ISSN 2989-0071

APJE

Asia Pacific Journal of Evaluation

Volume 3 | Issue 02 | 2025

A Journal by
Asia Pacific Evaluation Association

APEA
Asia-Pacific Evaluation Association

ABOUT THE ASIA PACIFIC EVALUATION JOURNAL (APJE)

The Asia Pacific Journal of Evaluation (APJE) is the official journal of the Asia Pacific Evaluation Association. The journal publishes fully referred research papers and focuses on promotion of theory, methodology and practices of monitoring and evaluation. The journal primarily focuses on creating and promoting scholarly work highlighting the challenges and achievements in M&E in the region of Asia and the Pacific. APJE is aimed at readers in social work, nonprofits, academia, public sector, consulting firms, and donor agencies.

The Transformative Potential of Participatory Community-led Evaluation and Research: A Case from the Rural Philippines

Asia Pacific Journal of Evaluation,
Vol. 03, Issue 02,
July: 2025

Elene Cloete[1], Valerie Lejarde[2]
Elvis Gatchalian[3], Bernadette Estoesta[4]

Abstract

This practice note shares the experiences, reflections, and contributions of local stakeholders participating in a community-based WASH research initiative in the rural Philippines. The women leaders' reflections revealed that participatory research and evaluation conducted within the community by its own members has the potential to foster capacity-building and empowerment, cultivate a profound understanding of the community's felt issues and situations, promote the relevance and utilization of the evaluation results and data, and establish a sense of trust and community cohesion. Collaborative and participatory research and evaluation, therefore, represent an empowering and transformative alternative to traditional top-down versions.

Keywords: *Community-led Development; Community-led Monitoring and Evaluation; Rural Philippines; Community-led WASH Research*

[1] Outreach International
[2] Outreach Philippines, Inc.
[3] Outreach Philippines, Inc.
[4] Outreach Philippines, Inc.

I. Introduction

Community empowerment that is rooted in transformative and liberatory processes unfolds when practitioners champion dialogical, dynamic, and non-dominant approaches to community development (Lejarde, 2023). However, when research and evaluations are conducted in a top-down and prescriptive way, there is little room for dialogue, collaboration, and mutual engagement.

There are alternatives to such evaluation and research practices. Fetterman's Empowerment Evaluation, for example, prioritizes the active engagement of all stakeholders. The foundation of Empowerment Evaluation lies in the idea that individuals empower themselves, emphasizing that the "selection of inside facilitators increases the likelihood that the process will persist and be internalized in the system, creating opportunities for capacity building" (Fetterman, 1995, p. 181). Therefore, it promotes a scenario where participants acquire the knowledge and experience to conduct the evaluation, fostering empowerment rather than dependency on external evaluators (Fetterman, 1995).

Another evaluation alternative, Mertens's Transformative Evaluation, promotes social justice (Mertens, 2008) by making power hierarchies explicit and then using the evaluation outputs and experience to address such power differences (Mertens, 2021). Evaluation through a transformative lens engenders a deliberate inclusion of those in vulnerable and marginalized communities, employing culturally responsive methods, and addressing power disparities (Mertens, 2021).

Similar trends are active among applied researchers working within the international development sector. Community-based Participatory Research (CBPR) and Participatory Action Research (PAR), for example, prioritize equal participation and ownership, reciprocity, and co-learning among all stakeholders, with social change as the goal (Hacker 2017). By emphasizing inquiry, Opsina, Burns, and Howard (2021) prioritize learning, insisting on a learning-action-learning cycle as a research process and goal. Research approaches that champion meaningful participation, inclusion, and voice of all stakeholders, therefore, also hold the potential of not only empowerment but also strengthening communities' learning and collective action.

Transformation through research and evaluation activities does not happen automatically. Instead, organizations committed to such an agenda should create opportunities for participatory evaluation, explicitly emphasizing the importance of aligning evaluation practices with the diversity and exigencies of program contexts (Chouinard 2013, p. 248). This shift is crucial for cultivating a more democratic and culturally responsive program environment (p. 249). Therefore, by reframing their evaluation and research activities as opportunities also to empower and transform *all* stakeholders, organizations and funders alike, can strengthen

commitments to "shift the power"[5] and decolonize aid and development agendas. But what does cultivating and then promoting local ownership over data look like? And if such practices are implemented, how do local stakeholders perceive of it? What do they think about it, especially if they have never participated in monitoring, evaluation, and research activities?

Our practice note responds to these questions. In what follows, we introduce a recent collaborative WASH research initiative between two Philippines-based community groups, a local nonprofit and U.S.-based WASH researchers. We share themes emerging from reflection sessions with the local research team members, during which they described their experiences, major takeaways, and challenges. Our practice note is not about the WASH study and its results, even though it sets the stage for our discussion. Instead, it focuses on the grassroots researchers' experiences. While external research and evaluations are relevant to certain contexts, we argue that within the realm of community-led development, participatory research and evaluation activities hold empowering and transformative potential.

II. SIKAP, PSBA, and WASH

Outreach Philippines, Inc. (OPI) facilitates community-led development (Cloete and Dasig Salazar, 2022). Drawing from its Participatory Human Development (PHD) methodology, OPI practitioners organize communities to identify and then seek solutions for their poverty-related challenges. Corresponding solutions manifest as community-managed and own projects, with both issues and projects reflecting context-specific needs. To manage their projects sustainably, communities need social infrastructure. OPI practitioners, therefore, support communities in establishing community-based organizations (CBOs), with grassroots leaders and organizational systems needed for ongoing collective action. Such organizations include San Isidro Kaakibat sa Pag-unlad Association (SIKAP) and Pag-asa ng Sitio Bimmangon Association (PSBA), where co-authors Elvis and Beng have facilitated community organizing since 2019 and 2020, respectively.

III. SIKAP

Poor sanitation has been a pressing concern in San Isidro, Nueva Ecija Province, where SIKAP is located. To address this issue, the organization conducted an in-depth analysis of inadequate sanitation's impact on households. Recognizing the issue's urgency, SIKAP then designed and implemented a sustained community-led

[5] The Shift the Power Movement started in 2016 following the #Shiftthepower social media hashtag shared by The Global Fund for Community Foundations (GFCF) and their intentions of moving the Global Community away from "top-heavy and top-down systems of international development and philanthropy" toward more equitable people centered development (Global Fund for Community Foundations, n.d.)

sanitation loan project in 2021 (Toizer and Cloete, 2024). They built 20 new toilets and upgraded 26 existing units, leading to significant change, especially for women (Sorcher et al., 2023).

In October 2023, SIKAP, alongside WASH scholars, revisited sanitation again. Acknowledging the importance of continuous improvement, they were curious about enhancing their sanitation unit's efficiency and effectiveness, including evaluating their existing sanitation project. SIKAP, OPI, and the WASH researchers decided to implement a collaborative research project to scope existing needs, usages, and threads of WASH in their community. SIKAP identified ten women leaders to join the research team. The full team then reflected on their overarching research questions before refining the content of an existing and standardized survey tool. Following informal conversations and transect walks the research teams also decided to include questions related to fecal sludge management practices and the impact of climate change on sanitation. After finalizing the research tool, the SIKAP researchers received some additional training on data collection, including how to use Kobo Toolbox for both surveying and photo documentation, the researchers then practiced the survey among themselves and piloted the tool among an initial 15 households. The U.S.-based team reviewed the results for quality purposes.

To ensure a comprehensive sample, the SIKAP team's goal was to survey almost all households within their immediate surroundings (about 300 households). However, as awareness spread, households beyond the initial sample expressed interest in participating. The local team, happy about the interest shown by their fellow community members, expanded the sample to around 400 households.

IV. PSBA

PSBA, located in Cordon, Isabela Province, embarked on a parallel journey to confront their community's sanitation concerns. Acknowledging sanitation as a critical issue, they also conducted an analysis and initiated a community-led sanitation loan project in 2023. PSBA supported the construction of 59 new toilets and upgraded 18 existing systems, reinforcing safer and sanitary conditions for their primary women members. Inspired by the success of SIKAP's participatory research, an inquiry arose about PSBA's interest in a similar endeavor in their community. Responding affirmatively, 10 women volunteered to collaborate. Using the same survey employed by SIKAP, the PSBA team underwent training sessions to familiarize themselves with the survey tool, practicing it among themselves. Leveraging Kobo Toolbox for data collection and management, they deployed the survey, covering all 328 households within their community.

Notably, both SIKAP and PSBA have been conducting community-led monitoring and evaluation activities following OPI

capacity-strengthening programming. Therefore, the transition from participating in the community-led evaluation, including using applications like Kobo Toolbox, to the WASH research went smoothly.

The WASH scholars played a crucial role in supporting these communities in the data analysis process. After obtaining the evaluation results, SIKAP and PSBA validated the data with their fellow community members and local authorities. Significantly, this initiative has not only equipped the local groups with valuable information about their sanitation projects but has also provided them with a profound understanding of the WASH conditions within their communities. The findings have allowed these groups to identify deficiencies in their sanitation projects, notably the absence of access points for their latrines. In response, since the completion of this participatory evaluation, these community-based organizations have embarked on initiatives to encourage their members to upgrade their tanks, ensuring the incorporation of access points for future emptying purposes.

V. Finding and Discussions

Scholars and practitioners who are advocating for alternatives to mainstream evaluation and research argue that such alternatives hold potential beyond just gaining empirical evidence on project and program successes. They can be transformative, emancipatory, democratic, and empowering. But what do those participating in collaborative evaluation and research initiatives think about such activities, particularly those stakeholders typically excluded from such activities? Elvis and Beng used these questions as guide questions to organize and then facilitate two reflection sessions. Because we followed a constructivist and exploratory design, we used a data collection approach that would allow for open dialogue and discussion. In this case, reflection sessions with open-ended guide questions felt most appropriate. Elvis and Beng invited the leaders of SIKAP and PSBA (27 women leaders, which were all the women who participated directly in the research initiative) to participate in these sessions so that we could learn more about their experiences and challenges and to what extent they found the research and emerging data relevant. These conversations lasted about two hours. The leaders, Elvis, and Beng documented the women's consent and recorded their reflections (about 10 minutes of edited video recordings). We then conducted a thematic analysis, coding the reflections and then collapsing these codes into categories and, eventually, broader themes. These include *Strengthening Capacity, Knowing the Situation Better, Using the Data, Discovering the Self, and Fostering Trust and Community Cohesion.*

VI. Strengthening Capacity

One of the most prominent themes from the women's reflections is how their involvement in the research strengthened their capacity. The first area pertains to using *digital technology*. Using data management platforms like Kobo Toolbox marks a notable change from using pen and paper. In 2020, OPI's practitioners started to introduce communities to Kobo Toolbox as part of its community-led monitoring and evaluation programming. Members were trained in its usage, particularly for monitoring and evaluating their water, food, and sanitation projects. The training actively involved community leaders and members (Sorcher et al., 2022). Strengthening the capacity of leaders from marginalized communities to use technology for their development demonstrates an effort toward digital inclusion.

Leaders also shared how the initiative increased their *research capacity*. As one leader said, "I gained new knowledge, especially in formulating questions to gather sufficient and accurate data" (Leader 2). A significant aspect of this collaborative research initiative was involving community members in formulating the overarching research and survey questions and collecting data with the support of the WASH researchers and OPI team. This inclusiveness deviates from traditional top-down evaluation and research practices. Instead, it reinforces the exercise of local self-determination (Lowe & Wilson, 2017) and values the community's knowledge and agency. Additionally, this method serves as an empowering act of evaluation, where experts willingly share their knowledge and expertise with the community to "upskill and empower rather than keeping their knowledge to themselves and taking the data with them when they leave" (Cheek, 2007; Smith, 2012, as cited in Kelly, 2021, p.617). This enables the community to develop evaluation skills (Kelly, 2021) and ensures sustainability as the community retains these skills even after the experts depart.

