scope

Contemporary Research Topics

learning & teaching 15

October 2025

Article

https://doi.org/10.34074/scop.4015004

DESIGNING A CONCEPTUAL TOOL FOR AI-SAFE ASSESSMENT IN VOCATIONAL EDUCATION

Bruno Balducci, Mairead Fountain and Ana Terry

Published by Otago Polytechnic Press.

Article https://doi.org/10.34074/scop.4015004

DESIGNING A CONCEPTUAL TOOL FOR AI-SAFE ASSESSMENT IN VOCATIONAL EDUCATION

Bruno Balducci, Mairead Fountain and Ana Terry

INTRODUCTION

Recent progress in generative artificial intelligence (GenAI) is compromising the validity of non-exam assessments across many sectors of education, including higher education and vocational education and training (VET). As AI evolves, the security of these assessments becomes increasingly threatened. In addition, as this technology becomes more widespread, the need for professional and vocational courses to use and teach the use of AI continues to grow. Hence it becomes increasingly important for maintaining the validity of programmes to assess students in the same skills they have developed through their learning (Moorhouse et al., 2023).

The constant erosion of assessment security is creating a "wicked problem" (Stolterman, 2008) for all stakeholders in education, including vocational assessment designers. Such problems, according to Stolterman (2008), require a "designerly" approach to human-computer interaction through a disciplined and rigorous process. What we are proposing here is a design solution encompassing both human-human and human-AI interaction, one that will enable assessment security to be achieved through authentic non-exam tasks in a wide range of fields, and where both student-student collaboration and the legitimate use of AI are made possible.

This article reports on a research project aimed at creating a conceptual instrument to help assessment designers. Our project was focused on the following questions:

> What is the most user-friendly, efficient, and widely applicable conceptual design tool we can create to enable legitimate AI use and prevent AI misuse in non-exam assessments?
>
> What improvements and insights can be gained from trialling this instrument by means of a Research-through-Design (RtD) approach?
>
> What new knowledge on vocational assessment can be generated by such an approach?

In this article, we describe the methodology and process through which we achieved our research aims, as well as the resulting conceptual framework and design tool (available at: https://aisafedesign.com), and how this can be used in VET.

LITERATURE REVIEW

In the context of vocational assessment, the problem of GenAI, although ill-defined, can be apprehended in terms of "AI misuse." The literature largely describes this as academic misconduct or cheating, against which the standard response is to use academic integrity policies whose aim is to "raise awareness and discourage misuse" (Xie et al., 2023, p. 80). From a design perspective, such misuse may be understood as any use of AI in assessment work that prevents the full and fair assessment of students' learning, while legitimate use may be seen as the direct opposite.

With AI's continuing development, this problem of misuse will become ever more pernicious. In the long term, the only viable way of protecting assessments will be through their design. Yet little attention has so far been given to this approach as compared to standard academic integrity procedures. It is therefore important to make the distinction between the two approaches clear. Henceforth, the term "AI-safe" in this article refers to assessments in which AI use by students may or may not be assessed and whose security against AI misuse rests entirely on the way they are designed.

The question now arises as to which guiding principles to use in AI-safe assessment design. A number of ideas have been put forward in the literature, the most common being to make assessment tasks personalised, creative, innovative, reflective, or critically thoughtful (Jürgen et al., 2023; Michel-Villarreal et al., 2023; Perera & Lankathilaka, 2023). These suggestions may be implemented to varying extents according to the field of practice. However, since AI-generated material can appear to have any of these qualities, they do not in themselves prevent AI misuse. Other factors need to be considered.

The most important concept for AI-safe assessment design is authenticity. Writing before the public release of ChatGPT in 2022, Sotiriadou et al. (2020, p. 2134) noted that "authenticity of assessment may hold the key to reducing academic misconduct." Their argument that authenticity is vital not only for skill development and employability, but also to prevent misconduct in assessments, can be readily applied to the problem of AI misuse. Sotiriadou et al. (2020) establish their conception of authenticity with reference to the more comprehensive work of Gulikers et al. (2004), whose focus on the skills and knowledge required for professional life is ideally suited to non-exam vocational assessment.

