Measuring What Heteronormative Systems Erase

Queer & Transgender Pedagogy, Context-Sensitive Evaluation, and Responsible AI in Pakistan & Afghanistan

Waqas Halim
Director, AI4Edu Lab, Toronto

Glocal 2026 | Interactive Evaluation Workshop
Duration: 60 Minutes
Audience: Evaluators, Researchers, Policymakers, Educators, Development Practitioners

Opening Question

Imagine two schools.
Both report: 95% attendance, 90% retention, and strong examination results.
School A: Students feel safe, respected, and able to express themselves.
School B: Students conceal identities, avoid participation, and experience exclusion.
To an evaluator, the schools appear identical.

Are they?

The Central Problem

Educational systems are becoming increasingly sophisticated at measuring access, attendance, and completion.
Yet they remain surprisingly weak at measuring belonging, dignity, safety, recognition, and agency.

The Result

We know more about participation than experience.

The Inclusion Measurement Gap

Definition

The gap between what institutions report and what learners experience.

Institutions Measure

- enrollment
- attendance
- completion

Learners Experience

- inclusion
- exclusion
- safety
- belonging
- dignity

Why This Matters

Educational systems shape futures.
Evaluation systems shape educational systems.
AI increasingly shapes evaluation systems.

Therefore
AI increasingly shapes futures.

The Problem of Visibility

We often assume visibility is positive.
Yet visibility can produce benefits and risks at the same time.

Benefits

- recognition
- resources
- protection

Risks

- surveillance
- exposure
- stigmatization

Visibility Is Not Neutral

Evaluators face a dilemma:

Too Little Visibility
People disappear from systems of evidence and policy.

Too Much Visibility
People may be exposed, targeted, or harmed.

Evaluation Challenge

How much visibility is enough?

Three Drivers of Invisibility

Structural
Institutional design can normalize binary assumptions.

Administrative
Data systems can reproduce narrow categories.

Epistemic
Ways of knowing can ignore experiences that do not fit existing frameworks.

Consequence
Absence from data often becomes absence from policy.

Heteronormativity as Infrastructure

Heteronormativity is not simply prejudice.

It is embedded within:

- forms
- databases
- indicators
- policies
- curricula
- evaluation systems

Evaluation Question

How do these assumptions shape what we measure?

Inclusion Lens

For every indicator ask:

1. Who becomes visible?
2. Who remains invisible?
3. Who may be harmed?
4. What assumptions exist?
5. What evidence is missing?

Use this lens before finalizing indicators, surveys, interview guides, or AI-assisted analysis plans.

Why Context Matters

Evaluations never occur in a vacuum.
Evidence is produced inside political, cultural, institutional, legal, and social contexts.

Context shapes:

- participation
- safety
- disclosure
- evidence quality
- interpretation

Pakistan Timeline

2009
Supreme Court recognition of rights of transgender citizens.

2011
NADRA recognition and identity documentation reforms.

2017
Census inclusion of transgender category.

2018
Transgender Persons (Protection of Rights) Act.

2023
Legal contestation following Federal Shariat Court decision affecting certain provisions.

Evaluation Question
Which of these milestones can educational evaluations actually detect?

Recognition Does Not Equal Inclusion

Legal Recognition

Institutional Inclusion

Social Acceptance

Educational Experience

These are not the same thing.

Evaluator Task
Assess lived realities rather than policy assumptions.

Afghanistan Case Study

A donor requests evidence about marginalized learners in a constrained environment.
Data collection may create risk.
Data absence may create invisibility.

Central Question

What should the evaluator do when complete evidence cannot safely be collected?

Group Exercise

Task

In small groups, decide what constitutes sufficient evidence when complete evidence cannot safely be collected.

Discuss

- What information is essential?
- What information is merely useful?
- What information could create harm?
- What proxy evidence could be safer?
- What decision will this evidence inform?

Identity Concealment

Why do people conceal identity?