VII.Using the Data

Community-led and participatory evaluation ensures that the results actively contribute to the continuous efforts and progress of the communities, aiding them in addressing their identified issues and aspirations. Over the past five years, PSBA and SIKAP leaders have been addressing their identified poverty issues, including sanitation concerns, leading to the initiation of projects involving the construction of latrines and water pumps. The reflections of three different leaders underscore the future tangible impact of their research data on activities related to resolving sanitation issues and enhancing the effectiveness of their sanitation projects. They shared that, "We can use the data in our next mobilizations" (Leader 1) and that they will "use [the] gathered data as a basis for writing project proposals and request letters to resolve [their] discovered sanitation issues (Leader 3)." The president of one of the

groups reflected that they "plan to use the data obtained from the research as a basis for creating the next plan to address issues related to water, sanitation, and health" (Leader 2).

Community-owned ownership means ownership not only over the tangible components of development initiatives and the process, but also the associated *data* of such processes and initiatives. The leaders' reflections underscore how they are perceiving such ownership, something crucial to ensure meaningful community development (Levine and Griñó 2015). They are realizing their rights and freedom to use their data for their own communities' benefits in ways that align with their capacities and goals. Such an approach nurtures community ownership — a vital marker of community empowerment — and ensures the data's relevance to individuals affected by poverty issues and project participants. This extends to their ongoing community-led initiatives, thereby increasing the likelihood of ongoing data utilization to resolve their felt poverty issues.

VIII. Knowing the Situation Better

As community members actively engaged in the research, they underwent a profound realization of their community's challenges and circumstances. This involvement led to heightened awareness regarding the prevalent sanitation and health issues affecting their locality. As articulated by the leaders:

> Because of this research, I understand better the situation of our community (Leader 2).
>
> We have learned that the problem regarding water, sanitation, and health is still extensive [in our community] (SIKAP leaders).

Furthermore, the leaders also realized that sanitation issues still affect many households within their respective communities. As they shared:

> We learned not only about ourselves but also about our community. During our data collection, I observed that many [of us] struggle to build proper toilets, and some are still unable to have their own. Regarding access to clean water, most households don't have their wells; there's only a communal well, and fetching water becomes a challenge to adequately meet the clean water needs (Leader 4).

Active participation in the research allows community members to immerse themselves even deeper into their communities' contexts and environment. This immersive approach exposes them to their communities' WASH challenges and increases their awareness of the

magnitude and breadth of such challenges. Furthermore, integrating research activities with experiential learning has deepened their understanding of community realities and strengthened their dedication to sustaining community development activities (Coughlin et al. 2017; Kolb 2014).

IX. Discovering the Self

Participatory research and evaluation activities can transform community members' sense of self, specifically women. Evidence of such a transformation emerged during the women's reflections. As one of the woman leaders articulated, her engagement in this evaluation activity made her realize something about herself and her co-leaders.

> As a researcher, I learned that I can indeed be a researcher. Previously, I thought the term "researcher" was reserved for those with higher education, for professionals. It turns out that even someone like me, a high school graduate, can become a researcher. All that's needed is an interest in learning and participating in activities like this. I believe my colleagues feel the same way (Leader 5).

Empowering individuals involves providing opportunities for learning and applying acquired knowledge into practical action. By providing such opportunities, community leaders can also start to recognize their own ability, efficacy, and agency. Affording them the chance to assert their right to play a pivotal role in their personal and community development is what truly empowers a community leader (Lejarde, 2023). Furthermore, Nawaz (2013) contends that "empowerment is something that has to be realized by the self" (p. 29) and that a woman can only empower herself as soon as she develops "strong feelings about herself" (p.29). The women leaders in the case study affirmed that they have learned something about themselves and their communities and manifested commitment through collective and continuous actions. Therefore, the participatory approach to evaluation, where the participants assume active roles and tasks, can provide transformative experiences that can empower the people.

X. Fostering Trust and Community Cohesion

Trust is paramount in eliciting authentic and accurate data from community members. In contrast to external enumerators and researchers, the internal researchers, or in this case the community leaders, have already established positive reputations in their community from past development activities. Their familiarity breaks the ice during data collection. It

encourages people to be more open and transparent when conveying data to internal evaluators, resulting in more relevant and reliable evaluation results. As affirmed by these leaders:

> We saw the trust of the majority of Bimmangon residents in us; I know that they need assistance to [resolve] their sanitation [issues] (Leader 6).

> Since the SIKAP Association is already known in our community, residents' participation in future research will likely be more accessible (Leader 2).

Participatory research and evaluation facilitate interaction among community members, thereby enhancing engagement in development activities. Leaders' reflections also revealed such interaction. As one leader shared, “One of the things I appreciated from this research is the opportunity to interact with the residents in our community, especially those I haven't met before” (Leader 1).

Participating in the WASH research activity also enables the SIKAP and PSBA leaders to develop shared values and goals, particularly sustaining their collective actions to address poverty issues. Fortified by experiential learning and capacity building, most leaders expressed their commitment to continuing efforts to resolve their sanitation issues.

> We discovered that the issues concerning water, sanitation, and health are more extensive than initially thought. Thus, we are still planning to take further action on this matter (SIKAP leaders).

> The entire PSBA organization will continue to work collectively to address our felt issues here in our community, especially those problems related to cleanliness, water, and sanitation, not only for our organization but also for future generations (Leader 5).

Participatory research and evaluation fosters community cohesion by creating opportunities for members to interact, garner support in evaluation activities, and actively contribute to development goals. Such interaction was also present in the work of SPPC et al. (2024) and their community-led monitoring and evaluation activities. Consistent engagement in these participatory evaluations within the community can elevate trust, cooperation, and a sense of belonging among members, thereby promoting and strengthening community cohesion.

XI. Conclusion

An emphasis on alternative forms of evaluation and research is timely because it closely aligns with global trends to decolonize aid, methodologies, and practice. Similarly, the push within the development sector toward localization and local ownership is rippling through the full sector with organizations, practitioners, and evaluators curious about how, where, and when they can adjust their methodologies. And as Levine and Griñó (2015) argue, this push toward local ownership should be visible throughout the full program cycle, with monitoring and evaluation practices being no exception. The local researchers and their experiences featured in this case study not only exemplify the ownership Levine and Griñó are arguing for, but it also evidences how such inclusion can have a more profound impact on local researchers than just knowing more about the success and impact of interventions. As the women's reflections confirm, participating in understanding and learning more about their communities and interventions transformed their minds about what their own capacities are about research and evaluation. Their reflections actively break down perceptions around who researchers are, because as the women have experienced, they too can evaluate, research, and discover more about their communities. After all, it is their communities, their neighbors and their challenges. By sharing our findings, we are encouraging practitioners and implementing organizations to take seriously the value of community-led and owned monitoring, evaluation, and learning and consider including it in their work.

By sharing this case study, we aimed to add to a growing body of case studies confirming the value of community-owned and led research and evaluation. One example of such work includes SPPC et al. (2024), featuring the work of another Philippines-based community group and their efforts to monitor and evaluate their newly implemented water system. Additionally, with this case study, we wanted to add the experiences of those directly involved in such research - in this case, the women from SIKAP and PSBA - to document the transformative contribution of collaborative and participatory evaluation and research practices. Too often, practitioners and researchers hope that their practices yield transformative results. Still, too frequently, they fall short of including local experiences in their follow-up reflections on how things went. Through this case study, we wanted to address this shortcoming by asking the local researchers from SIKAP and PSBA directly what it was like to be integral stakeholders and participants in research of and about their communities.

Our case study is not without limitations. First, it captures the perceptions of only the leaders directly involved in the research. However, the experiences of the actual research participants are not included here, i.e. those who responded to the surveys, the U.S-based WASH researchers and the practitioners

supporting the leaders in their broader community-level activities. Knowing what these other actors think about locally-led research and evaluation can deepen the conversation about why and how the sector should incorporate community-based and owned monitoring and evaluation into their practices. For example, it would be interesting to learn more from the survey participants about what it felt like to be surveyed by their fellow community leaders instead of external enumerators. By adding their voices to the analysis, our case study would have shed a broader and more representative light on the value of inclusive and community-led monitoring, evaluation and research. Future research into the locally perceived value of community-led research should, therefore, pay close attention to including a diverse set of voices.

A second limitation pertains to the collection of the leaders' perceptions. Due to time limitations on both the authors' and local community leaders' side, we collected only one set of reflections, which we then used for our analysis. A second round of conversations following our analysis would have allowed us to expand on the themes yielded during the first round, which, in turn, could have added additional depth to our findings and discussion. Planning for enough time, or considering dynamic data collection practices can therefore strengthen any future research on how local stakeholders perceive community-led monitoring, evaluation, and research.

A final limitation pertains to the context-specific nature of our case study. In this case, the perceptions featured here are specific to individuals who participated in the community-led interventions of Outreach Philippines Incorporated (OPI), their sanitation projects, and who live in the Philippines. While the findings and suggestions made in this case study are contextually specific, the goal of sharing such findings is to encourage practitioners and researchers further to broaden the collection of case studies on transformative evaluation to eventually allow for broader systematic reviews and meta-analyses and, therefore, generalizations on the impact of community-led and owned monitoring, evaluation, and learning.

XII. Recommendations

Through this case study, and by exemplifying the experiences of SIKAP and PSBA leaders in the rural Philippines, we hope to inspire practitioners and implementing organizations to consider and explore community-led and owned monitoring, evaluation, and learning. If they are committed to an agenda of transformation, localization, and shifting the power, one sure place to start is activating local ownership over program and project data. To ensure rich learning outcomes about implementing community-led monitoring, evaluation, and research, our recommendations to implementing organizations are to

1) Simply begin. Often, the status quo and that which is familiar feels safer and easier. But without starting, practitioners and their local community partners will never truly gain the experience of leading their own research activities.
2) Include and continue to include all stakeholders in all decision-making processes.
3) Make sure all stakeholders are clear on the rationale of the research. Such clarity will ensure local buy-in and strengthen ownership of the process and activities.
4) Guard against complicated and self-serving research goals and objectives. All should align with stakeholders' needs and intentions.
5) Facilitate ongoing check-ins with all stakeholders. Not only will this allow opportunities for all researchers to reflect on their activities, but also moments of course-correction. Check-in moments can also include data analysis to consider data quality and, if needed, include the actions needed to address quality concerns.

References

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation, 34*(2), 237–253. https://doi.org/10.1177/1098214013478142

Cloete, E., & Dasig Salazar, A. (2022). "To be one with others": exploring the development of community leadership in the Rural Philippines. *Development in Practice, 32*(6), 826-839. https://doi.org/10.1080/09614524.2022.2065244

Coughlin, S. S., Smith, S. A., & Fernández, M. E. (2017). *Handbook of community-based participatory research.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190652234.001.0001

Fetterman, D. M. (1995). In Response to Dr. Daniel Stufflebeam's: Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where It Needs to Go. *Evaluation Practice, 16*(2), 179–199. https://doi.org/10.1177/109821409501600207

Global Fund for Community Foundations. (n.d.) *Community Philanthropy and #ShiftThePower.* Blog, Accessed March 27, 2024. https://globalfundcommunityfoundations.org/what-we-stand-for/shiftthepower/

Hacker K. (2017) *Community-based participatory research.* London (UK): Sage. https://doi.org/10.4135/9781452244181

Kelly, L. M. (2021). A clash of values: Deep-rooted discord between empowering, participatory, community-driven development and results-focused, evidence-based evaluation. *Community Development, 52*(5), 607–623. https://doi.org/10.1080/15575330.2021.193610

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development.* Pearson Education.

Lejarde, V. A. (2023). Transformative Communication for Empowerment: Silence and Participatory Development in a Marginalized Community in Nueva Ecija. *PCS Review.* 279 - 315.

Levine, C. & Griñó, L. (2015). *Local Ownership in Evaluation. Moving from Participant Inclusion to Ownership in Evaluation Decision Making.* Report: Interaction.

Lowe, T., & Wilson, R. (2017). Playing the game of outcomes-based performance management. Is gamesmanship inevitable? Evidence from theory and practice. *Social Policy and Administration,* 51 (7), 981–1001. https://doi.org/10.1111/spol.12205

Mertens, D. M. (2021). Transformative Research Methods to Increase Social Impact for Vulnerable Groups and Cultural Minorities. *International Journal of Qualitative Methods*, 20. https://doi.org/10.1177/16094069211051563

Nawaz, F. (2013). Power, empowerment, and participatory development: Conceptual linkages. *Open Journal of Social Science Research, 1*(2), 26-30.