As well as a characteristic of the assessment itself, authenticity is defined by Gulikers et al. (2004, p. 70) as the "relation to the criterion situation derived from professional practice" (in other words, a relevant situation that students may be confronted with in the workplace). This is identified through: (a) the assessment task, (b) the physical context, (c) the social context, (d) the assessment result, and (e) the assessment criteria. Because of its alignment with principles of assessment design, this framework remains the most useful instrument available for the study of authenticity in VET. It is their resemblance to real-life tasks, and their physical and social contexts, that makes assessments authentic. However, in the age of GenAI, there is also a need for a similar framework to enable AI-safe design.

From a human-centric perspective, the authenticity of an assessment task depends on whether it resembles comparable human activity in real life. Here the evidence of learning to be assessed will be found in the way that problems are solved by students using their knowledge and skills. In a world where such activity is undergoing radical transformation, assessment can only be authentic if it reflects this change. This is doubly important in vocational education, where it is essential to keep up with developments in the workplace.

As indicated in the framework by Gulikers et al. (2004), context also plays an important role. Whether or not AI use is essential to the criterion task, assessment designers need to consider the context in which the kind of work to be assessed is normally done. Given the vast quantity of material accessible to GenAI, we need to consider the scope of each assessment context. The more restricted this scope, the less data, information, or commentary will be available online, thereby preventing AI misuse altogether or limiting AI use to certain tasks. What is more, context-specific design enhances the authenticity of assessment.

As well as the physical context, Gulikers et al. (2004) see social context as a separate dimension of assessment design, which may include (some degree of) collaboration. In this case, the need to collaborate with their peers will impose constraints on what students may or may not do in order to complete their assessment. In addition to facilitating the exchange of ideas, collective decision-making and accountability to a designated peer group will provide a measure of security.

Child and Shaw (2016) note that, for optimal assessment, process must be distinguished from outcome. The ability to do this has important implications for the way students are assessed (Gulikers et al., 2004). Different criteria must be identified for evaluating student performance both in the process of assessment work and in the final product resulting from this work. It follows that the evidence on which performance in the process is judged cannot be based on the product itself but needs to come from other sources.

The process then leads to the outcome or product which is to be assessed separately. This is of course where a lot of AI use or misuse will have occurred. AI is not only able to produce material that students may submit for assessment but can remove the need to carry out preliminary work through which such material is generated. Therefore, non-exam assessment is unlikely to be safeguarded against AI misuse through a one-dimensional approach. It is the combined application of the principles we have discussed that will provide the kind of security on which such assessment once depended.

METHODOLOGY AND DESIGN

Research-through-Design (RtD) is defined by Zimmerman et al. (2010, p. 310) as an "approach that employs methods and processes from design practice as a legitimate method of inquiry." In view of the research aims outlined above, it therefore provided an appropriate methodology for our project. We found that pragmatism, identified in a survey of paradigms for RtD as the most closely aligned to quality indicators (Prochner & Godin, 2022), was best suited to our requirements. Prochner and Godin (2022, p. 7) summarise this approach as "the view that reality can be changed for the better." Our own research seeks to improve assessment practices in order to prevent GenAI from invalidating grades awarded to students. As well as alleviating a wicked problem in VET, our framework could also contribute to the development of knowledge relating to assessment design.

Both Roggema (2017) and Zimmerman et al. (2010) highlight the iterative and reflective nature of RtD. This suggested that the best design process for addressing GenAI in assessment would be one where the initial problem is closely studied, while the solution to be developed is reviewed throughout. Hence, we planned a sequence of tasks for learning and teaching (L&T) staff and subject matter experts (SMEs) in VET using the phases in Roggema's (2017) methodological model, as follows:

Pre-design phase (October 2023)

Stage 1: Investigate and research (from a "designerly" perspective):

- risks posed by GenAI to non-exam assessment security.
- useful assessment design concepts for AI risk-mitigation.
- needs of assessment designers in vocational disciplines.

Design phase (November 2023–September 2024)

Stage 2: Develop initial framework / assessment design tool.

Stage 3: Trial v1 design tool with focus groups.

Stage 4: Data analysis / review of conceptual tool.

Stage 5: Trial v2 design tool using Qualtrics survey.