- safety
- stigma
- family concerns
- distrust
- institutional consequences

Critical Principle

Concealment is often a rational safety strategy, not a failure of participants to cooperate.

The Disclosure Paradox

The populations most important to understand are often least likely to disclose.

Critical Principle

Absence of disclosure absence of population.

Evaluation Implication

Low disclosure may indicate low trust, high risk, or unsafe conditions - not low need.

Concealment Is Data

Many evaluators treat concealment as missing data.
A better interpretation: concealment may itself be evidence.

Evidence of:

- risk
- fear
- distrust
- exclusion
- lack of institutional safety

Do not ask only: Why did people not disclose?
Ask: What made disclosure unsafe?

Common Evaluation Challenges

Challenge	Why It Matters	Response
Concealment	Hidden populations	Measure experiences, not only identities
Fear of Harm	Ethical risk	Use risk-benefit assessment
Sampling Bias	Missing voices	Use multiple recruitment pathways
Political Sensitivity	Contested findings	Document assumptions transparently
AI Bias	Computational distortion	Validate outputs with local expertise

Five Questions That Keep Evaluators Awake at Night

1. What cannot safely be measured?
2. What is being mistaken for absence?
3. What matters but is not measured?
4. How much evidence is enough?
5. When should AI not be used?

Why Conventional Evaluation Fails

Traditional model:
Input

Output

Outcome

Problem
Human lives and social systems rarely behave this way.

Realist Evaluation

Core Question

What works?
For whom?
Under what conditions?
Why?

Use realist evaluation when context strongly shapes whether an intervention succeeds or fails.

The CMO Framework

Context

Mechanism

Outcome

Programs do not work.
Mechanisms work within contexts.

Evaluator Question
What context activates which mechanism to produce which outcome?

Example

Teacher training

Inclusive practices

Increased participation

Improved belonging

But the same training may produce different outcomes across schools depending on leadership, safety, norms, and trust.

Contribution Analysis

Move beyond:
Did it cause?

Toward:
Did it contribute?

Use contribution analysis when outcomes have multiple causes and direct attribution is unrealistic.

Contribution Story Builder

Intervention

Change

Outcome

Alternative explanations

Contribution claim

A good contribution claim is plausible, evidence-based, and honest about uncertainty.

Outcome Harvesting

The most important outcomes are often unexpected.

Outcome Harvesting is useful when:

- change is emergent
- outcomes were not predefined
- direct attribution is difficult
- the context is politically or socially sensitive

Outcome Harvest Exercise

Ask:

- What changed?
- Who changed?
- Why does it matter?
- How do we know?
- What evidence supports this?

Participant Output

One outcome their existing indicators may miss.

Productive Uncertainty

Responsible evaluators acknowledge uncertainty.
Not everything important can be measured.
Not everything measurable is important.

Evaluation Principle
The goal is not perfect knowledge. The goal is sufficient, ethical, useful evidence for decision-making.

Access Is Not Inclusion

A learner may be:

present
enrolled

Yet still:

excluded
unsafe
silenced

The inclusion question begins where access indicators end.

Four Domains of Inclusion

Access
Can learners enter?

Participation
Can learners engage?

Belonging
Do learners feel accepted?

Agency
Can learners influence their environment?

Capability Lens

Drawing on Sen and Nussbaum.

Traditional Question
Did learners attend?

Capability Question
What freedoms became possible?

Can learners participate safely, express themselves, pursue aspirations, and influence decisions affecting them?

Better Indicators

Measure:

- belonging
- dignity
- participation
- agency
- perceived safety
- trust in reporting systems

Not simply attendance.

AI as an Epistemic Filter

AI does not merely process information.

AI increasingly determines:

- what becomes visible
- what gets summarized
- what gets categorized
- what becomes actionable

Evaluation Question

What happens when evaluators begin seeing the world through AI-generated summaries?

What AI Does Well

AI can support:

- qualitative coding
- translation
- synthesis
- pattern detection
- document review
- rapid comparison across large text datasets

Use AI as analytical support, not as final interpretation.