Ospina, S. M., Burns, D., & Howard, J. (2021). Introduction: Navigating the complex and dynamic landscape of participatory research and inquiry. *The Sage handbook of participatory research and inquiry*. London: Sage.

Samahan para sa Pagsulong at Pagunlad ng Capaculan (SPPC), Bobadilla, N. & Veda. G. (2024). Our Data, our Process: Re-imagining the Community-led Monitoring and Evaluation Discussion. *Asia Pacific Journal of Evaluation*, 2(1), 70-91.

Sorcher, R., Cloete, E., Salazar, A. D., Gatchalian, E., & Gonzales, J. (2023). Life has become more colourful: the social and economic outcomes of community-led sanitation initiatives in the rural Philippines. *Journal of Water, Sanitation and Hygiene for Development, 13*(6), 464-473. https://doi.org/10.2166/washdev.2023.042

Sorcher, R. Cloete, E. Bobadilla, N, & Penagunda, R. (2022). Piloting Community-led M&E Initiatives in Rural Philippines. *USAID Learning Lab Blog*. https://usaidlearninglab.org/community/blog/piloting-community-led-me-initiatives-rural-philippines

Swantz, M. L. (2008). Participatory action research as practice. *The Sage handbook of action research: Participative inquiry and practice*, 31-48.

Taylor, J., & Soal, S. (2003). *Measurement in developmental practice: From the mundane to the transformational*. Sao Paolo: Instituto Fonte.

AI In Evaluations: Opportunities, Implications, And Pitfalls – A Practitioner Review

Asia Pacific Journal of Evaluation,
Vol. 03, Issue 02,
July: 2025

Abhirup Bhunia[1]

Abstract

This practice note examines the rapidly evolving role of artificial intelligence (AI) in international development evaluation, highlighting both its transformative impact and its limitations. The note assesses current applications of AI across diverse use cases, organizations, geographies, and sectors; highlights evolving skills and attributes evaluators will need to cultivate to navigate the future of evaluation; and calls for continued exploration of both AI's untapped potential and the guardrails necessary to ensure its responsible use. A review of current applications indicates that AI is rapidly transforming evaluation practice, most notably by driving remarkable gains in speed and efficiency, but also analytical dependability albeit with occasional variability in consistency. Beyond mechanical automation, it has demonstrated analytical prowess in evaluative work, producing progressively more reliable outputs and, at times, surfacing insights that human analysts might overlook. GenAI analysis outputs now rival those of human evaluators. In at least one study, an AI-driven analysis tool outperformed established qualitative data analysis software. These type of advances are noted across sectors, contexts, organizations, and methodological approach (qualitative and quantitative), suggesting scalability. The growing adoption of enterprise AI tools among development actors is also a notable trend. However, limitations such as AI's bias (i.e., existing bias reinforcement), data quality issues, and algorithmic opacity underscore the continuing importance of human judgment, contextual understanding, and ethical oversight. As AI becomes more capable and dependable, evaluators must continuously upskill to use AI effectively, stay informed about its development, and understand the implications for its expanding presence in our shared 'evaluation futures'.

Keywords: *AI, artificial intelligence, LLM, evaluation, evidence synthesis, AI-augmented evaluation, MEL, evaluator competencies*

[1] Senior Independent Evaluation Expert

I. Introduction

In this practice note I explore the rapidly changing role of artificial intelligence (AI) in international development evaluation. I examine, with optimism, the promise of AI (and the consequent need for evaluators to upskill), while also striking cautionary notes where necessary. As AI evolves, its integration into evaluation is reshaping how data is collected, analyzed, interpreted, and how evidence is synthesized. AI now supports nearly the full suite of evaluation tasks, enabling evaluators to work more efficiently and productively, with real-world applications already demonstrating capabilities once unimagined. Beyond automating mundane tasks such as interview transcription, AI provides increasingly reliable support in high-value added tasks such as data analysis and evidence synthesis. These capabilities call for a shift in evaluators' skillsets and mindsets: one emergent skill is effective "prompting" for generative AI tools, while a key mindset is the ability to apply judgment and collaborate with AI as if it were a team member. Limitations of AI in evaluation mirror broader critiques, including algorithmic bias, unrepresentative training data, and opaque outputs. While AI-powered tools are rapidly improving efficiency and accuracy, AI still cannot be deployed unthinkingly, and human oversight remains essential. My appraisal of current practice suggests that AI will complement rather than replace the human expertise underpinning high-quality evaluation. Nevertheless, its role will only continue to expand, making it critical for evaluators to adapt and upskill in step with these transformations.

II. AI's Scalable, Efficiency-Boosting, And Increasingly Accurate Performance In Evaluations And Evidence Synthesis

For evaluators, AI promises to expand both the depth and scope of practice by improving efficiency and precision. The high-profile United Nations Evaluation Group Strategy 2025–2032 (UNEG, 2025), approved at the 2025 Annual General Meeting, explicitly identifies AI as a "potential evaluation influence," emphasizing its role in "contemporary evaluation approaches," its capacity for "resource efficiencies," and the need to "harness benefits while addressing risks."

In evaluation, AI is most visible in automating or optimizing tasks, from accelerating data analysis to generating insights. Its impact spans both quantitative and qualitative data, especially as evaluations increasingly contend with large, complex datasets. AI's ability to mine and synthesize vast data volumes has proved particularly valuable for evidence synthesis, where it improves speed, accuracy, and scale, though human expertise and oversight remain essential. For instance, a UNICEF (2024) case study showed that by converting 631 evaluation reports across sixty-four outcomes into machine-readable form, AI reduced text volume by 92%, allowing semi-automated text mining, searching, and highlighting to identify areas of

maximum impact. The process significantly reduced time and effort while maintaining reliability. Similarly, the World Bank (2023) applied machine learning to automate classification of implementation challenges in its private-sector projects. Using a taxonomy of 5 categories and 51 subcategories, the model identified key performance factors with ~70% accuracy comparable to human coders – but faster and more scalable. However, performance depended on training data quality and ongoing supervision, limiting immediate plug-and-play deployment.

Recent studies also highlight GenAI's potential and constraints. Clark et al. (2025) assessed its role in systematic reviews, finding tools "not yet reliable" for searching and screening but useful for simpler tasks like extracting metadata and assessing risk of bias in randomized trials. They struggled with nuanced judgments in non-randomized studies, reinforcing the need for human oversight despite rapid improvements. Nordmann et al. (2025), in a digital health scoping review of 407 articles, compared summaries generated by chatbots (ZenoChat, ChatGPT 3.5, ChatGPT 4.0, and ChatFlash) with those of two human researchers. Chatbots demonstrated strong contextual recognition, with one outperforming humans on completeness. Accuracy was high for simpler tasks but varied by model architecture, training, and prompt specificity, underscoring the importance of refined prompt engineering.

Automating tedious research tasks is a major driver of AI adoption in monitoring and evaluation. LateriteAI, for example, has developed large language model (LLM) based tools to extract survey data and code qualitative inputs, identifying groupings and meanings. Since 2022, UNDP's Independent Evaluation Office has used its in-house AI-powered analytics platform (AIDA) with a chatbot interface to browse over 6,000 of its reports, perform sentiment analysis, and generate visualizations. World Bank (2022) pilot tested AI-assisted content analysis across 392 project reports in sixty-four countries, where supervised machine learning replicated manual coding for high-level categories at scale. However, fine-grained labelling required expert pre-processing, limiting efficiency gains. Unsupervised methods showed strength in generating novel themes and uncovering patterns invisible to human analysts, though still requiring validation for interpretive depth.

In quantitative analysis, AI supports statistical modelling and pattern recognition. Raimondo, Ziulu, and Anuj (2023) tested GPT-based applications across tasks, finding ChatGPT effective at writing R scripts and improving code clarity, though highly dependent on precise prompting. In qualitative analysis, traditionally more reliant on human interpretation, AI's ability to identify themes in large textual datasets is increasingly recognized. Yang (2025) compared GPT-4 codes for sixty qualitative interviews with human coding, finding them reliable and scalable for theory-based evaluations. Hitch (2024)

argued that AI-augmented reflexive thematic analysis offers real benefits, demonstrated through ChatGPT examples, but cannot substitute the nuanced, contextual interpretation provided by human researchers.

Overall, AI now performs many tasks akin to a research assistant with machine speed and precision. Yet across evidence synthesis, quantitative, and qualitative domains, a scan of ongoing practice does reveal a continuing role of human judgment, oversight, and contextual understanding in ensuring evaluative quality.

III. Automation Of Mundane, Tedious Evaluation Tasks, Ai-Enhanced Traditional Data Analysis, Visualization Tools, And New-Gen Ai-First Tools

One of the most straightforward applications of AI in evaluation is automating repetitive, low-value tasks such as transcribing interviews, which has traditionally been time-consuming and resource-intensive. AI-powered tools like Otter.AI and auto-transcription features in Google Meets and Microsoft Teams now generate real-time transcripts with minimal human input, while other platforms seamlessly process audio from in-person interviews and focus groups. The efficiency gains are substantial. Evaluators save hours of manual data entry and can redirect effort toward deeper analysis; a benefit magnified in large-scale evaluations with many interviews or focus groups where human transcription would be slow and costly.

Beyond transcription, AI is increasingly integrated into mainstream analytical and visualization platforms. Atlas.TI, Tableau, and Power BI now embed AI features that automate analysis and enhance visualization. Atlas.TI's AI-powered auto-coding, for instance, categorizes qualitative data into machine-identified themes, reducing coding time and enabling evaluators to focus on synthesis, while Tableau uses AI to detect patterns and suggest visualizations that sharpen communication of findings.

New "AI-first" tools go even further. In early 2023, James Goh from MIT invited me to test an early version of their new tool AILYZE, and I was struck by its capability even then. Two years later, Harold (2025) rigorously evaluated AILYZE using anonymized transcripts from a Newcastle University study on vaccination decision-making and found it to significantly outperform NVivo – long considered the gold standard by evaluators as a qualitative analysis software. AILYZE reduced analysis time by 73% and was preferred in 76% of expert comparisons by human evaluators, excelling at surfacing nuanced themes such as civic responsibility, socioeconomic barriers, and complex relationships with healthcare authorities that NVivo often missed. The critical thing to note in this study is not just an obvious efficiency gain, but also AI's ability to serve as what the author calls "a catalyst for expanding the scope, depth, and impact of qualitative research across disciplines." While reinforcing AI's role as a powerful augmentation, it is also emphasized that an optimal approach would entail blending AI's capabilities with human interpretation and reflexivity.

These developments illustrate how AI is being embedded across every stage of evaluation. Yet they also underscore a steep learning curve: evaluators must keep pace with rapidly evolving tools to fully harness their potential.

Current AI use in evaluation leads to several clear conclusions:

- Efficiency gains from speed are undeniable, affecting how evaluation teams are resourced and highlighting new human skills evaluators must develop.
- AI applications across evidence synthesis, portfolio analysis, and both quantitative and qualitative analysis are viable, scalable, and reliable with expert human input and oversight.
- Leading academic and policy institutions widely apply AI in evaluation and evidence generation, reflecting its growing legitimacy.
- Human expertise remains indispensable, though AI accuracy is rapidly improving and its capabilities fast expanding.
- Quality of AI results depends also on training data, model architecture, and chatbot type.

The next section will discuss the new skills and attributes evaluators need to collaborate effectively with AI. Increasing enterprise-level AI integration – like UNDP's AIDA or the World Bank's supervised and unsupervised language models – further underscores the urgency for evaluators to master these skills.

IV. The Implications For Evaluator Skillsets And Capabilities

Alibaba chairman Joe Tsai recently said, "Equity research analysts can be completely replaced ... you can ask it to put together a report on Nvidia or Apple, and it'll do a great job." He adds, thankfully, "…then humans can apply their judgment and make better recommendations." Schopf (2025) echoes this view in the context of financial analysis. These insights are equally relevant for development evaluations, as discussed earlier. Given AI's rapid and significant advances, evaluators must now develop two essential skills: a) the art and science of prompting, and b) AI literacy and model understanding.