Stage 6: Data analysis / design of project website / conceptual tool review.

Stage 7: Trial v3 design tool / website through individual consultation.

Stage 8: Data analysis / redesign of conceptual tool.

Post-design phase (post-September 2024)

Stage 9: Website launch.

Stage 10: Research publication, conference presentation.

Project participants

- Research, data analysis, design: L&T specialists in the project team.
- Website development: L&T specialist, SME in project team.
- Trials: L&T specialists, SMEs.

Trials procedures

v1: Reflection questions and group discussions.

v2: Assessments analysis and detailed questionnaire.

v3: One-to-one criterion-based project evaluations.

Overall, this process included a series of feedback loops or "cycles of testing and refinement" (Reeves, 2006, p. 59). Our plan allowed us to share documentation at each stage with the L&T specialists in the project team to develop appropriate protocols and guidelines. Data analysis and reflections in project team meetings were recorded and circulated for review and redesign purposes. Ethics approval for this research was granted by the Otago Polytechnic Auckland International Campus (OPAIC) Research Ethics Committee. Informed consent was obtained from all participants.

Our methods needed sufficient transparency of procedure and acceptable interpretation of data in order to meet research quality standards. Prochner & Godin (2022, p. 8) recommend the following evaluation framework:

a) *traceability* (what was done during the research)

b) *interconnectivity* (links between important concepts and elements)

c) *applicability* (useable knowledge)

d) *impartiality* (positioning of researchers' bias)

e) *reasonableness* (rationale for choices made).

When applied to our research plan, these aims were met in the following ways:

a) We ensured traceability by recording procedures and decisions and circulating these records to project team members.

b) For interconnectivity, team members engaged in exploratory discussions on the nature of the project framework.

c) For applicability, we held practical trials with SMEs / L&T staff.

d) For impartiality, we used feedback loops to foster reflectivity in our data handling.

e) For reasonableness, we described in our records all the key choices made and justified them with reference to project objectives.

The main limitation of the project was the timeframe, which did not allow a full evaluation of the tool's applicability before publishing our findings. Impact is key in pragmatic enquiry, as this approach assumes that society can be changed for the better through research. However, a full appraisal of our project in terms of its impact on assessment practice and adaptability to future AI must inevitably remain beyond the scope of this article.

FINDINGS AND PROJECT OUTCOME

From our literature review, we were able to identify the most important concepts for dealing with the negative effects of GenAI. Drawing from our own experience of assessment design, we then devised a new framework for our conceptual tool. We were aware from previous research (Balducci & Sultana, 2024) that such a framework may contain too few or too many categories to be effective. Hence, we focused on what seemed the most coherent and promising concepts with which to develop a prototype.

The first trial

Having drafted our framework, we were ready to design a prototype (v1) using these categories: specificity, practicality, collaboration, process workflow, and information flow. Our purpose was to provide a clear definition for each category in order to avoid any confusion that may result from the broader and even multiple meanings of the concepts on which these categories were based (context, authenticity, collaboration, process, and product). We also designed a set of guiding questions (samples are discussed below) and examples to illustrate the meaning of each category, derived from actual assessment practice in vocational courses delivered at Otago Polytechnic and Auckland International Campus.

We now had an instrument ready for testing. The trial collected responses from 14 participants (seven lecturers in different disciplines and seven SMEs). Participants were asked to consider what it would be like to use the tool to evaluate assessments they were familiar with. Qualitative data was gathered from the trial transcripts and the participant feedback forms, with particular emphasis on the use of terminology, conceptual definitions or explanations, and guiding questions.

Overall, participants reported that the conceptual tool was clear and useful. Their discussions indicated that v1 was fundamentally compatible with good practice in assessment design and could readily be used across different disciplines. Some doubts were expressed, however, relating to our use of terminology. Some participants struggled to unpack the meaning of questions where "data" and "information" were used interchangeably. The following example of such a question will suffice: To what extent is task achievement dependent on the use of information and/or data that has been entirely generated in the context of the task itself? As a result, we decided to replace the term "data" with "information" throughout. Students might need to submit unprocessed data as part of their work (for instance, in the form of appendices), but it is hard to imagine an assessment where they would not be required to do something with this data (and therefore treat it as meaningful information).