What AI Does Poorly

AI struggles with:

- context
- lived experience
- ethical reasoning
- cultural interpretation
- political sensitivity

AI can help organize evidence, but it cannot carry moral responsibility.

AI Risks

Key risks for inclusion evaluation:

- bias amplification
- surveillance
- false objectivity
- exclusion
- misinterpretation of local terms
- over-reliance on computational outputs

Risk Principle
AI output must be validated with human and contextual expertise.

AI Decision Checklist

Before using AI ask:

- Necessity - Why use AI?
- Safety - Could harm result?
- Representation - Whose voices may be missing?
- Validation - How will outputs be checked?
- Accountability - Who remains responsible?

If accountability is unclear, do not use AI for sensitive evaluation judgments.

Future Reflection

Spivak asked:

Can the subaltern speak?

Evaluators may increasingly ask:

What becomes computationally visible?

Keep this as a provocation: AI may not only reproduce bias; it may shape what counts as recognizable evidence.

Toolkit 1: Indicator Audit

Purpose

To identify what an indicator reveals, conceals, and unintentionally normalizes.

What participants will do

1. Select one indicator from their own evaluation work.
2. Identify what the indicator currently measures.
3. Identify what it misses.
4. Revise it or add a complementary indicator.

Prompts

- What does this indicator reveal?
- What does it conceal?
- Who becomes visible?
- Who remains invisible?

Output

One audited indicator and one proposed improvement.

Toolkit 2: Inclusion Lens

Purpose

To examine whether an evaluation question or indicator reproduces exclusion.

What participants will do

1. Choose one evaluation question.
2. Apply the five inclusion questions.
3. Rewrite the question to reduce exclusion and risk.

Five inclusion questions

1. Who becomes visible?
2. Who remains invisible?
3. Who may be harmed?
4. What assumptions are embedded?
5. What evidence is missing?

Output

One revised evaluation question.

Toolkit 3: Risk-Benefit Matrix

Purpose

To decide whether sensitive data collection is ethically justified.

What participants will do

1. Identify one proposed data collection activity.
2. Rate the evidence value: low or high.
3. Rate participant risk: low or high.
4. Decide: proceed, add safeguards, simplify, or avoid.

Matrix

High Value + Low Risk = Proceed
High Value + High Risk = Safeguards Required
Low Value + Low Risk = Simplify
Low Value + High Risk = Avoid

Output

One risk-informed data collection decision.

Toolkit 4: Safe Evidence Framework

Purpose

To generate useful evidence without unnecessary exposure or extraction.

What participants will do

1. Identify the evidence they want.
2. Ask whether it is necessary for the decision.
3. Identify possible harms from collection.
4. Consider safer alternatives such as experience-based questions, proxy indicators, or anonymized methods.

Key Questions

- Is this information necessary?
- Could collection increase risk?
- Can experiences be measured instead?

Output

One safer evidence-generation strategy.

Toolkit 5: CMO Mapping

Purpose

To understand why an intervention may work differently across contexts.

What participants will do

1. Choose one intervention.
2. Identify the context in which it operates.
3. Identify the mechanism it is expected to activate.
4. Identify the outcome it may produce.

Template

Context Mechanism Outcome

Example

Supportive school leadership teacher confidence safer classroom participation

Output

One CMO map for an inclusion-related intervention.

Toolkit 6: Contribution Story Builder

Purpose

To build credible claims when direct attribution is impossible.

What participants will do

1. Name the intervention.
2. Identify the immediate change it may have produced.
3. Identify intermediate changes.
4. Identify the observed outcome.
5. List alternative explanations.
6. Draft a cautious contribution claim.

Template

Intervention Change Outcome Alternatives Contribution Claim

Output

One evidence-based contribution statement.

Toolkit 7: Outcome Harvest Worksheet

Purpose

To capture meaningful changes that were not predicted in advance.

