ASCILITE 2025

Future-Focused:

Educating in an Era of Continuous Change

Towards scalable curriculum mapping: Comparing human and GenAI alignment of CLOs to professional standards

Zachery Quince

Centre for Teaching and Learning, Southern Cross University

Curriculum mapping is a critical component of accreditation and continuous improvement in education. However, aligning Course Learning Outcomes (CLOs) to professional standards such as Engineers Australia's Stage 1 Competency Standard remains time-consuming and labour-intensive. This study evaluates the potential of generative AI (GenAI) to support curriculum mapping by comparing automated outputs with expert human judgement. A stratified sample of 141 (10%) first-year CLOs from a national dataset was analysed using both manual review and GenAI (ChatGPT 4o). Each CLO was mapped to the 16 EA Stage 1 Competencies, and outcomes were classified as Match, Manual Only, GenAI Only, or Neither. Overall agreement was 81.2% with particularly strong alignment in professional and personal attributes. Mismatches were most common in technical competencies, where GenAI over-mapped based on keywords or under-mapped due to limited contextual understanding. These findings suggest GenAI can support curriculum review at scale but requires expert oversight for nuanced or discipline-specific outcomes. The work in progress study contributes to the growing literature on AI in education and offers practical insights into hybrid approaches for accreditation and curriculum design.

Keywords: Curriculum mapping, Generative AI, Engineers Australia, Accreditation, Engineering

Introduction

Curriculum mapping is central to ensuring constructive alignment process and accreditation requirements in higher education. Mapping enables educators to align learning outcomes with professional standards, such as the Engineers Australia (EA) Stage 1 Competency Standard, which outlines 16 competencies across technical knowledge (element 1), application (element 2), and professional attributes (element 3) required for graduate accreditation. This alignment provides visibility into how skills are developed across programs. Mapping also offers a critical audit mechanism, helping to identify curriculum gaps and ensure that students are being prepared for contemporary engineering practice (Carew et al., 2013; Quince et al., 2023).

Despite its importance, the process of curriculum mapping is time-consuming, labour-intensive and prone to variability. Manual thematic mapping of Course Learning Outcomes (CLOs) requires expert interpretation, which can introduce inconsistency and bias (Sumsion & Goodfellow, 2004; Uchiyama & Radin, 2009). Previous work (Quince et al., in Press a) involving the manual mapping of over 1,400 first-year engineering CLOs revealed an imbalance between technical and professional skill coverage, but also highlighted the limitations of scale in conducting such reviews by hand.

Recent advances in generative AI (GenAI) offer a potential alternative. Large language models (LLMs) such as ChatGPT have demonstrated the ability to analyse, summarise and thematically align complex educational content (Sakaguchi et al., 2025). Prior research has explored the application of GenAI to professional skills frameworks and found that while GenAI can generate plausible thematic mappings, it often lacks the nuanced judgment required for accurate contextual interpretation (Quince et al., in Press b). These findings suggest that GenAI may serve as a complementary tool, supporting expert-led mapping processes with rapid, repeatable and explainable outputs. While this challenge is well-documented, there remains a lack of validated approaches for scaling CLO-to-competency mapping without compromising interpretive accuracy. This study addresses that gap by evaluating the alignment performance of a GenAI tool (ChatGPT 4o) against expert judgement, offering new evidence for its potential use in accreditation workflows.

This study investigates whether GenAI can effectively map a subset of first-year engineering CLOs to the EA Stage 1 Competency Standard to answer the following questions.

1. To what extent does GenAI align with expert human judgement in mapping CLOs to EA Stage 1 Competencies?
2. What patterns of agreement or mismatch can be identified, and what do they reveal about the strengths and limitations of GenAI in this context?

Methodology

More than 1,400 CLOs from the first-year engineering courses across Australia were collected during the first quarter of 2024. Earlier work identified an imbalance in curriculum emphasis, revealing that technical competencies were represented at approximately twice the rate of professional or personal skills (Quince et al., 2023), which prompted this data collection. The CLOs were manually aligned to the EA Stage 1 Competency Standard by domain experts, forming a validated baseline dataset for this study. This baseline was documented, with more details on the data collection methods (Quince et al., in Press a). The manual mapping was validated by multiple reviewers that has been shown to be a valid method by (Quince et al., 2023; Cooper et al., 2024)

For the current investigation, a stratified sample was drawn to enable comparison with a GenAI automated mapping approach. Every tenth CLO from the original dataset was selected, resulting in a total of 141 CLOs (representing 10% of the original set). These CLOs were then uploaded to a GenAI tool (ChatGPT 4o). After employing standard prompt engineering techniques to iteratively develop and test the prompt for accuracy and reliability, batches of 20 CLOs were then processed for mapping using the validated approach. A batch segmentation approach was undertaken as previous studies have shown that large dataset processing is a limitation of the tools (Naik et al., 2024). The mapping process involved detecting indicative language patterns and matching them to the published definitions and descriptors of the competencies. Each automated mapping was accompanied by a generated rationale, explaining the basis for the selection and providing transparency for comparison.

