



Delft University of Technology

## Developing and Piloting an Approach to Evaluate Educational Innovation at Course Level

Engelbrecht, Erna; Rooij, Remon; Specht, Marcus; Strobel, Johannes

### DOI

[10.1109/EDUCON60312.2024.10578656](https://doi.org/10.1109/EDUCON60312.2024.10578656)

### Publication date

2024

### Document Version

Final published version

### Published in

2024 IEEE Global Engineering Education Conference (EDUCON)

### Citation (APA)

Engelbrecht, E., Rooij, R., Specht, M., & Strobel, J. (2024). Developing and Piloting an Approach to Evaluate Educational Innovation at Course Level. In *2024 IEEE Global Engineering Education Conference (EDUCON)* (IEEE Global Engineering Education Conference, EDUCON). IEEE.  
<https://doi.org/10.1109/EDUCON60312.2024.10578656>

### Important note

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

***Green Open Access added to TU Delft Institutional Repository***

***'You share, we take care!' - Taverne project***

**<https://www.openaccess.nl/en/you-share-we-take-care>**

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

# Developing and Piloting an Approach to Evaluate Educational Innovation at Course Level

Erna Engelbrecht  
Faculty of Architecture and  
the Built Environment  
Delft University of  
Technology  
Delft, The Netherlands  
0000-0001-8731-347X

Remon Rooij  
Faculty of Architecture and  
the Built Environment  
Delft University of  
Technology  
Delft, The Netherlands  
0000-0002-2784-1892

Marcus Specht  
Faculty of Electrical  
Engineering, Mathematics  
and Computer Science  
Delft University of  
Technology  
Delft, The Netherlands  
0000-0002-6086-8480

Johannes Strobel  
College of Education,  
STEM Education  
The University of Texas at  
El Paso  
El Paso, USA  
0000-0002-2124-1116

**Abstract**— Evaluation of educational innovation at the course level is dominated by frameworks that are linear, built on simple models, lack grounding in theory, and are generally not flexible enough to provide actionable recommendations for improvements. Moreover, the guidelines provided by existing literature are difficult to transfer to other contexts as most of these are aimed at specific innovations only. Therefore, there is a need for a new approach that is flexible enough to enable adaptation to different contexts, and that draws on the strengths of existing evaluation traditions. This paper forms a starting point for a larger research initiative for developing a comprehensive innovation evaluation framework. The context of the paper is a project based within the four universities of technology in the Netherlands (4TU) dedicated to strengthening collaboration in engineering education and research needed to take on global societal problems. The potential reach of our framework has implications for the wider field. This paper describes the results of an exploratory literature review to draw on the strengths of various evaluation theories to develop a new approach to evaluating educational innovation in courses. Value, Methods, and Use are the main components identified for this approach. Secondly, we developed a workshop to pilot this approach. Lastly, the workshop was presented to educational researchers, engineering educators, and educational advisors. We share lessons learned from the workshop and conclude with descriptions of future research to refine the framework to prepare its application in real innovation initiatives.

**Keywords**— *educational innovation, innovation evaluation, evaluation theory*

## I. INTRODUCTION

With technological and industrial advances taking place at an increasingly fast pace, our educators are working hard to keep their courses relevant and up to date. This is evident from the large body of research through which such initiatives are evaluated (see, for example, [1], [2], [3], [4]). However, existing educational innovation models are difficult to find and often fall short of expectations [5], [6]. What makes it more difficult is that the research is still fragmented and lacks consensus and theoretical underpinning, which limits transfer to other contexts [5], [7], [8]. Therefore, we need to develop a new approach that is flexible enough to enable adaptation to different contexts, and that draws on the strengths of existing evaluation traditions.

To take on this challenge, we aim to address the following research question: How can literature-driven evaluation methods contribute to the development of practical educational innovation evaluation approaches? We first describe how we conducted the literature review to develop

this initial evaluation approach. Then, we explain how we piloted the application of the evaluation approach. Lastly, we discuss our results and conclusions for moving forward toward refining our framework for evaluating educational innovation.

## II. METHODOLOGY

### A. Reviewing the literature

In conducting the exploratory literature review, a simple search term was used, ‘education innovation evaluation framework,’ searching IEEE, Web of Science, Scopus, and Google Scholar. Based on relevant search results, additional papers were identified through their reference lists to expand our view of the available literature. We then narrowed down publications for inclusion based on how well they relate to our broader research aim of developing an innovation evaluation framework, outlined in our working definition of innovation:

*Determining the (potential) impact of transformation that resulted from an intervention that was introduced to a new context for improved decision-making in educational practice.*

This working definition is broadly constructed and can include beliefs, practices and resources [9]. The definition will be refined as the main research initiative progresses.

