The Journal of Teaching and Learning for Graduate Employability

ISSN: 1838-3815 (online) Journal Homepage: https://ojs.deakin.edu.au/index.php/jtlge/

Before we automate WIL evaluation, we need to ask what we're actually measuring

David Fulcher [1]

Corresponding author: David Fulcher (dfulcher@uow.edu.au)
[1] University of Wollongong, New South Wales, Australia

Abstract

Community-engaged Work-Integrated Learning (WIL) is increasingly positioned as a mechanism for advancing graduate employability and social responsibility, yet the frameworks used to evaluate it remain anchored to countable proxies: employment rates, satisfaction scores, and placement completions. This provocation argues that these metrics are structurally misaligned with what WIL practitioners actually know about placement quality: knowledge that is relational, contextually rich, and routinely invisible in institutional reporting. As Generative AI enters WIL evaluation, this misalignment becomes urgent. Automating existing frameworks will not resolve their limitations, and risks further marginalising the evidence that most reliably signals whether community-engaged WIL is working. Drawing on research into practitioner agency in learning analytics, this provocation calls for evaluation frameworks that legitimise practitioner judgement, centre community partner perspectives, and deploy GenAI in service of relational knowledge rather than as a substitute for it. Effective WIL evaluation is not primarily a measurement problem, it is a question of whose evidence counts.

Keywords
work-integrated learning, learning analytics, evaluation frameworks, practitioner knowledge, generative AI, community engagement, graduate employability

A prior question

Universities are under genuine and growing pressure to scale Work-Integrated Learning. Community-engaged WIL is increasingly positioned as a mechanism for addressing graduate employability, sustainable development, and social responsibility simultaneously - an ambitious brief (Wright et al., 2022; Zegwaard et al., 2020). In response, institutions are investing in the infrastructure to deliver WIL at scale and, inevitably, to demonstrate that it works (Cook, 2021). Generative AI is now being drafted into this effort, promising to streamline placement reporting, summarise student reflections, and surface patterns across large cohorts that no coordinator could manually track (Young et al., 2025). The efficiency gains are real. But before institutions accelerate down this path, a prior question deserves serious attention: are we measuring the right things in the first place?

What the metrics actually capture

Current WIL evaluation frameworks tend to reach for what is countable. Graduate employment rates, typically the percentage of students securing work within six months of placement, are widely cited as headline indicators of WIL success (Cook, 2021). Student satisfaction scores, collected through post-placement surveys, provide a second layer of evidence (Young et al., 2024). Placement completion numbers, such as hours logged and positions filled, round out the picture.

These metrics are not without value. But experienced WIL practitioners know they are incomplete proxies at best, and actively misleading at worst. Employment rates reflect graduate labour market conditions and student socioeconomic background as much as they reflect the quality of any placement (Jackson & Bridgstock, 2021). Satisfaction scores capture surface impressions, not depth of learning. Throughput numbers say nothing about whether the work was meaningful or merely busy-work. Table 1 makes this gap visible.

Table 1. Common countable WIL metrics and practitioner-valued insights that are often unreported.

Common WIL Metric (Countable)	**Practitioner-Valued Insight (Often Unreported)**
Graduate employment rate	Quality of student transformation and growth during placement, observed by supervisors but absent from formal reports
Student satisfaction score	Depth of the student–mentor relationship and genuineness of community partnership reciprocity
Number of placements completed	Actual community impact, and whether the work was meaningful or merely make-work

The gap here is not a technical problem awaiting a better dashboard. It is structural. What gets measured reflects what institutional accountability frameworks were designed to capture — not what community-engaged WIL was designed to achieve (Peach et al., 2012; Rook & Sloan, 2021).

Practitioners already hold the evidence

Here is what is rarely acknowledged in the evaluation literature: WIL practitioners are not working in an evidence vacuum. They hold rich, contextual knowledge about the quality of placements. This is knowledge that is real, professionally grounded, and routinely invisible in institutional reporting (Young et al., 2025).