The manual and GenAI mappings were both converted into a binary matrix format for comparison, with 16 columns corresponding to the EA competencies. A cell was marked where a competency had been mapped to a CLO. The two matrices were then compared cell by cell and each instance was classified into one of four categories: 'Match', where both manual and GenAI mappings agreed; 'Manual Only', where the competency was assigned by the human reviewer but not by the GenAI tool; 'GenAI Only', where it was assigned by the GenAI but not manually; and 'Neither', where neither approach selected the competency.

Finally, results were analysed both at the level of individual competencies and within their broader domains. Total agreement was calculated as the percentage of competencies where both systems either included or excluded the same competency. Mismatch types were also quantified, enabling the identification of over-mapping by the GenAI system (GenAI Only) and under-mapping (Manual Only). These patterns were used to evaluate the strengths and limitations of the automated approach relative to expert judgement.

Results

A total of 141 CLOs were mapped manually and via a GenAI tool to the 16 Engineers Australia Stage 1 Competencies. Each CLO could be mapped to multiple competencies, resulting in 2,256 individual comparison points across each method. The comparison matrix classified each cell as one of four categories: Match, Manual Only, GenAI Only, or Neither. Table 1 summarises the alignment outcomes across all 16 EA Stage 1 Competencies, including match rates and mismatches categorised by mapping method.

Table 1

Alignment of Manual and Automated (GenAI) Mappings to EA Stage 1 Competencies

	Match						**Mismatch**					
	Both Mapped		Neither Mapped		Total		Manual only		GenAI only		Total	
	n	%	n	%	n	%	n	%	n	%	n	%
Overall	189	**8.38**	1643	**72.83**	1832	**81.21**	324	**14.36**	100	**4.43**	424	**18.79**
1.1	27	19.15	52	36.88	79	56.03	33	23.40	29	20.57	62	43.97
1.2	17	12.06	87	61.70	104	73.76	33	23.40	4	2.84	37	26.24
1.3	30	21.28	50	35.46	80	56.74	58	41.13	3	2.13	61	43.26
1.4	0	0.00	138	97.87	138	97.87	2	1.42	1	0.71	3	2.13
1.5	2	1.42	117	82.98	119	84.40	5	3.55	17	12.06	22	15.60
1.6	3	2.13	130	92.20	133	94.33	0	0.00	8	5.67	8	5.67
Element 1		**9.34**		**67.85**		**77.19**		**15.48**		**7.33**		**22.81**
2.1	47	33.33	49	34.75	96	68.09	34	24.11	11	7.80	45	31.91
2.2	9	6.38	44	31.21	53	37.59	84	59.57	4	2.84	88	62.41
2.3	20	14.18	99	70.21	119	84.40	11	7.80	11	7.80	22	15.60
2.4	2	1.42	128	90.78	130	92.20	5	3.55	6	4.26	11	7.80
Element 2		**13.83**		**56.74**		**70.57**		**23.76**		**5.67**		**29.43**
3.1	4	2.84	132	93.62	136	96.45	4	2.84	1	0.71	5	3.55
3.2	15	10.64	90	63.83	105	74.47	36	25.53	0	0.00	36	25.53
3.3	1	0.71	139	98.58	140	99.29	1	0.71	0	0.00	1	0.71
3.4	5	3.55	124	87.94	129	91.49	8	5.67	4	2.84	12	8.51
3.5	3	2.13	128	90.78	131	92.91	9	6.38	1	0.71	10	7.09
3.6	4	2.84	136	96.45	140	99.29	1	0.71	0	0.00	1	0.71
Element 3		**3.78**		**88.53**		**92.32**		**6.97**		**0.71**		**7.68**

At the domain level, the highest overall agreement was observed in Element 3 (Professional and Personal Attributes), with 92.3 per cent of comparisons classified as Match or Neither. Element 1 (Knowledge and Skill Base) had an agreement rate of 77.2 per cent, and Element 2 (Engineering Application Ability) showed the lowest agreement at 70.6 per cent. A visual example of this can be seen in Figure 1.