### B. Development of the workshop

Based on the literature, a few key points were extracted. These included the importance of a thorough understanding of the Innovation itself, the Purpose (reason or goal) of the innovation, Predictions (assumptions about what might go wrong during implementation), the Implementation process, Value (the evaluation criteria), Methods (for data collection and evaluation), Use (what the results from the evaluation will be used for), and Information sharing (how the results can be shared with others).

### C. Piloting the approach in the workshop

The evaluation approach and method for application were piloted in a workshop at Delft University of Technology, which provided us with the opportunity to prepare for more research where we intend to further develop and refine the larger evaluation framework. There were 28 active workshop participants, including educational researchers, engineering educators, and educational advisors. The participants were divided into groups of four. Each participant was assigned a persona: 2 x Academics (Educators), 1 x Educational Advisor and 1 x Teaching Assistant. Each persona had a different role assigned and embodied a different attitude towards education and the evaluation of innovations.

---

Funded by the 4TU Centre for Engineering Education

Each group received a set of nine cards containing prompting questions, pens and an A0-size paper, folded into nine blocks. Each card was allocated a block on the paper for the participants to respond to the questions, which then guided them in the development of their own evaluation plan for their particular innovation. The different topics indicated on the cards included *Innovation*, *Purpose*, *Predictions*, *Implementation process*, *Value*, *Methods*, *Use*, and *Information sharing*. Once each card is addressed, the group will have developed a very basic plan for evaluating their chosen innovation which would ideally display alignment between Value, Methods and Use, and is indicated in the last remaining block, *Conclusions*.

In this instance of the workshop, their plans were captured on a simple A0 sheet, but could alternatively be done on an adapted version of the business model canvas [10], an impact assessment map [11], or something similar. Feedback on the workshop content and the evaluation approach was collected through observation and informal discussions with the participants. Only the evaluation plans from the groups ( $n = 5$ ) were included for this study where all participants in the group agreed to sign the consent forms. Participants' feedback and input during the observations were also included for improving the workshop.

### III. RESULTS AND DISCUSSION

#### A. The evaluation approach

Although there is plenty of *evaluation research* available, their scope is mostly limited to a single innovation, while their evaluation criteria are confined to evaluating either the features, fidelity of implementation, student perceptions, or effect on student learning. The findings from these evaluation studies were found to be context-specific, pertain only to the specific qualities of the innovations under investigation and less oriented for application at scale. Studies that purposefully addressed methodologies for evaluating innovation are difficult to come by and so we decided to develop a new approach to evaluating innovation that would suit our context. We turned to the field of Programme Evaluation as a starting point of a development of an our framework.

At first glance, the Kirkpatrick model (evaluating reaction, learning, behavior and results) [12] and Theory of Change (looking at inputs, activities, outputs, outcomes and impacts) [13] seem ideal for evaluating innovation, and many have used these models for that purpose [14], [15], [16]. Indeed, there are many elements that we considered for the development of our framework. However, these models are limited by their linearity (step-by-step processes) and simplicity seen in the lack of consideration of interactions between elements. The dynamic nature of innovation requires a broader perspective on value and criteria for measuring impact. There is risk of oversimplifying its intricacies and its context of implementation. A simple cause-and-effect relationship cannot be assumed and instead, a more systematic approach is needed [17], [18].

Davidson [6] offers a more comprehensive view on program evaluation, though heavily based on Scriven's Key Evaluation Checklist [19]. This framework for evaluation includes a collection of checkpoints grouped under three main parts: Foundations (with checkpoints Background and Context; Descriptions and Definitions, and others); Sub-evaluations (including Process Evaluation; Outcome Evaluation etc.); and lastly the Conclusion. Each checkpoint

is detailed and extensive. Additionally, with its linear progression from step 1 to 11, its rigidity does not accommodate flexibility and sensitivity to the emerging nature of educational innovation.

Christie and Alkin [20] categorized three main evaluation approaches to evaluation in their Evaluation Theory Tree. The Evaluation Theory Tree maps authors and their evaluation approaches within three main branches: Valuing, Methods and Use (or Utility). In the Use tradition of evaluation, evaluations are conducted to inform decision making and the results of evaluations are therefore used to fulfill specific needs of the stakeholders who commission the evaluation [21]. Within the tradition of Valuing, on the other hand, the evaluator determines the value (or merit) of the subject under evaluation, whether objective or subjective [20]. The Methods branch emphasizes rigorous knowledge construction [20] and favors well-designed methods for data collection and analysis [21].