Some questions relating to how students may collaborate with each other were also deemed confusing. For example: Does the student carry out an individual review of their own task achievement? Participants began to wonder if this might be about getting students to take personal instead of collective ownership. In response to this feedback, we decided that the purpose of these questions had been misunderstood. The one above was in fact

aimed at identifying individual contributions to a group task in order to assess the performance of each student. Evidently, the questions (or the context in which they appeared) needed to be rethought so that the type of teamwork involved could be clearly defined.

After due consideration, we also concluded that "collaboration" was still an appropriate umbrella term that can refer to different types of arrangements for the production of assessed work. However, we agreed that not all vocational disciplines offer the same scope for collaborative activity. As Gulikers et al. (2004) have argued about authentic assessment, "if the real situation demands collaboration, the assessment should also involve collaboration, but if the situation is normally handled individually, the assessment should be individual" (p. 74). We would therefore allow for partial collaboration or individual work in the conceptual tool.

The second trial

The next version (v2) incorporated changes made after the first trial. Although there were only 12 respondents, this version was tested more rigorously than v1 because the participants included L&T specialists from an external organisation and SMEs in more diverse disciplines who completed a more extensive questionnaire to evaluate v2 in their own time.

The evaluation questionnaire was in two parts. Questions one to five asked for feedback on the five key concepts, while questions six to ten sought feedback on the conceptual tool as a whole. Since it was mostly aimed at qualitative data, the main quantitative data of value that we gathered was in answer to: What is your general feeling about this tool? The responses were 33 percent "positive," 17 percent "quite positive, 33 percent "neutral, and 17 percent "slightly negative." No respondents chose "negative." The reasons given for the "slightly negative" responses were that the language used was too complex and the tool was difficult to apply. Our priorities thus became to simplify the use of language throughout, and make v3 more accessible and user-friendly.

Many of the other questions were broad and open-ended (for example, What is your feedback relating to…?). For each one, respondents had the space to explain their answers, which enabled us to collect sufficient qualitative data for our research. Once anonymised, this data was analysed by members of the project team, first independently, then in consultation with one another. As experienced practitioners in assessment design, we could interpret our data from a learning and teaching perspective in order to decide whether action was needed to improve v2, and if so, what improvements could be made. These decisions were grouped into different types of implementation, and supporting evidence from the survey was numbered accordingly:

1. changes to the design tool's content.
2. simplified use of English.
3. changes to the graphic design.
4. new material for the website.

The data gathered confirmed the need to focus on priorities identified through the question on the participant's general feeling about the tool. Unlike v2, the next version would be more of a redesign than a series of tweaks. The findings were circulated to all project team members for consideration from graphic design and website design perspectives. They were subsequently reviewed and action points were drawn up, as summarised below:

A) Write a preamble to introduce the conceptual tool.

B) State the assumptions underlying its design.

C) Write annotations to clarify important points of detail.

D) Provide more generic examples of how to apply the framework.

E) Include suggestions for making assessments more AI-safe.

F) Improve graphics to illustrate the relations between key concepts.

G) Use simpler English and avoid terminology (or define its use).

In a few instances, where there was the possibility of making alterations to the conceptual tool itself or adding extra material that would influence the way it is perceived, it was decided to have: i) a streamlined version on the website, and ii) a fuller version as a downloadable PDF.

The following thoughts from respondents had the most influence on how the conceptual framework or design tool were redeveloped and presented in v3:

- It is unclear whether assessment security or academic integrity is the aim.
- Some users might not see there is a framework involved.
- How the key concepts relate to each other needs explanation.
- This tool seems to be about eliminating AI use altogether.
- The questions only apply to some (not all) types of assessment.
- To understand the questions, you need to know their underlying purpose.
- The examples are too disjointed or discipline-specific.
- The graphics don't convey any conceptual information.

What these responses brought home to us was how the different elements of the tool would have to be fully integrated into our website design, while the presentation of its purpose, basic principles, and possible applications ought to address the concerns and expectations of our targeted audience more directly.