Effective prompting means crafting precise, context-aware queries for AI chatbots. Well-structured prompts lead to more accurate, relevant AI outputs that evaluators can use in their analysis and recommendations. This skill requires balancing knowledge of the evaluation context with an understanding of AI's technical capabilities and limits. For instance, when using AI to identify themes from many interview transcripts, evaluators must specify target themes or key evaluation questions. Vague prompts yield broad, unfocused results; overly narrow prompts may miss important nuances. Thus, AI becomes a powerful tool only when evaluators skillfully guide it.

Raheja and Srinivasa (2023) outline practical methods to improve prompting:

- Chain-of-Thought prompting: Encourages AI to generate step-by-step reasoning, boosting interpretability and accuracy. This is also vital when linking evidence to evaluative judgments.
- Self-Consistency: Samples multiple reasoning paths and selects the most consistent answer to improve accuracy.
- Least-to-Most prompting: Breaks complex problems into simpler subproblems, solving sequentially. This is helpful in addressing complex evaluation questions for example.
- Tree of Thoughts: Explores multiple reasoning paths like a decision tree, useful for analyzing causal factors in evaluations.
- Reasoning via Planning: Treats the AI as both a world model and reasoning agent to strategically explore key questions.

Lo (2023) offers the CLEAR Framework for Prompt Engineering: Concise, Logical, Explicit, Adaptive, and Reflective prompting. This means:

- Being concise and specific,
- Maintaining logical flow and clear instructions,
- Experimenting with different prompt formulations (adaptive), and
- Evaluating and refining prompts based on AI outputs (reflective).

Similar advice on prompting approach appears in Raimondo, Ziulu, and Anuj (2023) in their review of AI use in evaluations at the World Bank's DIME unit.

Prompting is not a one-off action but an iterative, reflective process. Just as inter-researcher triangulation refines qualitative analysis through dialogue, reflective prompting allows evaluators to treat GenAI as a co-evaluator enabling comparing, refining, and challenging its outputs as they would with a human colleague. This demands not only evaluation skills and contextual knowledge but also a sound grasp of AI itself.

Effective collaboration with AI hinges on literacy in how it works, especially explainability. Understanding data inputs, training methods, and decision processes helps evaluators interrogate AI outputs meaningfully, much like they critically examine evidence. Without such literacy, explainability risks being superficial. With it, evaluators can judge whether explanations are credible, engage more effectively with developers, and use AI tools with greater confidence and discernment.

Below I provide an evaluation scenario from a health program showing how AI literacy helps understand explainability in evaluation contexts:

- Scenario: An AI tool analyzes interview transcripts to identify key themes explaining why a health program succeeded or failed.
- With AI literacy: The evaluator understands that the AI's outputs depend on program data, but also prompt phrasing. Therefore, s/he further clarifies: "Which transcripts or phrases most influenced the theme identification? Could some themes be over- or under-represented? Why?"
- Explainability role: The AI provides a rationale showing it weighted

frequently mentioned barriers like staff shortages and community engagement challenges most heavily.

- Benefit: The evaluator can critically assess if these explanations align with their domain knowledge, ground reality in given context, or if important nuances were missed, improving trust and guiding follow-up analysis.

Tully, Longoni, and Appel (2025) argue that individuals with higher AI literacy, i.e., a basic understanding of algorithms, data training, and computational models, are better able to engage critically with AI. Instead of being dazzled by AI's outputs, these types of AI-literate users contextualize results realistically, fostering healthy scepticism. In evaluators, this mindset will enable better critical interrogation of AI-generated analyses.

Importantly, this level of AI literacy does not require deep technical expertise but rather the ability to mentally model how AI might generate outputs and a degree of comfort with how rapidly AI is evolving. Evaluators who visualize these mechanisms can better detect errors, challenge assumptions, and interpret results appropriately. For example, newer AI models aim for greater transparency through built-in interpretability, moving beyond "black box" systems. Sun, Oikarinen, and Weng (2024) introduce the Concept Bottleneck Large Language Model (CB-LLM), which balances scalability with clearer, traceable reasoning paths. This transparency is especially valuable in evaluation contexts, for instance when assessing if models use strategies like Automatic Concept Correction (ACC) to handle ambiguous terms. Since LLMs do not truly understand meaning but rely on probabilistic patterns, awareness of these features strengthens evaluators' scrutiny. While effective prompting remains the most critical user skill, foundational AI literacy enhances evaluators' capacity to use AI tools responsibly and insightfully. In fact, prompt engineering is a great way to ask the model to explain reasoning more explicitly. Newer GPT models for example are equipped with the ability to provide such answers.

From a broader perspective, Seker et al. (2025) find that AI literacy is strongly predicted by digital literacy and data literacy, with factors like English proficiency and AI usage frequency also being significant determinants. This raises important questions about skill inequality and who is best positioned to benefit from AI advances – is it evaluators in the global North with better tech access, higher English proficiency, and stronger digital infrastructure? Though crucial, this topic is beyond the scope of this discussion, but I suggest exploration in that direction in a future edition.

V. Limitations of AI

AI integration in the evaluation field is not without limitations, and these cut across evaluation tasks. For instance, transcription tools often misinterpret non-Western English accents. This is only an example, but reflective of a broader issue: AI models are trained on data that is not bias-free, and non-representative, which can lead to the perpetuation of racial, gender, class, or ethnic biases in analysis and evidence synthesis. These risks are compounded by the lack of algorithmic transparency. While AI bias has been widely studied in fields like healthcare

and public services, it remains underexplored in evaluations. There is a growing focus on fairness and ethical AI, but the evaluation field needs its own documentation of how biases manifest in the evaluation practice. It is important to note that the type of biases most widely damaging for valuations directly emanate from the wider AI biases rooted in algorithms and training data. There might still be value in specifically testing AI tools critically and identifying where and how biases emerge in evaluative work.

VI. Conclusion and Suggestions For The Evaluation Practice

As we have seen, AI can transform how evaluations are conducted, but its value will be realized only if evaluators develop the skills to use it effectively. From analysis to insight generation, AI can increase efficiency and accuracy, yet human judgment, critical thinking, and contextual understanding remain indispensable (yet). To work effectively with AI as an emerging "team member," evaluators must cultivate skills in prompting and build progressively deeper understanding of AI itself, and how its capabilities grow. The ability to collaborate and "pair" with AI is critical. Equally important is demonstrating evaluator's "value-add" beyond AI's obvious and growing capabilities, showing not that one competes with AI, but leverages it to amplify and enhance the evaluation field and practice itself. Below are a few concluding suggestions for the evaluation practice.

Explore AI in high-value evaluative tasks: AI's performance in high-value functions such as generating evidence-based recommendations remains underexplored. Unlike consumer recommendation engines (e.g., Spotify, YouTube), which rely on pattern recognition and reinforcement learning to suggest content based on past behaviour, evaluative recommendations require contextual understanding, creativity, foresight, and ethical judgment. While the underlying architectures differ, both systems operate as input-output engines, and with thoughtful prompting, AI can potentially assist evaluators by synthesizing evidence into draft recommendations. But this should be compared against human evaluator outputs to assess performance in several contexts to generalize a conclusion on performance.

Map current capabilities and degree of human interface: Practitioners should conduct a situational analysis to map AI's role across the evaluation cycle from design, tool development, and data collection, to analysis, synthesis, sensemaking, evaluative judgment, and recommendations. Capabilities might be classified as "established," "emergent," or "untested," while supervision needs could range from "light" to "significant." Such rubrics can guide systematic reviews of AI's use in evaluation.

Study AI risks and biases in evaluation: The risks of rapid AI integration require close study, particularly how biases emerge in evaluative work. This calls for evaluators to apply AI consistently, document instances of bias, and compare these against purely human approaches to assess when AI might amplify or accentuate biases, mislead, and discriminate.

References

Bravo, L., Hagh, A., Joseph, R., Kambe, H., Xiang, Y., & Vaessen, J. (2023). Machine learning in evaluative synthesis: Lessons from private sector evaluation in the World Bank Group. IEG Methods and Evaluation Capacity Development Working Paper Series. Independent Evaluation Group. Washington, DC: World Bank.

Cameron, H. (2025, March 18). Breaking up with NVivo: AI delivers faster, deeper, and better qualitative insights. SocArXiv. https://osf.io/preprints/socarxiv/

Clark, J., Barton, B., Albarqouni, L., et al. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. Research Synthesis Methods, 16(4), 601–619. https://doi.org/10.1017/rsm.2025.16

CNBC. (2025, March 12). Alibaba's Tsai on AI research: Analysts can be completely replaced. CNBC.https://www.cnbc.com/2025/03/12/alibabas-tsai-on-ai-research-analysts-can-be-completely-replaced.html

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023, August 21). GPTs are GPTs: An early look at the labor market impact potential of large language models. arXiv. https://doi.org/10.48550/arXiv.2303.10130

Franzen, S., Quang, C., Schweizer, L., Budzier, A., Gold, J., Vellez, M., Ramirez, S., & Raimondo, E. (2022). Advanced content analysis: Can artificial intelligence accelerate theory-driven complex program evaluation? IEG Methods and Evaluation Capacity Development Working Paper Series. Independent Evaluation Group. Washington, DC: World Bank.

Hitch, D. (2023). Artificial intelligence augmented qualitative analysis: The way of the future? Qualitative Health Research, 34(7), 595–606. https://doi.org/10.1177/10497323231217392

Leo, S. L. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. The Journal of Academic Librarianship, 49(4), 102720. https://doi.org/10.1016/j.acalib.2023.102720

Nordmann, K., Sauter, S., Stein, M., et al. (2025). Evaluating the performance of artificial intelligence in summarizing pre-coded text to support evidence synthesis: A comparison between chatbots and humans. BMC Medical Research Methodology, 25, 150. https://doi.org/10.1186/s12874-025-02532-2

Raimondo, E., Ziulu, V., & Anuj, H. (2023, August 23). Fulfilled promises: Using GPT for analytical tasks. World Bank Independent Evaluation Group (blog). https://ieg.worldbankgroup.org/blog/fulfilled-promises-using-gpt-analytical-tasks

Raheja, T., & Srinivasa, J. (2025, June 17). Mastering the art of prompting for large language models: Reducing hallucination and improving reasoning. Outshift by Cisco (blog). https://outshift.cisco.com/blog/mastering-the-art-of-prompting-for-large-language-models-reducing-hallucination-and-improving-reasoning

Schmidt, L., Addis, P., Mattellone, E., O'Keefe, H., Nabiyeva, K., Huynh, U. K., Dehingia, N., Craig, D., & Campbell, F. (2024, August 28). Evaluation synthesis analysis can be accelerated through text mining, searching, and highlighting: A case study on data extraction from 631 UNICEF evaluation reports. medRxiv (preprint). https://doi.org/10.1101/2024.08.27.24312630v3

Schopf, M. (2025, April 15). Outperformed by AI: Time to replace your analyst? Find out which GenAI model does it best. SSRN. https://ssrn.com/abstract=5222427

Seker, O., Kwon, K., & Kocak, O. (2025). Exploring researchers' artificial intelligence (AI) literacy: The mediating role of digital literacy and data literacy between 21st century skills and AI literacy. Information Development, 0(0). https://doi.org/10.1177/02666669251336368

Stoelinga, D. (2023, March 24). Developing LateriteAI – Technology to automate tedious research tasks. Itad (blog). https://www.itad.com/article/developing-lateriteai-technology-to-automate-tedious-research-tasks/

Sun, C. E., Oikarinen, T., & Weng, T. W. (2024, July 5). Crafting large language models for enhanced interpretability. arXiv. https://doi.org/10.48550/arXiv.2407.04307

Tully, S. M., Longoni, C., & Appel, G. (2025). Lower artificial intelligence literacy predicts greater AI receptivity. Journal of Marketing, 89(5), 1–20. https://doi.org/10.1177/00222429251314491

United Nations Development Programme. (2023, September 20). New analytical capabilities launched for UNDP IEO's AI powered platform, AIDA. UNDP Independent Evaluation Office. https://www.undp.org/evaluation/press-releases/new-analytical-capabilities-aida

Integrating Indigenous Evaluation Insights into the Master of Arts in Development Evaluation Program: Lessons for Culturally Responsive M&E Education

Asia Pacific Journal of Evaluation,
Vol. 03, Issue 02,
July: 2025

Almas Fortunatus Mazigo[1]

Abstract

This paper examines the integration of Indigenous evaluation insights into the Master of Arts in Development Evaluation (MADE) program at the University of Dar es Salaam to promote culturally responsive and ethical practices. The MADE curriculum was collaboratively reviewed with Indigenous communities, educators, and evaluation experts, resulting in a transformative program that incorporates Indigenous philosophies, participatory methods, and culturally defined success indicators. Ethical frameworks grounded in reciprocity and community data ownership ensure evaluations benefit communities and address historical inequities. Experiential learning further prepares students to work authentically in diverse contexts. Graduates are thus equipped to design evaluations that amplify the voices of marginalized individuals, promote equity, and uphold cultural integrity. This initiative exemplifies efforts to decolonize monitoring and evaluation education, aligning with African Evaluation Principles and global development goals, and sets a precedent for reimagining evaluation approaches that advance social justice.