Although the three traditions are seen as separate branches, their arrangement also reflects their relation to one another. They are arranged in such a way that each branch (primary emphasis in evaluation approach) touches the adjacent branches (secondary emphasis in evaluation approach).

We argue, however, that the three approaches cannot be separated from one another. Evaluation, as with domain specific research, should be approached with the same level of scholarship and therefore scientific rigour should be evident in our evaluation *methods* [22], [23], by choosing the most appropriate data collection and reporting methods based on the criteria for evaluation (value). The chosen evaluation methods should facilitate the intended utility (*use*) of the evaluation results. What the evaluation will be used for should determine the evaluation criteria (*value*). The evaluation criteria (value) in turn, should be determined by the purpose of the evaluation results, i.e. what they will be used for (*use*).

The purpose of the evaluation should be clearly defined so that the results can be *used* deliberately to improve educational practice and outcomes. Our units of measurement should be determined by what we, and our stakeholders, attach *value* to. The value we attach to the innovation should be produced and expressed using valid and reliable *methods*.

Value indicators such as sustainable educational change, student performance, professional development of educators, or our success in teaching transversal skills come to mind. As an overarching goal, in our evaluation approach, the three branches of evaluation should be aligned.

Innovation is complex due to its unpredictable and volatile nature and so we need evaluation methods that are inherently flexible to allow for responsiveness and emergence, instead of constraining ourselves with predefined evaluation parameters and prescribed criteria for measurement. Once the novelty of an innovation wears off parameters and criteria might stabilize and be refined by the evaluator. Although 'innovations' might take the form of 'programs' and vice versa, it is at this point that our evaluation approach diverges from program evaluation.

#### B. Application of the innovation approach

While working through the card deck, some participants ordered the cards in a sequence that made most sense to them. Our framework intentionally supports this approach, since one

would choose the starting point that most reflects one's approach to evaluation. For example, receiving a top-down request for statistics on performance indicators, the *methods* would be a good starting point to make sure the requested data is collected and communicated. If used for professional development of teaching staff, for example, the evaluation can focus on collecting data that provides evidence of reflection and progress in skills development.

Not all groups followed the encouragement and instruction to iteratively move between Value, Methods and Use to ensure alignment, and future instructions will need to make this more explicit. Furthermore, the same color will be assigned to these cards, and in a different tone than the cards that proceed and follow them. The brainstorming questions on these cards will also be simplified to prevent cognitive overload, enable participants to work through them more efficiently, and to reduce any possible overlap between the cards.

Due to the low number of groups and extensive need for improvement of the workshop, no strong conclusions can be drawn yet on the content provided by the participants. Still, three participants expressed their interest in using the framework for evaluating their own innovations. Participants also expressed their agreement on the flexibility of the approach to be used on various innovations, from educational games to technologies.

#### IV. LIMITATIONS

A number of participants expressed their agreement on the flexibility of the approach to be used on various innovations and wish to continue using the framework. However, due to the low number of groups and extensive need for improvement of the workshop, no strong conclusions can be drawn yet. Since the results from the literature review for this study were limited by the exploratory nature of our literature review, a more systematic review will follow.

#### V. CONCLUSIONS

This paper addressed the need for a theory-based, practical approach to refine existing innovation evaluation practices, and could serve as a starting point for the development of a more comprehensive evaluation framework. With much innovation taking place at course level, the framework in development will aim to help educators reflect on the value of their educational innovations and to enable more visibility on the educational transformation that meets the demands of industry and society. The workshop materials are currently under review and will be made available for adaptation and use later on within the larger research initiative.