Of the suggestions which were not incorporated directly into the redesign, three in particular did help us clarify what needed to be conveyed to potential users. The first related to content, while the other two were about application:

A) Critical thinking should be included as a key concept.

B) A model assessment would help to show how to apply the framework.

C) A list of criteria for assessment security would be useful.

We decided to respond to these suggestions through our FAQs, respectively:

A) In a section entitled "The thinking behind our conceptual framework," we would establish that context is the relevant key concept under which critical thinking may be subsumed. We would do this in such a way as to recognise the value of critical skills in themselves, whilst pointing out that assessment security is entirely dependent on the context in which such skills are applied.

B) We brainstormed ways of reworking our examples of AI-safe assessment into short and simple ideas aimed at different disciplines, thereby allowing for diverse practices within VET.

C) We would also continue to avoid any suggestion that our tool might be used as a checklist. Assessment design can proceed in any order or fashion that suits the designer. While we took on board the suggestion of highlighting inter-relationships between key concepts, we also needed to acknowledge that assessment design is a creative activity that should not be confined to standardised procedures.

The third trial

After redesigning the tool according to the findings of the second trial, we wrote annotations and organised material for the website. We started by changing some of our terminology to make it more recognisable to users and defining our key concepts as simply as we could.

Our concepts	Is your assessment …	What we mean
	context-specific?	The context is clearly defined and relatively limited in its scope.
	authentic?	The work done is similar to what professionals often need to do.
	collaborative?	There is a joint effort requiring equal commitment from everyone.
	process-driven?	How work is done is as important as its end product.
	generative?	New information, ideas, artefacts, designs etc. are produced.

Figure 1. AI-safe assessment design: Key concepts (Balducci et al., n.d.).

We then built a WordPress website, which was reviewed in one-to-one consultation by three evaluators in academic management using these criteria:

- The framework is clear, coherent, and sufficiently broad.
- The tool is user-friendly, efficient, and widely applicable.
- The website is well-designed, informative, and easy to navigate.

Since, according to our evaluators, these criteria were all met, this last trial enabled us to establish that our design was generally fit for purpose.

Discussions with evaluators also yielded suggestions for fine-tuning the website, leading to the following changes:

- a clear statement that it is primarily intended for VET;
- equal focus on AI misuse *and* legitimate AI use, and
- emphasis on AI-safe design as risk mitigation (not 100 percent security).

Finally, we rewrote the section on how the concepts relate to each other ("How to use AI-Safe?") so as to clarify the nature of the security that AI-safe design can provide. We identified the following areas for assessment designers to consider (Balducci et al., n.d.):

1. a specific context for learners to work in;
2. authentic work for learners to do in this context;
3. ways in which collaboration might enable them to do this work;
4. an observable process for working (partly) in collaboration, and
5. the type of content to be generated through this process.

DISCUSSION AND CONCLUSION

Institutional responses to the wicked problem of GenAI in assessment have generally taken the form of academic integrity policies and procedures, together with guidelines and recommendations for assessment redesign which are mostly aimed at minimising AI misuse (Evangelista, 2025; Moorhouse et al., 2023). In this project, we set out to develop a conceptual tool that addresses both misuse and legitimate use of AI by tertiary vocational students in their assessment work. This aim was eventually achieved by means of a design process with separate trials and feedback loops.

Our Research-through-Design approach enabled us to make a range of improvements from insights gained by trialling three successive versions in order to have an instrument in our purpose-built website which fulfilled the aims of our project. One such improvement was to include our rationale for the five key concepts in our framework. Another main improvement was the elaboration of interrelationships between these concepts. By discouraging compartmentalised thinking, we can render the framework more suitable as an instrument for coherent and global reflection on the user's own assessment design.

These insights into the design process culminated in the definition of the final key concept in the framework, namely generativeness. Here the concept is to be understood, not simply as the product generated by the assessment process, but one that can be identified as essentially the result of human (as opposed to artificial) intelligence. An assessment becomes "generative" through the application of other concepts in the framework (a possible exception being collaboration, which may not always apply). Therefore, each step taken by the designer in relation to context and other key concepts will contribute to the validity of the assessment content generated by students.