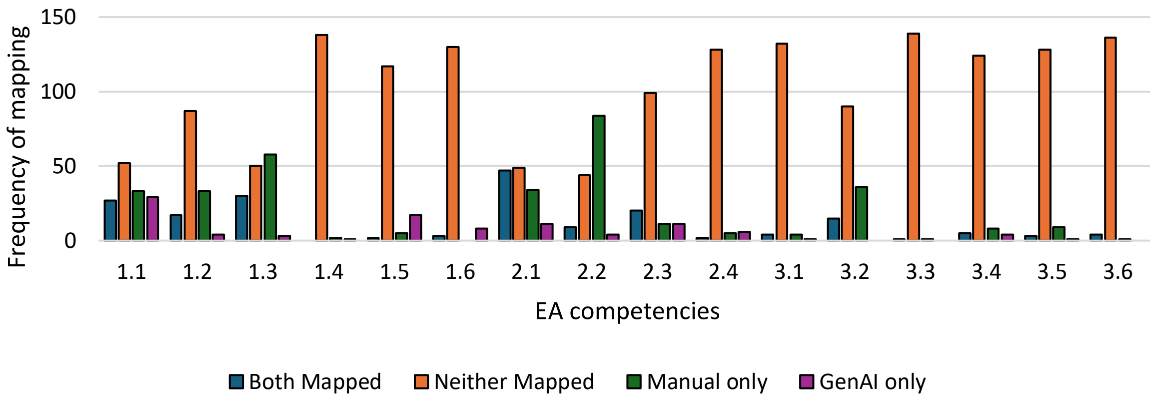

Figure 1. Classification results for the stratified sample broken into the 16 competencies

The highest overall agreement rates were recorded in 3.1 (Ethical conduct and professional accountability), 3.2 (Communication), and 3.6 (Team membership and leadership), all exceeding 95 per cent total agreement. In contrast, competencies such as 2.2 (Application of tools and resources) and 1.3 (Specialist knowledge) showed higher proportions of mismatch, particularly GenAI Only instances. Categorising the CLOs into the areas of teaching 'Engineering', 'Physics', 'Mathematics' and 'Other', demonstrated little variation in the ability for GenAI to show inconsistencies as viewed in Figure 2.

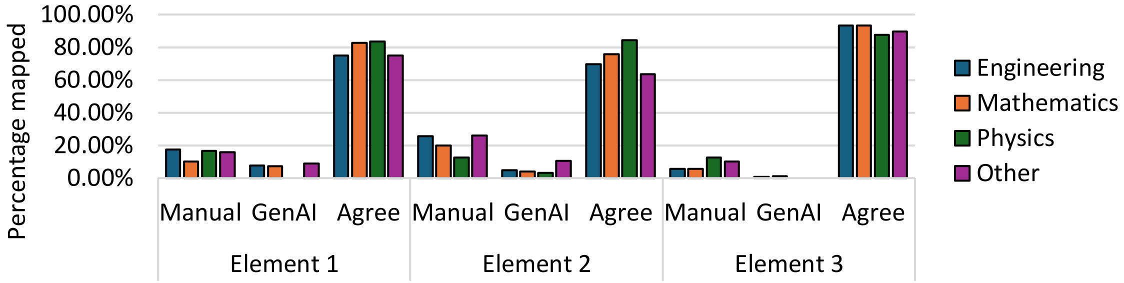

Figure 2. Agreement, manual and GenAI mapping outcomes for discipline CLO's.

Discussion

The comparison between manual and GenAI-based mapping revealed both promising alignment and notable limitations. With an overall agreement rate of 81.2%, the automated method was able to replicate a significant portion of expert decision-making. However, the presence of both GenAI Only and Manual Only mismatches, each accounting for approximately 9.5% of comparisons, highlights the complexity of interpreting CLOs at the semantic level. Agreement was strongest in the third domain of the EA Stage 1 Competencies, which encompasses professional and personal attributes. Competencies such as ethical conduct, communication, and teamwork displayed high match rates, likely due to the clear and consistent language typically used in CLOs that address these areas. These results suggest that GenAI is well-suited to identifying professional outcomes that are expressed explicitly.

In contrast, the lowest agreement was observed in Element 2 (Engineering Application Ability), particularly in Competency 2.2, which relates to the use of tools, techniques, and resources. Here, the GenAI system frequently assigned mappings where human reviewers had not, indicating a tendency to over-map based on surface-level keyword detection. Where CLOs using verbs such as "use," "implement," or "apply" were often mapped to 2.2, even in cases where the outcome did not directly relate to engineering tools or technical processes. This highlights a limitation of the automated approach, where context is essential but not always adequately interpreted.

A different type of limitation was seen in Competency 2.3 (Systematic synthesis and design of solutions), where Manual Only mapping was more common. This indicates an under-mapping tendency by the GenAI tool in cases that required interpretive or design-based judgement, where the intent of the CLO was clear to a human reader but lacked the terminology required to trigger a GenAI match. These patterns suggest that while GenAI can effectively map CLOs with well-defined language, it struggles with nuance. Interestingly, Competency 1.5 (Knowledge of engineering design principles) was not mapped by either method, suggesting a potential gap in curriculum coverage rather than a limitation of the mapping process. This may warrant further investigation into how and where design principles are introduced in the early years of the curriculum.