Keywords: *Development Evaluation, Indigenous Evaluation, Culturally Responsive Evaluation, Monitoring and Evaluation Education, Decolonizing Evaluation, Ethical Evaluation Practices*

[1] Dar es Salaam University College of Education (DUCE), Tanzania, mazigo.almas@udsm.ac.tz

I. Introduction

Monitoring and Evaluation (M&E) is central to evidence-based decision-making in development. However, conventional approaches remain rooted in Western paradigms that often marginalize Indigenous perspectives. Growing recognition of these limitations has created both a scholarly and ethical imperative to integrate Indigenous insights into M&E education and practice.

This paper examines how Indigenous evaluation philosophies, methods, and ethics were systematically embedded into the Master of Arts in Development Evaluation (MADE) program at the University of Dar es Salaam. The focus is on curriculum transformation—how program design and pedagogy were reshaped to foster culturally responsive, ethically grounded, and contextually relevant training.

By documenting this process, the paper offers valuable lessons for educators, curriculum designers, and institutions seeking to decolonize M&E education. Grounded in Tanzania but linked to broader debates, the analysis engages with the African Evaluation Principles (AfrEA, 2021) and aligns with global development goals.

Section II describes the philosophy and principles of Indigenous evaluation. Section III examines international case studies that integrated Indigenous perspectives. Section IV explains the MADE curriculum review. Section V examines the ethical and cultural implications, as well as global comparisons. Section VI concludes with insights on how the MADE model can inspire systemic change in M&E education.

II. Indigenous Evaluation: Philosophy, Principles, and Aspects

Indigenous evaluation, rooted in worldviews, knowledge systems, and cultural practices, challenges conventional Western frameworks that favor standardized metrics and objectivity. At its core, it is relational—linking evaluation to land, community, spirituality, and accountability through relationships (Wilson, 2008). Success is understood holistically, encompassing physical, mental, spiritual, and social well-being, rather than being limited to isolated outcomes or economic indicators (Durie, 1994). This approach calls for evaluations that are not only technically rigorous but also culturally meaningful and socially just.

A defining feature of Indigenous evaluation is its insistence that communities determine their own priorities, standards, and measures of success. This contrasts with external, donor-driven approaches and affirms the right of Indigenous peoples to shape their own knowledge systems and protect their cultural integrity (Chilisa, 2012). In this way, evaluation becomes an exercise in self-determination rather than external judgment.

The above philosophy translates into a set of interlinked principles and practices:
First, Indigenous evaluation values **cultural contexts and ways of knowing.** Oral traditions, storytelling, lived experiences, and spirituality are recognized as valid and rigorous forms of knowledge (Kovach, 2009; Bartlett et al., 2012). Immersion in community life allows evaluators to build trust and ensures that findings reflect Indigenous epistemologies rather than external assumptions.

Second, it emphasizes **reciprocity and relational accountability**. Evaluations must contribute back to communities, avoiding extractive practices. Relational accountability means evaluators are answerable not only to funders but also to people, land, and spirit, transforming evaluation into a moral and relational practice where integrity is measured by sustained respect (Wilson, 2008).

Third, Indigenous evaluation is **community-led and participatory.** Communities co-create purposes, methods, and interpretations, ensuring alignment with local values and ownership of results. For instance, the Alaska Native Tribal Health Consortium co-designed health evaluations with Indigenous communities that prioritized Indigenous perspectives on wellness (Barnhardt & Kawagley, 2005).

Fourth, it rests on **ethical and data sovereignty frameworks**. Ethics extend beyond procedural clearance to affirm community sovereignty over knowledge and information. The OCAP® principles (Ownership, Control, Access, and Possession) safeguard Indigenous authority over how information is collected, stored, and used, countering exploitative research practices (Schnarch, 2004).

Fifth, success is understood in terms of **holistic well-being**. Models such as the Māori *Te Whare Tapa Whā* frame health as a balance across physical, mental, spiritual, and family dimensions, demonstrating that imbalance in any one of these domains undermines collective well-being (Durie, 1994).

Finally, Indigenous evaluation advances **two-eyed seeing (*Etuaptmumk*)**, which draws on the strengths of both Indigenous and Western knowledge systems. In Nova Scotia, fisheries projects combined Indigenous ecological knowledge with scientific data, yielding sustainable and culturally grounded outcomes (Bartlett et al., 2012; Reid et al., 2020).

Collectively, these philosophies and principles reframe evaluation as a community-empowering process. They dismantle colonial legacies that have long marginalized Indigenous voices (Smith, 2012), center relational accountability, and affirm self-determination (Chilisa, 2012). For M&E educators and practitioners, they offer pathways to inclusive, ethical, and contextually relevant practice. As the M&E field continues to evolve, Indigenous evaluation enriches and diversifies the discipline, positioning it as a tool not only

for accountability but also for healing, justice, and transformation.

III. Integrating Indigenous Evaluation Insights into M&E Curricula and Practices

Organizations that train and practice M&E increasingly incorporate insights from Indigenous evaluations. The literature discusses various motivations for integration, highlights Indigenous evaluation insights prioritized by organizations, and describes the diverse strategies employed by organizations involved in M&E training and practice.

III.1 Rationale for Integrating Indigenous Evaluation Insights

The integration of Indigenous evaluation insights into M&E curricula and practice has transformative potential.

First, it **addresses historical marginalization**. Colonial practices privileged Western paradigms while devaluing Indigenous knowledge systems, erasing cultural narratives and reinforcing power imbalances (Smith, 2012; Chilisa, 2012; Tuck & Yang, 2012). Integrating Indigenous perspectives challenges these "settler epistemologies" and ensures that communities become active participants in evaluations that affect them. This shift is pivotal for decolonizing M&E education and practice.

Second, it **enriches and diversifies paradigms**. Indigenous paradigms broaden the scope of evaluation by emphasizing relational accountability, cultural resilience, and holistic well-being (Wilson, 2008; Chilisa, 2012). Unlike conventional approaches that focus narrowly on measurable outputs, Indigenous methodologies emphasize the interconnections among people, the environment, and spirituality. Adopting these approaches enriches understandings of success and impact, producing more nuanced and contextually relevant evaluations.

Third, it **enhances cultural responsiveness**. Culturally responsive evaluation incorporates respect, reciprocity, and co-creation with communities (Mertens, 2012; Kovach, 2009). Embedding Indigenous perspectives fosters trust and ownership, making evaluations more inclusive and reflective of community priorities and values. This positions M&E as a tool for advancing equity and social justice.

Fourth, it **advances global development goals**. Global frameworks, such as the UN Sustainable Development Goals (SDGs), emphasize inclusivity and the principle of "leaving no one behind." Incorporating Indigenous insights ensures development initiatives remain culturally relevant and effective, validating Indigenous worldviews while addressing local challenges (Estrella & Gaventa, 1998; Mertens et al., 2013).

Fifth, it **promotes ethical and equitable practices**. Indigenous approaches redefine evaluation ethics by rejecting extractive, top-down models in favor of co-creation and data sovereignty. The OCAP® principles (Ownership, Control,

Access, and Possession) exemplify this shift, affirming community authority over knowledge (Schnarch, 2004). By prioritizing reciprocity and relational accountability (Kovach, 2009), Indigenous evaluation fosters trust and collaboration, directly addressing historical inequities.

In summary, integrating Indigenous insights equips M&E educators and practitioners to design evaluations that are more inclusive, ethical, and effective. Grounded in community values and experiences, such evaluations promote social justice, strengthen participation, and ensure that monitoring and evaluation truly respond to the needs of all stakeholders.

III.2 Embraced and Prioritized Indigenous Evaluation Insights

Organizations that have incorporated Indigenous evaluation insights into their M&E curricula or practices have intentionally selected and prioritized several key elements as follows.

III.2.1 Relational and Holistic Principles, Models, and Frameworks

A key feature of Indigenous evaluation is its holistic perspective, recognizing that all aspects of life—physical, emotional, spiritual, social, and environmental—are interconnected. This approach contrasts with the often-segmented nature of Western evaluation methods, which typically focus narrowly on outcomes or outputs. As Wilson (2008) notes, relationships are central to Indigenous worldviews, and evaluations should prioritize these relationships rather than relying solely on abstract metrics. One notable example is Durie's *Te Whare Tapa Whā* model from New Zealand. This model conceptualizes well-being as a house with four walls: physical health (*taha tinana*), mental health (*taha hinengaro*), spiritual health (*taha wairua*), and family health (*taha whānau*). By incorporating such models into M&E curricula, students can think beyond material indicators and conduct evaluations attuned to cultural and communal contexts (Durie, 1994). The *First Alaskans Institute*, which evaluates programs addressing Indigenous youth education, employs a framework that includes spiritual connections to the land and ancestors. Their reports demonstrate the importance of including relational aspects when evaluating long-term community impacts, making a compelling case for inclusion in educational curricula (Barnhardt & Kawagley, 2005).

III.2.2 Participatory Methods and Community-Led Approaches

Indigenous evaluation methods that emphasize participatory practices empower communities to be co-creators rather than just subjects of the evaluation process. This approach is rooted in the belief that communities possess the deepest knowledge of their own needs, goals, and metrics for success. Chilisa (2012) highlights that participatory techniques—such as storytelling, mapping, and community consensus—are particularly effective in Indigenous contexts. These methods not only

produce culturally rich data, but also validate and empower communities by amplifying their voices. For instance, storytelling as an evaluation tool captures lived experiences and historical narratives that are often overlooked in quantitative research. The Yarning Circle methodology, used in Australia to evaluate Aboriginal health and education programs, is a powerful example of this approach. Rooted in traditional dialogue and consensus practices, it fosters trust and creates a culturally safe space for participants (Dudgeon et al., 2014). By including these methods in the M&E curriculum, we can inspire students with the potential of culturally relevant participatory techniques. Furthermore, Estrella and Gaventa (1998) demonstrate that adaptable participatory evaluation frameworks and methods have been successfully applied in Indigenous development projects worldwide, including in regions in Latin America and Sub-Saharan Africa.

III.2.3 Ethical Guidelines and Cultural Accountability

Ethics in Indigenous evaluation extend beyond institutional protocols to include cultural values, respect for community sovereignty, and relational accountability. Kovach (2009) argues that ethical guidelines in Indigenous contexts should prioritize reciprocity, ensuring that evaluation processes and outcomes directly benefit the community. The First Nations principles of OCAP® (Ownership, Control, Access, and Possession) provide a strong ethical framework for this purpose. These principles advocate for Indigenous ownership and authority over data collection, interpretation, and dissemination. By incorporating OCAP® principles into the monitoring and evaluation (M&E) curriculum, students learn to navigate complex ethical issues related to evaluations involving Indigenous populations (Schnarch, 2004). In Canada, Indigenous communities have successfully utilized OCAP® principles to negotiate research partnerships that align with their cultural priorities. Canadian universities, such as the University of Victoria, have increasingly integrated Indigenous values and ethical frameworks into graduate-level curricula, demonstrating the feasibility of embedding such guidelines in education (Battiste, 2013; Kirkness & Barnhardt, 1991; Pidgeon, 2016). In addition, Mertens (2012) emphasizes the importance of transformative evaluation, which aligns closely with Indigenous ethics by aiming to challenge systemic inequities and ensure meaningful participation by marginalized groups.