#### REFERENCES

- [1] H. Robles, K. Burden, and K. Villalba, "A Socio-Cultural Approach to Evaluating and Designing Reading Comprehension Apps for Language Learning," *International Journal of Mobile and Blended Learning*, vol. 13, no. 1, pp. 18–37, Jan. 2021, doi: 10.4018/IJMBL.2021010102.
- [2] M. Yap, "A study of TESCA an Indonesia' higher education e-readiness assessment model," in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, Nov. 2015, pp. 1–7. doi: 10.1109/ICITSI.2015.7437720.
- [3] L. Chamba-Eras and J. Aguilar, "Augmented Reality in a Smart Classroom—Case Study: SaCI," *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, vol. 12, no. 4, pp. 165–172, 2017, doi: 10.1109/RITA.2017.2776419.
- [4] J. C. F. De Winter, D. Dodou, and A. H. A. Stienen, "ChatGPT in Education: Empowering Educators through Methods for Recognition and Assessment," *Informatics*, vol. 10, no. 4, p. 87, Nov. 2023, doi: 10.3390/informatics10040087.
- [5] L. Caley, S. J. Williams, I. Spernaes, D. Thomas, D. Behrens, and A. Willson, "Frameworks for evaluating education programmes and work related learning: a scoping review," *JWL*, vol. 33, no. 6, pp. 486–501, Jun. 2021, doi: 10.1108/JWL-09-2020-0157.
- [6] V. C. Tassone, H. J. A. Biemans, P. den Brok, and P. Runhaar, "Mapping course innovation in higher education: a multi-faceted analytical framework," *Higher Education Research & Development*, pp. 1–15, Oct. 2021, doi: 10.1080/07294360.2021.1985089.
- [7] Y. Cai, "From an Analytical Framework for Understanding the Innovation Process in Higher Education to an Emerging Research Field of Innovations in Higher Education," *The Review of Higher Education*, vol. 40, no. 4, pp. 585–616, 2017, doi: 10.1353/rhe.2017.0023.
- [8] S. Maghsoudi, C. Duffield, and D. Wilson, "Innovation Evaluation: Past, Current Models and a Framework for Infrastructure Projects," *International Journal of Innovation Science*, vol. 7, no. 4, pp. 281–298, Dec. 2015, doi: 10.1260/1757-2223.7.4.281.
- [9] M. Fullan, *The new meaning of educational change*, 4. ed. London: Routledge, 2007.
- [10] Business Model Alchemist and pharma industry, "Business Model Canvas". Accessed: March 17, 2023. [Online]. Available: <https://www.strategyzer.com/blog/5-tools-for-new-business-success>.
- [11] S. Verdonshot and D. Stevens, *Impact onderzoek: Samen in kaart brengen wat een leertraject oplevert*. Utrecht, Netherlands: Libertas Pascal, 2021.
- [12] J. D. Kirkpatrick and W. K. Kirkpatrick, *Kirkpatrick's Four Levels of Training Evaluation*. in Titolo collana. Alexandria, VA: ATD Press, 2016.
- [13] "Center for Theory of Change: Setting Standards for Theory of Change," *TOC Origins*. Accessed: Oct. 22, 2023. [Online]. Available: <https://www.theoryofchange.org/what-is-theory-of-change/toc-background/toc-origins/>.
- [14] M. Paull, C. Whitsed, and A. Girardi, "Applying the Kirkpatrick model: Evaluating an Interaction for Learning Framework curriculum intervention," *Issues in Educational Research*, vol. 26, no. 3, pp. 490–507, 2016.
- [15] R. Hains-Wesson, "Something went terribly wrong: failing successfully," *HIGHER EDUCATION RESEARCH & DEVELOPMENT*, vol. 41, no. 3, pp. 729–742, Apr. 2022, doi: 10.1080/07294360.2021.1872056.
- [16] M. B. Cahapay, "Kirkpatrick Model: Its Limitations as Used in Higher Education Evaluation," *International Journal of Assessment Tools in Education*, vol. 8, no. 1, pp. 135–144, 2021.
- [17] C. Ives, L. McAlpine, and T. Gandell, "A Systematic Approach to Evaluating Teaching and Learning Initiatives in Post-secondary Education," *Canadian Journal of Higher Education*, vol. 39, no. 2, pp. 45–76, 2009.
- [18] W. Trochim et al., "The Guide to the Systems Evaluation Protocol," 2015, doi: 10.13140/RG.2.1.5003.2809.
- [19] M. Scriven, "The Checklist Imperative," *Evaluation*, vol. 2019, no. 163, pp. 49–60, Sep. 2019, doi: 10.1002/ev.20374.
- [20] M. C. Alkin and C. A. Christie, Eds., *Evaluation roots: theory influencing practice*, 3rd ed. New York: Guilford Publications, 2023.
- [21] M. C. Alkin and C. A. Christie, "An evaluation theory tree," *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences*, Jan. 2013.
- [22] P. Pohlenz, "Innovation, professionalisation and evaluation: implications for quality management in higher education," *Quality in Higher Education*, vol. 28, no. 1, pp. 50–64, 2022, doi: 10.1080/13538322.2021.1951452.
- [23] M. Q. Patton, *How to Use Qualitative Methods in Evaluation*, 2nd ed. in *Program Evaluation Kit*, no. 4. Newbury Park, CA: Sage Publications, 1987.