What participants will do

1. Identify an unexpected change.
2. Describe who changed and how.
3. Explain why the change matters.
4. Identify evidence that supports the claim.
5. Decide whether the outcome should be added to the evaluation framework.

Questions

- What changed?
- Who changed?
- Why does it matter?
- How do we know?

Output

One harvested outcome and supporting evidence.

Toolkit 8: Capability Lens

Purpose

To shift evaluation from participation counts to meaningful freedoms.

What participants will do

1. Identify one existing access or participation indicator.
2. Ask what freedom the intervention is meant to expand.
3. Translate that freedom into a measurable or narratively assessable indicator.

Capability Questions

- Can learners participate safely?
- Can they express themselves?
- Can they pursue aspirations?
- Can they influence decisions?

Output

One capability-based indicator.

Toolkit 9: AI Evaluation Checklist

Purpose

To determine whether AI should be used in an evaluation and under what safeguards.

What participants will do

1. Choose one AI use case such as coding interviews, translating responses, or summarizing reports.
2. Assess necessity, safety, representation, validation, and accountability.
3. Decide whether to use AI, restrict its use, or avoid it.

Checklist

- Is AI necessary?
- Could AI create harm?
- Whose voices are missing?
- How will outputs be validated?
- Who is accountable?

Output

One AI use decision with safeguards.

Toolkit 10: Responsible Evaluation Canvas

Purpose

To synthesize the full workshop into one practical planning tool.

What participants will do

Complete a one-page canvas for a real or hypothetical evaluation.

Canvas Domains

- Visibility - Who becomes visible and who remains invisible?
- Safety - Could evidence generation increase risk?
- Inclusion - Whose experiences remain unmeasured?
- Context - What political, cultural, or institutional factors matter?
- Evidence - What evidence is sufficient and what is unnecessary?
- AI - Should AI be used and how will outputs be validated?
- Action - What decision will this evaluation inform?

Output

A draft Responsible Evaluation Canvas participants can take back to their organizations.

What Will You Take Back?

Participants should leave with:

- one revised indicator
- one revised evaluation question
- one risk assessment
- one AI assessment
- one organizational recommendation

This is a workshop, not only a presentation.

Personal Commitment

Prompt

One change I will make in my evaluation practice is:

_____

_____

_____

Participant Output

One individual practice change.

Organizational Commitment

Prompt

One evaluation practice my organization should reconsider is:

_____

_____

_____

Participant Output

One organizational recommendation to take back.

Five Principles for Inclusive Evaluation

1. Measure beyond access.
2. Treat concealment as information.
3. Prioritize safety over completeness.
4. Combine quantitative and qualitative methods.
5. Use AI as a tool, not an authority.

Tomorrow Morning Test

When you return to your organization, review:

- one indicator
- one survey question
- one AI-assisted process
- one assumption

Ask:

What does this make visible?
What does this make invisible?
Who could be harmed?

Key Takeaways

The Inclusion Measurement Gap is real.
Visibility is political.
Safety matters.
Context matters.
Concealment may be evidence.
AI requires human judgment.
Better evidence begins with better questions.

Resources

Evaluation

- BetterEvaluation
- EvalPartners
- Realist Evaluation - Pawson & Tilley
- Contribution Analysis - Mayne

Inclusion

- APA Guidelines for Psychological Practice with Transgender and Gender Nonconforming People (2015)
- APA Resolution on Gender Identity and Gender Expression (2021)
- UNESCO Inclusion Resources
- UNICEF Gender-Responsive Evaluation Resources

AI

- OECD AI Principles (2019)
- UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)

Discussion

What is the most difficult inclusion challenge in your context?

Possible areas:

- disclosure
- safety
- institutional resistance
- weak indicators
- AI use
- political sensitivity

Thank You

Waqas Halim
Director, AI4Edu Lab, Toronto

Fulbright Scholar | TEDx Speaker

Glocal 2026

Better evidence begins with better questions.