III.2.4 Decolonizing Methodologies

Decolonizing methodologies challenge the dominance of Western paradigms by prioritizing Indigenous knowledge systems, values, and ways of knowing. Smith (2012) argues that decolonizing evaluation involves rethinking power dynamics and placing Indigenous voices at the center of evaluation design and practice. A practical approach to decolonizing monitoring and evaluation

(M&E) education is to teach frameworks like "two-eyed seeing" (*Etuaptmumk*), introduced by Mi'kmaw Elder Albert Marshall. This framework emphasizes the importance of integrating both Indigenous knowledge systems and Western science in evaluations (Bartlett et al., 2012). It encourages students to move beyond binary thinking and to develop evaluation designs that honor multiple perspectives. In the environmental sector, *two-eyed seeing* has been effectively applied in water resource management projects in Canada. Evaluations employing this approach have successfully integrated Indigenous ecological knowledge with scientific data, yielding more comprehensive and culturally respectful outcomes (Reid et al., 2020). Teaching such methodologies prepares students to tackle complex, cross-cultural evaluation challenges.

III.2.5 Indigenous Language, Context, and Indicators of Success

Language plays a central role in Indigenous evaluation, functioning as both a cultural carrier and a medium for expressing knowledge systems. Success, when articulated through Indigenous languages, is often defined in ways that diverge from Western frameworks, privileging values such as harmony, resilience, and intergenerational well-being (Kirmayer, Simpson, & Cargo, 2009). Chouinard and Cousins (2007, 2009) emphasize the importance of culturally responsive and participatory evaluation frameworks, although they fall short of specifying specific success indicators. More recent scholarship demonstrates how such indicators can be co-developed with communities. For example, Donatuto, Campbell, and Gregory (2016) document Indigenous Health Indicators created by the Swinomish Tribe, which prioritize cultural use, community connection, and self-determination. Rountree and Smith (2016) describe strength-based well-being measures identified by Indigenous youth across North America, while Bourke et al. (2022) outline the development of the Mayi Kuwayu Survey, in which Aboriginal and Torres Strait Islander peoples defined culturally grounded domains of well-being.

Comparable approaches are evident in East Africa. Ondicho (2022) shows that evaluations of Maasai women's empowerment through cultural bomas tourism rely on locally meaningful measures—such as beadwork production, women's group participation, enhanced social status, and self-confidence—rather than solely on income metrics. Taken together, these cases illustrate how Indigenous and community-driven evaluations foreground cultural continuity, cohesion, and spiritual fulfillment, thereby providing a richer and more accurate account of program success than conventional Western indicators.

Integrating the abovementioned Indigenous evaluation aspects into the M&E curriculum enhances students' understanding of culturally responsive and ethically grounded evaluations. The

principles of relationality, participatory methods, ethical accountability, and decolonizing methodologies challenge traditional paradigms, providing future evaluators with the tools they need to navigate diverse and complex contexts.

III.3 Strategies for Integration

Integrating Indigenous evaluation insights into M&E curricula requires a multidimensional approach that ensures meaningful representation and respectful inclusion. Around the world, universities, training institutions, and development agencies have adopted strategies that illustrate both the opportunities and challenges of such integration.

III.3.1 Embedding in Core Courses

A common strategy is embedding Indigenous paradigms, ethics, and methods directly into core M&E courses, ensuring that all students—not just those in electives—develop competence in culturally responsive evaluation. For example, Canadian universities integrate the First Nations OCAP® principles into ethics modules, while in Aotearoa New Zealand, Kaupapa Māori evaluation is used to challenge Western-centric frameworks (Cram, 2016; Smith, 2012). Embedding such frameworks promotes equity and prepares students to critically examine and challenge dominant paradigms (Mertens, 2012).

III.3.2 Cross-Disciplinary and Faculty Development

Integration is strengthened when M&E programs collaborate with fields such as anthropology, Indigenous studies, and public health. This broadens students' perspectives and links evaluation to wider cultural and social contexts (Smith, 2012). Faculty training is equally essential. Initiatives such as co-led workshops with Indigenous scholars or mentorship programs for non-Indigenous staff (Chilisa, 2012; Dudgeon et al., 2014) help ensure that Indigenous content is taught with cultural competence, rather than through a Western lens.

III.3.3 Community Partnerships and Experiential Learning

Partnering with Indigenous communities provides students with direct opportunities to participate in Indigenous-led evaluations. These partnerships ground evaluation practice in community priorities and ethics. For example, the Alaska Native Tribal Health Consortium collaborates with universities to offer internships shaped by Indigenous ethical principles (Barnhardt & Kawagley, 2005). Such experiential learning strengthens cultural competence and builds relationships of trust.

III.3.4 Dedicated Courses and Case-Based Learning

Some institutions offer stand-alone courses on Indigenous evaluation, providing in-depth exploration of Indigenous paradigms, methods, and case studies (Bartlett et al., 2012). Others integrate case studies into mainstream courses to illustrate how Indigenous frameworks redefine the concept of success. The Yukon First Nations Education Project, for instance, emphasized intergenerational knowledge

transfer and cultural continuity as core outcomes, demonstrating how evaluations can prioritize community-defined indicators over standardized metrics (Canadian Council on Learning, 2007).

III.3.5 Co-creation and Digital Resources

Genuine integration requires co-creation of curricula with Indigenous communities, ensuring course content reflects Indigenous priorities and ethics (Tuck & Yang, 2012). The University of Queensland's collaboration with Aboriginal leaders offers a model of authentic co-design (Dudgeon et al., 2014). Digital tools and online modules also expand access, particularly for students who are remote learners. For instance, the First Nations Information Governance Centre offers OCAP® training online, supporting ethical data governance and Indigenous data sovereignty (FNIGC, 2019).

Taken together, these strategies demonstrate that integration is not a one-time exercise, but an ongoing process that requires institutional commitment, community engagement, and pedagogical innovation. Global experiences confirm that meaningful integration deepens cultural responsiveness, strengthens ethical practice, and better prepares graduates to engage respectfully and effectively in diverse contexts. These lessons offer a valuable foundation for understanding the rationale, scope, and processes of integration that will be explored in the context of the MADE program in the next chapter.

IV. Integration of Indigenous Evaluation Insights in the MADE Program

IV.1 The MADE Program in Brief

The University of Dar es Salaam (UDSM) introduced the Master of Arts in Development Evaluation (MADE) program in 2020. This unique program aims to (i) nurture and grow the field of development evaluation in Tanzania, and (ii) build the capacity of the country's much-needed development evaluation scholars and professionals (UDSM, 2024). The MADE curriculum covers internationally recognized competencies in the development evaluation profession, including professional practice, technical competence, contextual analysis, planning and management, and interpersonal skills. The program is accredited by the Tanzania Commission for Universities (TCU) and is delivered in person at the Dar es Salaam University College of Education (DUCE).

In 2023, a comprehensive mid-review of the MADE curriculum was conducted, with a specific focus on aligning it with the 2021 African Evaluation Principles and incorporating content on Indigenous evaluation. According to these principles, evaluation practices in Africa should (i) empower Africans, (ii) be technically robust, (iii) be ethically sound, (iv) be Africa-centric, and (v) be connected with

the world (AfrEA, 2021). Consequently, the MADE program is designed to cultivate a new generation of adept professionals who embody the core characteristics of (i) Africa-centric development evaluators, who appreciate and prioritize African contexts, (ii) culturally sensitive development evaluators, who are aware of and respect diverse cultural perspectives, (iii) gender-sensitive and responsive development evaluators, who advocate for gender equity and inclusivity, (iv) ethical development evaluators, who commit to integrity and the welfare of stakeholders, and (v) ethical development evaluation managers and commissioners, who uphold ethical practices while commissioning and managing evaluations.

The revised MADE program is a comprehensive 24-month journey comprising 15 core and optional courses that cover a wide range of topics in the field of development evaluation. These courses include DE 601: Theories and Practices of Development; DE 602: Development Planning and Management; DE 603: Paradigms and Approaches to Development Evaluation; DE 604: Monitoring and Evaluation Systems and Frameworks; DE 605: Research Methodologies in Development Evaluation; DE 606: Statistical Methods for Development Evaluation; DE 607: Development Evaluation Ethics; DE 608: Planning for Development Evaluation; DE 609: Evaluating Value for Money; DE 610: Gender-sensitive Monitoring and Evaluations; DE 611: Communication and Reporting Skills for Development Evaluators; DE 612: Meta-evaluation; DE 613: ICT-enabled Monitoring and Evaluation; DE 614: Practicing Development Evaluation; and DE 699: Dissertation. Students must complete at least 10 courses, design and conduct research, and write a dissertation, ensuring a thorough understanding of the field of development evaluation (UDSM, 2024).

The focus of this paper is on the curriculum transformation process undertaken during the 2023 mid-review—specifically, how Indigenous evaluative philosophies, methods, and ethics were embedded in the MADE courses. The analysis focuses on the pedagogical and design innovations that integrate Indigenous perspectives into the program's structure and content, with the goal of providing lessons for other M&E programs worldwide.

IV.2 The Process and Strategies for Integrating Indigenous Evaluation Insights

The 2023 mid-review of the MADE curriculum was a strategically planned and collaborative process designed to embed Indigenous evaluative philosophies, methods, and ethics into the program's structure and pedagogy. The focus was on transforming the program's curriculum content, teaching approaches, and experiential learning opportunities, thereby positioning the MADE program as a model for integrating

Indigenous perspectives into graduate-level M&E education.

This curriculum transformation process was guided by the understanding that evaluation training programs must prepare graduates to work competently and respectfully in diverse cultural contexts, and that such preparation requires more than token inclusions of Indigenous content (Chilisa, 2012; Smith, 2012). It also responded to the African Evaluation Principles (AfrEA, 2021), which emphasize Africa-centric, ethically sound, and contextually relevant evaluation practice.

IV.2.1 Stakeholder Engagement and Co-creation

The curriculum review process engaged a diverse range of stakeholders, including university lecturers specializing in M&E, current M&E students, experienced M&E professionals, evaluation managers, commissioners, representatives from Indigenous communities, and practicing Indigenous evaluators. This multi-stakeholder approach aligns with recommendations by Tuck and Yang (2012) and Kovach (2009), who emphasize the importance of co-creation and reciprocal relationships in educational design processes that incorporate Indigenous knowledge. The inclusion of Indigenous voices from the outset ensured that the integration was not extractive, but grounded in community priorities and epistemologies.

IV.2.2 Key Indigenous Evaluation Insights Integrated into the Curriculum

IV.2.2.1 Indigenous Perspectives on Evaluation

Indigenous peoples possess a unique understanding of the nature, purposes, roles, principles, and standards of evaluation. These Indigenous perspectives challenge and enrich dominant Western definitions of evaluation, emphasizing relational accountability, cultural continuity, and community benefit (Wilson, 2008). Aspiring development evaluation scholars and professionals must learn about these alternative and competing viewpoints. They should also explore how to uncover Indigenous evaluation perspectives through Indigenous languages, wise sayings, and proverbs. The **Indigenous Perspectives on Evaluation** are included in **Module 1: Fundamentals of Evaluation** of the **DE 603 course**. This provides students with the opportunity to learn about, appreciate, and compare competing explanations of the meanings, purposes, tasks, and standards of evaluation. Specifically, students examine the case studies of Indigenous perspectives on evaluation found in Swahili proverbs, as presented in Mazigo (2024) and Mazigo et al. (2024). This module equips students with the necessary skills and knowledge to apply Indigenous evaluation perspectives in their future work, making them feel prepared and competent.

IV.2.2.2 Indigenous Paradigm

The Indigenous paradigm provides a comprehensive philosophical framework that significantly shapes the understanding and practice of evaluation in various contexts. Like other paradigms, it consists of foundational beliefs that explore several essential questions, all aimed at fostering a deeper engagement with reality. According to Mertens and Wilson (2019, pp. 38–46), paradigms must include *an ontological belief* that responds to the question about the nature of reality, *an axiological belief* that responds to the question about the nature of ethics, *an epistemological belief* that responds to questions about the nature of knowledge and ways of knowing, and *a methodological belief* that responds to questions about gathering credible information about the phenomenon. The Indigenous paradigm provides suitable and relevant answers to these questions, enhancing our evaluative practices. The **Indigenous paradigm**, which reflects the evaluation philosophy of Indigenous peoples, and addresses the questions of what, how, and why evaluations are conducted, is included in **Module 3: Paradigms of Development Evaluation of DE 603**. This module offers students opportunities to learn about, compare, and contrast diverse ways of thinking and shaping evaluation practices rooted in several paradigms. In addition, students learn how the ontological, epistemological, axiological, and methodological beliefs inherent in the Postpositivist, Pragmatic, Constructivist, Transformative, and Indigenous paradigms influence evaluation practices (Mertens & Wilson, 2019). This module inspires students to apply the transformative potential of the Indigenous paradigm in their future work, making them feel inspired and motivated to make a real-world impact.

