Randomised Controlled Trial (RCT), is considered the strongest design for showing causation. In an RCT, two groups are created by random selection. One group receives the programme (treatment group) and the other does not (control group). If everything except the programme is similar between the two groups, then the difference in outcomes can be linked to the programme.

The key idea is that randomisation removes selection bias. This means that personal, social, or economic factors that influence attendance or dropout are equally distributed across both groups.

Important features of RCTs:

- **Random assignment:** participants or clusters are randomly allocated to treatment and control groups.
- **Highest internal validity:** internal validity means how confident we are that the programme caused the change. RCTs have high internal validity because both groups are similar at the start due to randomisation.
- **Clear counterfactual:** the control group represents what would have happened without the programme.

Example using AGEP:

If AGEP wanted to do an RCT, it could randomly assign some schools to receive the programme and other similar schools to not receive the programme. After one year, attendance in both sets of schools would be compared.

If the treatment schools show higher attendance compared to control schools, and both groups were similar at the start, then AGEP can claim that it caused the improvement.

Group	Receives AGEP?	Purpose
Treatment schools	Yes	To measure the effect of AGEP
Control schools	No	To represent what would happen without AGEP

RCTs are powerful, but they require strong planning and full cooperation from all stakeholders.

Why RCTs Are Hard in Real Life

Although RCTs are ideal for research, they are often not practical in social programmes. There are several reasons for this:

- **Ethical concerns:** It may not be acceptable to deny a useful programme to some groups just for research purposes.
- **Targeting logic:** Many programmes are designed to reach those who need it most. Random allocation may not match programme goals.
- **Operational issues:** Randomisation requires strict protocols, trained staff, and full cooperation from government and community.
- **Cost and time:** RCTs usually require large sample sizes, multiple surveys, and long follow-up periods.

Example from AGEP:

In Barabanki, AGEP may want to target girls who are at high risk of dropping out. Randomly assigning some high-risk girls to a control group without support may not be acceptable to families or schools. This creates ethical and operational barriers.

Also, government or donor agencies may expect the programme to operate in specific areas. In such cases, randomisation is not possible.

Because of these challenges, many evaluations use non-randomised designs that are more practical and still provide useful evidence.

Quasi-Experimental Designs (QED)

Quasi-Experimental Designs (QED) are used when randomisation is not possible. QEDs try to create a fair comparison between programme participants and non-participants by selecting groups that are similar in important ways.

QEDs do not use random assignment. Instead, they use matching or statistical control to reduce selection bias.

Common features of QEDs:

- No random assignment: groups are formed naturally or through matching.
- Use of comparison group: the comparison group acts as a counterfactual.
- Moderate internal validity: lower than RCT but still useful for causal inference.

Example using AGEP:

In Barabanki, AGEP schools could be compared to similar non-AGEP schools in nearby blocks. If the schools are similar at the start, differences in attendance after one year can suggest the effect of AGEP.

Design Type	Group Formation	Strength	Limitation
RCT	Random allocation	Very high internal validity	Hard to implement
QED	Matching or selection	Moderate internal validity	Unobserved differences may remain

QEDs are widely used in development programmes because they balance scientific credibility with practical feasibility.

Difference-in-Differences (DiD)

Difference-in-Differences (DiD) method compares changes over time between two groups. Instead of comparing the final values directly, DiD compares the difference in change between treatment and comparison groups.

This approach helps control for background changes that affect both groups.

Example using AGEP:

Group	Attendance Before	Attendance After	Change
AGEP schools	60%	72%	+12%
Non-AGEP schools	58%	65%	+7%

The programme effect is calculated as: 12% - 7% = 5 percentage points.

This means that although both groups improved over time, AGEP schools improved more. The extra 5 percentage points can be linked to AGEP.

Parallel trends assumption:

DiD assumes that if the programme had not been implemented, both groups would have followed similar trends. In simple terms, this means both groups were moving in similar directions before the programme started.

Advantages of DiD:

- Controls for time-related background changes
- Useful when baseline and follow-up data are available
- Suitable for programmes implemented at different times

DiD is widely used in policy and development studies because it is simple, intuitive, and practical.

Matching: A-priori and Propensity Score Matching (PSM)

Matching is used to make programme and non-programme groups similar at the start. The idea is to compare 'like with like' so that differences in outcomes are more meaningful.

There are two common matching methods: A-priori matching and Propensity Score Matching (PSM).

A. A-priori Matching (Design Stage)

A-priori matching is done before programme implementation. It involves selecting comparison groups based on known similarities.

- **Field-driven:** uses local knowledge to find similar areas or schools.
- **Transparent:** matching criteria are simple and understandable.
- Useful when data is limited: does not require complex statistics.

Example using AGEP:

AGEP could select non-AGEP blocks with similar poverty levels, school sizes, and gender ratios to serve as comparison blocks.

B. Propensity Score Matching (PSM)

Propensity Score Matching uses statistical models to calculate the probability of receiving the programme based on background characteristics.

Matching is done on covariates such as age, caste, socioeconomic status, and distance to school.

- Observed characteristics: PSM adjusts for factors we can measure.
- Cannot adjust for unobserved factors: motivation or parental support may still differ.

Key concepts in PSM:

Concept	Simple Meaning	AGEP Example
Covariates	Background features	Age, caste, distance to school
Baseline outcome	Starting level of result	Attendance before AGEP
Unobserved bias	Hidden differences	Motivation, parent attitude

PSM is useful when baseline covariate data is available. It helps reduce bias, but cannot remove differences that are not measured.

Cross-sectional vs Longitudinal Designs

Evaluation studies can collect data at one point in time or over multiple points in time. The choice affects what we can learn.

Cross-sectional design:

- Data is collected at one point in time
- Shows the current status, not change
- Cheaper and faster to conduct

Example: A survey in 2025 showing how many girls are currently attending school.

Longitudinal design:

- Data is collected from the same individuals over time
- Shows change and trends
- Helps identify who improved and who did not

Example: Tracking the same girls from 2024 to 2025 to see who stayed in school.

Type	What it shows	Strength	Weakness	AGEP Example
Cross-sectional	Current status	Quick and low cost	Cannot measure change	Attendance in 2025 only
Longitudinal	Change over time	Measures improvement	Costly, risk of dropouts	Attendance tracked 2024–2025

Other Evaluation Approaches (Beyond Attribution)

Not all evaluations aim to prove attribution. Some focus on learning, understanding pathways, or capturing unexpected results. These approaches are useful when programmes operate in complex settings with many actors.

Common approaches include:

Theory-based Evaluation

- Examines whether the programme's pathway of change is valid
- Useful for complex programmes with many activities

Example: Checking if awareness → parent support → attendance pathway holds for AGEP.

Contribution Analysis

- Tries to understand whether the programme contributed to the observed outcome
- Does not claim full responsibility

Example: AGEP may be one of several factors improving attendance along with government schemes and community efforts.

Outcome Harvesting

- Collects outcomes first, then investigates how the programme may have contributed
- Useful when results are not fully predictable in advance

Example: AGEP may influence girls to delay early marriage, even if this was not a stated objective.

These approaches are valuable for learning and adaptation, especially in real-world settings where strict attribution may not be possible.

DevInsights®
accentuating insights for social development

Contact us :

Phone :
+91 120-4217386

Email :
marcom@devinsights.co.in

Address :
A -1, LGF 1, Sector 59, Noida, Gautam Budh Nagar, Uttar Pradesh – 201301

Social Media :

Website :
https://devinsights.co.in/