More broadly, the results affirm the potential of GenAI as a tool for rapid curriculum mapping, especially when used to support rather than replace human judgement. Its ability to identify alignment at scale, generate rationale, and maintain consistency makes it an asset in early-stage curriculum audits. However, the observed mismatch patterns reinforce the importance of expert oversight, particularly in areas requiring contextual interpretation or professional judgement. These findings have practical implications for institutions seeking to streamline accreditation processes, offering a pathway to increase efficiency without compromising quality, provided human expertise remains. This case study demonstrated the applicability of utilising GenAI for curriculum mapping however, it revealed key points where future research lies. To apply the full data set to ensure the trends demonstrated in the stratified sample continue, and for this work to include other common GenAI tools such as ChatGPT and Copilot. Further discipline variation can also strengthen this method.

Conclusions

This study found that a GenAI tool was able to align first-year engineering CLOs to Engineers Australia Stage 1 Competencies with over 80% agreement when compared to manual expert mapping. While high alignment was observed in clearly defined professional competencies, discrepancies in technical and design areas highlight the need for expert oversight. These findings suggest GenAI has potential as a support tool for scalable curriculum mapping but should be used alongside human judgement to ensure accuracy and context sensitivity. Future work will expand on this process with several GenAI tools and models. The dataset will be fully analysed in future work to determine if there are any changes. Future work will also align the length and complexity of the CLO with the ability of GenAI to undertake the process.

Reference list

Carew, A. L., Hadgraft, R. G., & McCarthy, T. J. (2013). *Benchmarking graduate outcomes for undergraduate engineering education*. Proceedings of the 24th Annual Conference of the Australasian Association for Engineering Education. https://aaee.net.au/wp-content/uploads/2018/10/AAEE2013-Carew_Hadgraft_McCarthy-Benchmarking_graduate_outcomes.pdf

Cooper, B., Cowie, B., & Furness, J. (2024). Curriculum mapping as a boundary encounter: Meeting the demands of multiple agendas. *Educational Research for Policy and Practice, 23*(2), 229–250. https://doi.org/10.1007/s10671-021-09299-5

Naik, D., Naik, I., Naik, N. (2024). Imperfectly Perfect AI Chatbots: Limitations of Generative AI, Large Language Models and Large Multimodal Models. In: Naik, N., Jenkins, P., Prajapat, S., Grace, P. (eds) Contributions Presented at The International Conference on Computing, Communication, Cybersecurity and AI, July 3–4, 2024, London, UK. C3AI 2024. Lecture Notes in Networks and Systems, vol 884. Springer, Cham. https://doi.org/10.1007/978-3-031-74443-3_3

Quince, Z., Lidfors Lindqvist, E., & Chen, A. (in Press b). Evaluating GenAI's Ability to Map a Professional Engineering Skills Framework. *Research in Engineering Education Symposium (REES 2025), Johor Bahru, Malaysia.* Manuscript accepted for publication.

Quince, Z., Shari, S., & Cunningham, S. (in Press b). Rethinking Engineering Education for 2035: Progress Towards Closing Skill Deficiencies in Australian First-Year. *Research in Engineering Education Symposium (REES 2025), Johor Bahru, Malaysia.*

Quince, Z., Hills, C., & Maxwell, A. (2023). *Are engineering programs meeting the 2035 professional and personal requirements?* In *Proceedings of the 34th Australasian Association for Engineering Education (AAEE) Conference: Adapting to the Changing Expectations of Students and Industry*, Gold Coast, Qld, Australia. Engineers Australia. https://search.informit.org/doi/10.3316/informit.T2024110400009692147447182

Sumsion, J., & Goodfellow, J. (2004). Identifying generic skills through curriculum mapping: a critical evaluation. *Higher Education Research & Development*, *23*(3), 329–346. https://doi.org/10.1080/0729436042000235436

Sakaguchi, K., Sakama, R., & Watari, T. (2025). Evaluating ChatGPT in qualitative thematic analysis with human researchers in the Japanese clinical context and its cultural interpretation challenges: Comparative qualitative study. *Journal of Medical Internet Research, 27*, e71521. https://doi.org/10.2196/71521

Uchiyama, K. P., & Radin, J. L. (2009). Curriculum Mapping in Higher Education: A Vehicle for Collaboration. *Innovative Higher Education*, 33(4), 271–280. https://doi.org/10.1007/s10755-008-9078-8

Quince, Z. (2025). Towards scalable curriculum mapping: Comparing human and GenAI alignment of CLOs to professional standards. In Barker, S., Kelly, S., McInnes, R., & Dinmore, S. (Eds.), *Future Focussed. Educating in an era of continuous change*. Proceedings ASCILITE 2025. Adelaide (pp. 273-277). https://doi.org/10.65106/apubs.2025.2661