Whereas in the past, academic misconduct was limited to plagiarism and contract cheating, the development of GenAI has forced us to reconceptualise assessment design to ensure that we can differentiate between genuine student work and the activity of AI. With our framework and conceptual tool, we believe that we now have the means to do this.

Bruno Balducci joined the Otago Polytechnic Auckland International Campus in 2021. He supports academic staff to design, develop and deliver courses and assessments. Prior to this, he worked as a training and development manager, curriculum developer, teacher trainer and lecturer in the UK, Vietnam and New Zealand. He holds a research MA in English/French translation from the University of Leeds, a Cambridge ESOL Diploma in English Language Teaching to Adults (DELTA) and a Postgraduate Certificate in Further Education (PGCE). His current research interests include experiential learning, authentic learner activities and AI-safe assessment design.

https://orcid.org/0000-0002-0911-4114

Mairead Fountain is a Principal Lecturer in learning design and tertiary education at Otago Polytechnic. She leads the NZCATT programme and chairs the AI Steering Committee. With a background in neuroscience and educational technology, she is passionate about bicultural practice, inclusive assessment, and authentic learning especially for second-chance learners.

https://orcid.org/0000-0003-0575-5368

Ana Terry is a Principal Lecturer and Learning and Teaching Specialist at Otago Polytechnic. Her area of interest lies in AI's visualising capabilities and its role in the creative domain, particularly how its use impacts the creative process, cognitive development, and visual literacy. She also serves on the committee of the International Visual Literacy Association (IVLA).

https://orcid.org/0000-0001-9437-9522

REFERENCES

Balducci, B., Fountain, M., & Terry, A. (n.d.). *How to use AI-Safe*. AI-Safe Assessment Design. https://aisafedesign.com/how-to-use-ai-safe/

Balducci, B., & Sultana, K. (2024). Analysing modes of reflection in experiential learning: A sociocultural perspective on student peer editing. *SN Social Sciences*, *4*, Article 136. https://doi.org/10.1007/s43545-024-00937-2

Child, S., & Shaw, S. (2016). Collaboration in the 21st century: Implications for assessment. *Research Matters*, *22*, 17–22. https://doi.org/10.17863/CAM.100344

Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology*, *17*(1), Article ep559. https://doi.org/10.30935/cedtech/15775

Gulikers, J., Bastiaens, T., & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–86. https://doi.org/10.1007/BF02504676

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D., Thierry-Aguilera, R., & Gerardou, F. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences*, *13*(9), Article 856. https://doi.org/10.3390/educsci13090856

Moorhouse, B., Yeo, M., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers & Education Open*, *5*, Article 100151. https://doi.org/10.1016/j.caeo.2023.100151

Perera, P., & Lankathilaka, M. (2023). AI in higher education: A literature review of ChatGPT and guidelines for responsible implementation. *International Journal of Research and Innovation in Social Science, 7*(6), 306–314. https://dx.doi.org/10.47772/IJRISS.2023.7623

Prochner, I., & Godin, D. (2022). Quality in research through design projects: Recommendations for evaluation and enhancement. *Design Studies*, *78*, Article 101061. https://doi.org/10.1016/j.destud.2021.101061

Reeves, T. (2006). Design research from a technology perspective. In J. Van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 64–78). Routledge. https://doi.org/10.4324/9780203088364

Roggema, R. (2017). Research by design: Proposition for a methodological approach. *Urban Science*, *1*(2). https://doi.org/10.3390/urbansci1010002

Jürgen, R., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.23

Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, *45*(11), 2132–2148. https://doi.org/10.1080/03075079.2019.1582015

Stolterman, E. (2008). The nature of design practice and implications for interaction design research. *International Journal of Design*, *2*(1), 55–65.

Xie, Y., Wu, S., & Chakravarty, S. (2023). AI meets AI: Artificial intelligence and academic integrity: A survey on mitigating AI-assisted cheating in computing education. *SIGITE '23 Proceedings of the 24th annual conference on information technology education*, 79–83. https://doi.org/10.1145/3585059.3611449

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of research through design: Towards a formalization of a research approach. *DIS '10: Proceedings of the 8th ACM conference on designing interactive systems*, 310–319. https://doi.org/10.1145/1858171.1858228