IV.2.2.3 Indigenous Evaluation Approaches

Well-developed Indigenous evaluation approaches provide valuable guidance for designing and implementing evaluation assignments, making them preferable for some evaluation commissioners. Aspiring development evaluators should familiarize themselves with these approaches to succeed in markets where they are highly regarded. **Indigenous and African-rooted evaluation methods** are included in **Module 4: Development Evaluation Approaches of DE 603**. This module offers students the opportunity to learn about the unique features of various evaluation approaches, and the specific guidance they offer for designing and conducting evaluations. In addition, students will engage in constructing, critiquing, and refining evaluation approaches. Specifically, they learn the process of developing the Swahili Evaluation Approach, as detailed in the works of Mazigo (2024) and Mazigo et al. (2024).

IV.2.2.4 Indigenous Methods

Indigenous peoples utilize various methods and tools to inquire, gather evidence, and share knowledge. The notable ones are storytelling, mapping, and community consensus, which create

culturally rich data and amplify Indigenous voices (Chilisa, 2012). These methods and techniques complement Western-based research and evaluation methods. Furthermore, these methods are presented not as 'supplements' to Western techniques, but as rigorous, context-appropriate means of generating and validating knowledge.

Prospective development evaluators must learn about and practice these alternative research and evaluation methods and tools. **Indigenous methods and tools** for inquiry, evidence generation, and knowledge sharing are included in **Module 6: Indigenous Research Methodologies of DE 605**. This module enables students to learn about and practice different Indigenous methods and tools for generating and sharing evidence.

IV.2.2.5 Indigenous Evaluation Ethics

Indigenous people place a strong emphasis on ethical conduct, deeply valuing principles that recognize and uphold the rights, duties, and responsibilities of everyone involved in the evaluation process. The ethical guidelines specific to Indigenous contexts prioritize reciprocity, ensuring that evaluations are not merely procedural exercises, but instead are designed to yield tangible benefits for the community (Kovach, 2009).

In essence, Indigenous evaluation ethics have the potential to inspire a comprehensive approach to evaluations that not only aim to do no harm, but also actively confront adverse issues and generate positive impacts for individuals and communities (Van den Berg et al., 2022). This perspective stands in stark contrast to Western-based ethical frameworks, which, while effectively guiding practitioners on "doing no harm," often fall short in their ability to inspire "tackling bad" and "doing good" (Van Den Berg, 2022, p. 17).

The ethical landscape of Indigenous evaluations is complex and multifaceted, prioritizing community welfare and ethical responsibility more than simply adhering to procedural norms. Prospective development evaluators must understand these empowering Indigenous evaluation ethical frameworks. These **ethical frameworks and protocols** are included in **Module 4: Ethical Decision-making Frameworks in Evaluation Contexts of DE 607**. This module provides students with the opportunity to learn about and appreciate the practical guidance offered by various ethical frameworks on ethical behavior and evaluation.

IV.2.2.6 Exemplary Cases of Indigenous Evaluation Practices

Indigenous peoples regularly evaluate their planned and implemented socio-economic and political activities and events. Empirical research has documented notable cases of Indigenous evaluations in wedding ceremonies, burial rituals, farming, and fishing projects. These examples provide valuable lessons on how Indigenous communities address the challenges of

exclusion and marginalization during evaluation processes. They also highlight strategies for incorporating gender and human rights considerations into evaluations, and promoting democratic, inclusive, and gender-sensitive evaluation practices.

Prospective development evaluators should learn from and appreciate the evaluative thinking, methods, tools, and protocols of Indigenous peoples to engage with them effectively in future evaluation assignments. **Exemplary Cases of Indigenous Evaluation Practices** have been identified and included in the **sections of DE 603, DE 605, DE 610, and DE 614**. These materials allow students to learn about, appreciate, and critique Indigenous people's evaluative thinking, standards, methods, tools, protocols, and strategies for incorporating considerations of gender and human rights into evaluations.

IV.2.2.7 Experiential Learning and Participation in Indigenous Evaluations

Aspiring development evaluators must develop cultural competencies to effectively conduct evaluations within Indigenous communities. This can be achieved through experiential learning opportunities, allowing students to engage in community-led or Indigenous-led evaluations. By partnering with Indigenous communities, students can learn directly from community members and experts. Such partnerships offer valuable opportunities to understand and apply evaluation criteria, standards, protocols, designs, methods, and tools specific to Indigenous peoples. **Experiential Learning and Participation in Indigenous Evaluations** are integral components of **DE 614** and **DE 699**, unique courses that offer students exceptional opportunities to learn about the philosophy, principles, and practices of evaluation in collaboration with Indigenous communities. In addition, these courses provide students with opportunities to research Indigenous evaluation thoughts and practices, generating new insights and exemplary cases of Indigenous evaluation practices.

IV.2.3 Integration Strategies

The integration strategy for the MADE program intentionally embeds Indigenous content throughout the curriculum, rather than isolating it in elective modules, following best practice guidance from Cram (2016), Battiste (2013), and Pidgeon (2016). The Indigenous evaluation insights incorporated into the MADE curriculum are **woven into core courses**, ensuring they are not treated as peripheral knowledge. This approach reflects a commitment to embedding significant and valued elements of Indigenous evaluation across coursework, fostering structural rather than tokenistic inclusion.

Additional strategies include (i) inviting Indigenous evaluators as guest lecturers to provide lived perspectives; (ii) encouraging student research on Indigenous evaluation practices to expand the body of documented approaches, and (iii) establishing formal agreements with Indigenous

organizations to support sustained experiential learning.

The program invites Indigenous evaluators as guest speakers, offering students the chance to gain first-hand insights and perspectives from experienced practitioners in the field. This fosters a dynamic learning environment where theory and practice intersect. To supplement these experiences, the curriculum also encourages research into Indigenous evaluation thoughts and practices, fostering an atmosphere of ongoing learning that generates new insights and valuable lessons for all participants. Lastly, the program has established formal agreements with Indigenous communities and Indigenous organizations to provide experiential learning opportunities. This collaborative approach allows students to engage directly with Indigenous-led evaluation processes, enriching their understanding and appreciation of these practices.

IV.2.4 Outcomes of the Mid-Review Process

The mid-review resulted in a curriculum that is richer in cultural responsiveness, grounded in ethical principles that privilege community welfare, and capable of preparing graduates to navigate cross-cultural evaluation contexts. These changes illustrate how graduate M&E programs can operationalize the African Evaluation Principles (AfrEA, 2021) and global calls for decolonizing evaluation (Chilisa, 2012; Smith, 2012). While the MADE program is still in the early stages of implementing these changes, the integration process offers replicable lessons for other institutions seeking to bridge Indigenous and Western paradigms in evaluation education.

V. Discussion

The integration of Indigenous evaluation insights into the Master of Arts in Development Evaluation (MADE) program offers critical insights into the broader project of decolonizing higher education curricula, particularly within monitoring and evaluation (M&E) training. By embedding Indigenous philosophies, ethics, and methodologies throughout core courses, the MADE program moves beyond tokenistic inclusion toward a transformative educational model. This section examines the implications of the program's integration process in three key areas: bridging Indigenous and Western paradigms, addressing ethical and cultural considerations, and drawing comparative perspectives from global practices.

V.1 Bridging Indigenous and Western Paradigms

A central achievement of the MADE program lies in its intentional bridging of Indigenous and Western paradigms in evaluation. Western evaluation traditions—rooted in positivist and postpositivist epistemologies—often prioritize standardized metrics, linear logic models, and quantifiable outputs (Mertens & Wilson, 2019). While these approaches have made significant contributions to evaluation theory and

practice, they can marginalize community-based knowledge systems that prioritize relationality, spirituality, and holistic well-being (Durie, 1994; Wilson, 2008).

The MADE program's approach is consistent with the philosophy of *two-eyed seeing* (*Etuaptmumk*) articulated by Bartlett et al. (2012), which emphasizes drawing on the strengths of both Indigenous and Western knowledge systems. By embedding the Indigenous paradigm alongside postpositivist, constructivist, pragmatic, and transformative paradigms in DE 603, students gain a comparative, critical perspective on how ontological, epistemological, axiological, and methodological orientations shape evaluation practices. This exposure encourages students to recognize that methodological pluralism can enhance the validity and cultural relevance of evaluation.

Such paradigm-bridging is not merely academic—it prepares graduates to operate in complex, multicultural evaluation contexts, where hybrid approaches may yield the most socially legitimate and technically robust outcomes (Cram, 2016). For example, blending community-led storytelling and participatory mapping with conventional statistical analysis can meet both donor accountability requirements and community-defined measures of success. This approach aligns with the African Evaluation Principles (AfrEA, 2021), which call for Africa-centric, contextually grounded evaluations that remain connected to global standards.

V.2 Ethical and Cultural Implications

The integration process also highlights the ethical imperative of culturally grounded evaluation. Traditional Western evaluation ethics frameworks often emphasize "do no harm" principles, focusing on minimizing risk (Van den Berg, 2022). Indigenous evaluation ethics go further, centering on reciprocity, relational accountability, and community benefit (Chilisa, 2012; Kovach, 2009). These principles reframe evaluation not as a neutral or detached exercise, but as a relational process in which evaluators are accountable to the communities with whom they engage.

The MADE program operationalizes these Indigenous ethical frameworks into DE 607, ensuring that graduates are trained to respect data sovereignty and community autonomy (Schnarch, 2004). Embedding such frameworks directly into required ethics courses institutionalizes cultural accountability as a non-negotiable element of professional practice.

Moreover, integrating Indigenous ethics into experiential learning—such as the partnerships with Indigenous communities in DE 614—provides students with lived examples of navigating ethical complexities in real-world settings. This hands-on exposure echoes Kolb's (1984) experiential learning model, which emphasizes the transformation of abstract knowledge into practical

competence through concrete engagement.

By privileging ethical frameworks that actively 'tackle bad' and 'do good' (Van den Berg et al., 2022), the program equips graduates to challenge extractive, top-down evaluation practices. It positions evaluation as a tool for equity and empowerment, rather than control, contributing to the profession's role in advancing social justice (Mertens, 2012).

V.3 Comparisons with Global Practices

The MADE program's integration strategy stands out in the global landscape of Indigenous evaluation education. In many contexts, Indigenous content is offered as a discrete course or elective, such as the Te Whare Tapa Whā model (Durie, 1994), which has been widely used in New Zealand evaluations to embed Māori values of balance and well-being or elective courses on Indigenous perspectives within Canadian universities (Battiste, 2013; Pidgeon, 2016). While these are important advances, they may limit exposure to students who choose those electives, rather than ensuring all graduates develop competence in Indigenous evaluation approaches.

In contrast, the MADE program ensures universal exposure by embedding Indigenous perspectives across multiple required modules, including theory (DE 603), methods (DE 605), ethics (DE 607), gender-sensitive evaluation (DE 610), and practice-oriented courses (DE 614, DE 699). This approach aligns with the systemic integration advocated by Tuck and Yang (2012), who argue that decolonizing education necessitates fundamental structural shifts, rather than superficial curriculum adjustments.

Internationally, examples of similar systemic embedding are rare. Some parallels can be drawn with the University of Waikato's cross-disciplinary integration of Māori perspectives in New Zealand, or with collaborative evaluation models in Canada that pair non-Indigenous faculty with Indigenous mentors (Dudgeon et al., 2014). However, the MADE program's Africa-centric orientation and alignment with the African Evaluation Principles provide a unique regional lens that could serve as a model for other African universities.