An experienced WIL coordinator knows when a placement has been genuinely transformative. They know the student who arrived uncertain and left with a confidence that no satisfaction score could convey, who delivered something the community partner now relies upon. They know, equally, when a placement went through the motions: when the partnership was token rather than reciprocal, when the student completed the hours but nothing of consequence happened for anyone. A field supervisor's handwritten note, a community partner's comment in a debrief conversation, a coordinator's professional intuition accumulated across dozens of placements - this is evidence. It simply isn't legible to the frameworks designed to receive it.

This mirrors a finding from research into how university teachers engage with learning analytics more broadly: practitioners exercise significant agency in interpreting and adapting data tools to their contexts, drawing on professional judgement that the system itself cannot replicate (Fulcher, 2025). WIL coordinators do the same. The problem is not that they lack evidence, it is that the evaluation infrastructure is not built to hold the evidence they already have.

Why GenAI makes this urgent

Generative AI will not resolve this misalignment. In all likelihood, it will deepen it.

The most immediate applications of GenAI in WIL evaluation are predictable: automating placement reports, generating summaries of student reflective journals, flagging completion anomalies at scale (Young et al., 2024). These are genuinely useful capabilities, but they operate on the data that already exists. If the underlying evaluation framework privileges employment rates and satisfaction scores, GenAI will process employment rates and satisfaction scores more efficiently. The structural gap remains; it simply becomes harder to see.

There is a subtler risk. GenAI-generated evidence carries an institutional air of rigour that a field supervisor's handwritten note does not. Under accountability pressure, universities will gravitate toward outputs that look like data. The relational, contextual, practitioner-held knowledge that actually signals whether community-engaged WIL is working gets further marginalised. This is not because anyone decided it mattered less, but because the infrastructure was never built to hold it, and now a more sophisticated infrastructure has arrived that doesn't hold it either.

The question is not whether GenAI has a role in WIL evaluation. It does. But there are more promising applications than automating existing reporting: helping practitioners articulate qualitative observations systematically; supporting reflective documentation that makes tacit professional judgement legible; enabling community partners to contribute their own evidence of impact in forms the institution can actually receive (Australian Collaborative Education Network, 2024). These applications put GenAI in service of the relational, rather than against it. They require, however, that institutions first decide what they are trying to measure, and why.

A constructive provocation

Three things would help.

First, evaluation frameworks should explicitly legitimise practitioner judgement as a recognised form of evidence, not as anecdote to be noted and discarded but as professional knowledge with standing in institutional reporting. This is not a call for subjectivity over rigour; it is a call for a more sophisticated understanding of what rigour looks like in relational contexts (Cook, 2021).

Second, before designing any evaluation system, AI-augmented or otherwise, institutions should ask community partners and WIL practitioners what evidence of a good placement looks like to them. The literature on effective WIL partnership consistently emphasises reciprocity and relationship quality as central to outcomes (Peach et al., 2012; Rook & Sloan, 2021). Evaluation frameworks should reflect that centrality, not work around it.

Third, the arrival of GenAI in higher education is an opportunity to ask hard questions about evaluation, not to defer them. The urgency to scale is real. But scaling the wrong framework faster is not progress.

Key questions for future evaluation

To reorient WIL evaluation toward what practitioners and partners actually value, institutions might begin by asking:

- Did the placement change how the student understands themselves as a professional, not merely whether they secured employment six months later?
- Would the community partner describe the partnership as genuinely reciprocal, not merely whether the student completed their hours?

- What would a field supervisor say to a colleague unprompted, not merely what they recorded on a form?

These questions do not replace systematic data collection. They redirect it. They make space for the relational, contextual evidence that WIL practitioners already hold, and they provide a basis for evaluating GenAI tools against a standard that matters: not whether they automate what we already measure, but whether they help us surface what we have consistently failed to capture.