This systemic integration also offers a platform for innovation in applied evaluation practice. For instance, experiential learning placements with Indigenous communities mirror the Alaska Native Tribal Health Consortium's approach (Barnhardt & Kawagley, 2005), but with an Africa-specific contextualization that prioritizes Swahili evaluation approaches (Mazigo et al., 2024). By ensuring that students graduate with cross-cultural competencies and a repertoire of Indigenous and Western methods, the program enhances both employability and the potential for evaluations to be more contextually relevant and ethically sound.

V.4 Implications for Future Research and Practice

The integration of Indigenous evaluation insights into the MADE program offers fertile ground for advancing both scholarship and professional practice in M&E. Future research could follow graduates into their professional contexts to examine how they apply Indigenous paradigms, methods, and ethics in real-world evaluations (Mertens, 2012; Wilson, 2008). Such longitudinal studies would generate evidence of the program's effectiveness in producing culturally responsive practitioners, and reveal how blended Indigenous–Western approaches, such as two-eyed seeing (Bartlett et al., 2012), perform in diverse settings. Comparative analyses with other African and global M&E programs could also clarify which strategies are transferable and which require careful adaptation, contributing to a broader repertoire of regionally grounded Indigenous evaluation frameworks (Battiste, 2013; Smith, 2012; Cram, 2016; Pidgeon, 2016).

Strengthening and documenting community–university partnerships will be essential for sustaining this integration. The MADE program's experiential learning model provides students with opportunities to work alongside Indigenous evaluators, aligning with best practices in collaborative evaluation design (Kovach, 2009; Tuck & Yang, 2012). Further research could shed light on how co-creation processes, governance structures, and shared benefits can be sustained over time, ensuring that partnerships remain equitable and community-driven. As digital and blended learning formats become more prevalent, there is also potential to explore how virtual platforms might deliver Indigenous evaluation training without compromising cultural integrity, relational accountability, or the co-production of knowledge (Hartmann & Gone, 2016).

Beyond the classroom, the MADE experience could inform policy and accreditation standards by embedding Africa-centric and Indigenous content into the core competencies for professional M&E training (AfrEA, 2021). Disseminating the curriculum design, teaching resources, and case studies through open-access platforms and professional networks would not only amplify the program's reach, but also foster global dialogue on decolonizing evaluation education (Chilisa, 2012; Smith, 2012). In this way, MADE serves as both a living laboratory for culturally grounded evaluation, and a catalyst for systemic change, offering a model that is ethically robust, inclusive, and contextually relevant.

The MADE experience demonstrates that embedding Indigenous evaluation into graduate curricula is both feasible and scalable when approached strategically. Its replicability rests on three interlinked design features: integration across core courses, rather than confinement to electives, ensuring that all students—not only those with pre-existing interest—

develop cultural competence; co-creation with Indigenous knowledge holders and professional bodies, such as AfrEA, which safeguards authenticity and community trust; and alignment with internationally recognized ethical frameworks, including the African Evaluation Principles and the concept of *two-eyed seeing*, which provides a bridge between local grounding and global applicability. This combination enables the model to be adapted to various cultural contexts, while maintaining its integrity, providing a practical blueprint for universities and training institutions committed to decolonizing evaluation education.

VI. Conclusion

The integration of Indigenous evaluation insights into the Master of Arts in Development Evaluation (MADE) program represents a deliberate and transformative reorientation of how evaluation is taught, learned, and practiced. Rather than treating Indigenous content as an optional supplement, the program embeds Indigenous paradigms, ethics, and methodologies across its core curriculum. In doing so, it challenges entrenched Western-dominated assumptions in M&E education and advances a model that is contextually grounded, ethically robust, and globally relevant.

This transformation aligns directly with the African Evaluation Principles (AfrEA, 2021) and responds to global calls to decolonize evaluation (Chilisa, 2012; Smith, 2012). It equips graduates to design and implement evaluations that respect community-defined measures of success, uphold data sovereignty, and foster reciprocal, trust-based relationships. Through experiential learning and sustained partnerships with Indigenous communities, these competencies move beyond theory, preparing graduates to practice evaluation in culturally respectful and methodologically sound ways.

The MADE model also provides a replicable and adaptable framework for other institutions seeking to integrate culturally responsive approaches into professional M&E training. Its innovation lies not only in the breadth of integration across courses but also in the way it positions evaluation as an instrument for equity, empowerment, and social justice. As such, it has the potential to influence accreditation standards, inspire cross-institutional collaborations, and contribute to a global reimagining of evaluation practice.

By setting this precedent, MADE affirms that evaluation education can be academically rigorous, ethically grounded, and transformative in addressing historical inequities while helping to shape a more inclusive and just professional field.

Acknowledgements

I wish to formally acknowledge the valuable comments and insights provided by participants in the MADE review process, and the panelists of Session 10, "Integrating Indigenous Evaluations into the M&E Curriculum and Culturally Responsive Evaluation Practices," at the Academic Symposium on Monitoring and Evaluation Education, held in Colombo, Sri Lanka, from 2 to 4 December 2024. Furthermore, I am grateful for the financial support received from EvalIndigenous, which facilitated my attendance at the symposium, and enabled me to present the initial draft of this manuscript.

Declaration of Conflicting Interests

The author declares, with utmost sincerity, that he has no personal or financial interests that may have inappropriately influenced him in writing this article.

References

African Evaluation Association (AfrEA). (2021). *The African evaluation principles 2021 version.* https://alnap.hacdn.io/media/documents/The-African-Evaluation-Principles.pdf

Barnhardt, R., & Kawagley, A. O. (2005). Indigenous knowledge systems and Alaska Native ways of knowing. *Anthropology & Education Quarterly, 36*(1), 8–23.

Bartlett, C., Marshall, M., & Marshall, A. (2012). Two-eyed seeing and other lessons learned within a co-learning journey of bringing together Indigenous and mainstream knowledges and ways of knowing. *Journal of Environmental Studies and Sciences, 2*(4), 331–340.

Battiste, M. (2013). *Decolonizing education: Nourishing the learning spirit.* Purich Publishing.

Bourke, S., Lovett, R., Maher, B., Fane, J., Guthrie, J., Cox, L., & Walter, M. (2022). The development of the Mayi Kuwayu well-being survey: Measuring the wellbeing of Aboriginal and Torres Strait Islander peoples. *International Journal for Equity in Health, 21*(1), 36. https://doi.org/10.1186/s12939-022-01710-8

Canadian Council on Learning. (2007). *Redefining how success is measured in First Nations, Inuit and Métis learning* (Report on Learning in Canada No. 07). Ottawa, ON: Canadian Council on Learning. https://education.afn.ca/afntoolkit/wp-content/uploads/2021/04/5._2007_redefining_how_success_is_measured_en.pdf

Chilisa, B. (2012). *Indigenous research methodologies.* SAGE Publications.

Chouinard, J. A., & Cousins, J. B. (2007). *Culturally competent evaluation for Aboriginal communities: A review of the empirical literature. Journal of MultiDisciplinary Evaluation, 4*(8), 40–57. https://doi.org/10.56645/jmde.v4i8.30

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation, 30*(4), 457–494. https://doi.org/10.1177/1098214009349865

Cram, F. (2016). Lessons on decolonizing evaluation from Kaupapa Māori evaluation. *Canadian Journal of Program Evaluation, 30*(3), 296–312. https://doi.org/10.3138/cjpe.30.3.04

Donatuto, J., Campbell, L., & Gregory, R. (2016). Developing responsive indicators of Indigenous community health. *International Journal of Environmental Research and Public Health, 13*(9), 899. https://doi.org/10.3390/ijerph13090899

Dudgeon, P., Milroy, H., & Walker, R. (2014). *Working together: Aboriginal and Torres Strait Islander mental health and wellbeing principles and practice.* Australian Government Department of the Prime Minister and Cabinet.

Durie, M. (1994). *Whaiora: Māori health development* (2nd ed.). Oxford University Press.

Estrella, M., & Gaventa, J. (1998). *Who counts reality? Participatory monitoring and evaluation: A literature review* (IDS Working Paper 70). Institute of Development Studies, University of Sussex, Falmer, UK.

First Nations Information Governance Centre (FNIGC) (2019). *The First Nations principles of OCAP®*. https://fnigc.ca/ocap-training/

Kirkness, V. J., & Barnhardt, R. (1991). First Nations and higher education: The four R's—Respect, relevance, reciprocity, responsibility. *Journal of American Indian Education, 30*(3), 1–15. https://eric.ed.gov/?id=EJ438242

Kirmayer, L. J., Simpson, C., & Cargo, M. (2003). Healing traditions: Culture, community and mental health promotion with Canadian Aboriginal peoples. *Australasian Psychiatry, 11*(Suppl), S15–S23. https://doi.org/10.1046/j.1038-5282.2003.02010.x

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Prentice Hall.

Kovach, M. (2009). *Indigenous methodologies: Characteristics, conversations, and contexts.* University of Toronto Press.

Mazigo, A. F. (2024). The Swahili evaluation approach: Content and guidance for doing development evaluation. *African Evaluation Journal, 12*(2), a739. https://doi.org/10.4102/aej.v12i2.739

Mazigo, A. F., Mwaijande, F., Nguluki, I. M., & Mkombozi, M. (2024). Swahili wisdom for shaping development evaluation practices. *African Evaluation Journal, 12*(2), a738. https://doi.org/10.4102/aej.v12i2.738

Mertens, D. M. (2012). Transformative mixed methods: Addressing inequities. *American Behavioral Scientist, 56*(6), 802–813.

Mertens, D. M., & Wilson, A. T. (2019). *Program evaluation theory and practice: A comprehensive guide*. Guilford Publications.

Mertens, D. M., Cram, F., & Chilisa, B. (Eds.). (2013). *Indigenous pathways into social research: Voices of a new generation*. Left Coast Press.

Ondicho, T. G. (2022). Empowering Kenyan Maasai women through cultural bomas tourism. *Pathways to African Feminism and Development: Women's Economic Empowerment, 7*(1), 1–15. African Women Studies Centre, University of Nairobi. https://uonjournals.uonbi.ac.ke/ojs/index.php/aws/article/view/1618

Pidgeon, M. (2016). More than a checklist: Meaningful Indigenous inclusion in higher education. *Social Inclusion, 4*(1), 77–91. https://doi.org/10.17645/si.v4i1.436

Reid, A. J., Eckert, L. E., Lane, J. F., Young, N., Hinch, S. G., Darimont, C. T., & Marshall, A. (2020). Two-eyed seeing: An Indigenous framework to transform fisheries research and management. *Fish and Fisheries, 21*(4), 730–748.

Rountree, J., & Smith, A. (2016). Strength-based well-being indicators for Indigenous children and families: A literature review of Indigenous communities' identified well-being indicators. *American Indian and Alaska Native Mental Health Research: Journal of the National Center, 23*(3), 206–220. https://doi.org/10.5820/aian.2303.2016.206

Schnarch, B. (2004). Ownership, Control, Access, and Possession (OCAP®) or self-determination applied to research: A critical analysis of contemporary First Nations research and some options for First Nations communities. *Journal of Aboriginal Health, 1*(1), 80–95. https://doi.org/10.3138/ijih.v1i1.28934

Smith, L. T. (2012). *Decolonizing methodologies: Research and Indigenous peoples* (2nd ed.). Zed Books.

Tuck, E., & Yang, K. W. (2012). Decolonization is not a metaphor. *Decolonization: Indigeneity, Education & Society, 1*(1), 1–40.

University of Dar es Salaam (UDSM). (2024). *Reviewed curriculum for the Master of Arts in Development Evaluation (MADE) program.*

Van den Berg, R. D. (2022). Ethical guidance: Enabling evaluation's role in tackling bad and doing good. In R. D. van den Berg, P. Hawkins, & N. Stame (Eds.), *Ethics for evaluation: Beyond 'doing no harm' to 'tackling bad' and 'doing good'* (pp. 17–44). Routledge.

Van den Berg, R. D., Hawkins, P., & Stame, N. (Eds.). (2022). *Ethics for evaluation: Beyond 'doing no harm' to 'tackling bad' and 'doing good.'* Routledge.

Wilson, S. (2008). *Research is ceremony: Indigenous research methods*. Fernwood Publishing.