Unless we redefine what counts in WIL, automating evaluation with GenAI will only help us measure the wrong things faster.

Conflict of interest

The author has no conflicts of interest to declare.

Funding

This work received no external funding.

Statement on the use of Artificial Intelligence

AI tools were used for literature identification, argument development, and drafting support only. All intellectual contribution, critical synthesis, verification of sources, and final authorship responsibility rest with the author. Microsoft Copilot Researcher (OpenAI GPT, accessed March 2026) was used to assist with identification of potentially relevant literature; all sources were independently located, read, and verified by the author. Microsoft Copilot (GPT-4, accessed March 2026) and Claude (Anthropic, claude-sonnet-4-6, accessed March 2026) were used to support argument development, structural drafting, and refinement of clarity. All AI-generated suggestions were critically reviewed and revised by the author. No AI tool is listed as an author.

Contributions against CReDIT

David Fulcher: Conceptualisation; Methodology; Writing – original draft; Writing – review and editing.

References

Australian Collaborative Education Network. (2023). *ACEN response to Accord Discussion Paper*. https://www.education.gov.au/system/files/documents/submission-file/2023-04/AUA_tranche1_Australian%20Collaborative%20Education%20Network%20%28ACEN%29.pdf

Cook, E. J. (2021). Evaluation of work-integrated learning: A realist synthesis and toolkit to enhance university evaluative practices. *International Journal of Work-Integrated Learning*, *22*(2), 213–239. https://ro.ecu.edu.au/ecuworkspost2013/10403/

Fulcher, D. (2025). *Investigating how Australian university teachers' use learning analytics as part of their teaching practice* [Unpublished doctoral dissertation] University of Wollongong. https://doi.org/10.71747/uow-r3gk326m.29906354.v1

Jackson, D., & Bridgstock, R. (2021). What actually works to enhance graduate employability? The relative value of curricular, co-curricular, and extra-curricular learning and paid work. *Higher Education*, *81*(4), 723–739. https://doi.org/10.1007/s10734-020-00570-x

Peach, D., Larkin, I., & Ruinard, E. (2012). High-risk, high-stake relationships: building effective industry-university partnerships for Work Integrated Learning (WIL). Proceedings of the Australian Collaborative Education Network (ACEN), Australia, 230-236. https://acen.edu.au/wp-content/uploads/2015/09/ACEN-2012-National-Conference-Proceedings.pdf?x91341#page=230

Rook, L., & Sloan, T. (2021). Competing stakeholder understandings of graduate attributes and employability in work-integrated learning. *International Journal of Work-Integrated Learning*, *22*(1), 41–56. https://files.eric.ed.gov/fulltext/EJ1286230.pdf

Wright, C., Ritter, L. J., & Wisse Gonzales, C. (2022). Cultivating a collaborative culture for ensuring sustainable development goals in higher education: An integrative case study. *Sustainability*, *14*(3), 1273. https://doi.org/10.3390/su14031273

Young, K., McKenzie, S., & Harvey, M. (2024). WIL evaluation: It is both what you know, and who 'knows' what, that matters. *WIL in Practice*, *2*(1). http://wilinpractice.nafea.org.au/index.php/WILIP/article/view/23

Young, K., Bradbury, O., & McKenzie, S. (2025). PRISE: An explorative study of a workplace-based work-integrated learning data schema prototype. *International Journal of Work-Integrated Learning*, *26*(4), 801–815. https://files.eric.ed.gov/fulltext/EJ1492131.pdf

Zegwaard, K. E., Pretti, T. J., & Rowe, A. D. (2020). Responding to an international crisis: The adaptability of the practice of work-integrated learning. *International Journal of Work-Integrated Learning*, *21*(4), 317-330. https://researchcommons.waikato.ac.nz/server/api/core/bitstreams/ff323c7f-8d85-44dd-8b49-720b5fa94523/content