Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10036

A practical use case and IRR based validation of concept maps coding scheme

Pavan Kumar and Joshua Summers✉

University of Texas at Dallas, USA

✉ joshua.summers@utdallas.edu

ABSTRACT: Concept maps have been used to measure student learning and performance comparisons before and after interventions. However, these concept maps may vary in structure and content. The complexity of the maps are difficult to measure when multiple concepts are embedded within individuals nodes. A systematic coding scheme is evaluated for objectivity and reliability using 22 undergraduate researchers. The findings suggest that the coding scheme is suitable and will allow multiple different researchers to generate similar bi-partite graphs from the notional concept maps generated. Additional work is needed to ensure that the semantic content is not invalidated through the coding scheme.

KEYWORDS: information management, conceptual design, design cognition, design education

1. Motivation - concept naps: diversity and challenges

Concept maps are structured (Chang, 2007; Tergan et al., 2006; West et al., 2002), graph-based visual tools (Grundspenkis & Strautmane, 2010; Novak & Cañas, 2006; Patel et al., 2024; Strautmane, 2012) that depict relationships between concepts (McClure et al., 1999; Sun, 2004), attributes (Besterfield-Sacre et al., 2004), and ideas (Coffey et al., 2006; Cornwell, 1996; Kinchin et al., 2019), offering a comprehensive framework for understanding and analysing mental models (Chang, 2007; Moon et al., 2018). These maps can have different words in the vertices considered as factual data (Henry, 1974); and having different words (or just lines) connecting these vertices which called edges, considered as information (flow of organized data) (Gamble, Paul R., 2001). Thus, creating a complex system of information flow giving meaning and purpose to the original intent and mental model of participants, finally to be considered as their knowledge (Dewey, John, 1949). This approach facilitates the visualization of complex systems (Khajeloo & Siegel, 2022; Tergan et al., 2006), enabling users to map out how various elements interact and relate to one another (Schroeder et al., 2018). Concept maps are usually administered with a purpose (Torre et al., 2007) and accompanies several experiments (Gurupur et al., 2015; Markham et al., 1994) and studies (Rosas & Ridings, 2017; M Araceli Ruiz-Primo, 2004; van Bon-Martens et al., 2014) to help the researchers understand their participants and their thought processes. Hence it is a complex process-like engineering design which is often termed as a complex-socio-technical process (Milne & Leifer, 1999). During such complex-socio-technical processes, there are several variables and uncertainties that present itself as diverse challenges to be coded accurately (Lachner et al., 2018; Yin et al., 2005) and consistently (Schau et al., 2001). The problem is compounded when thousands of unique concept maps are to be coded by multiple researchers from diverse backgrounds. This need for a robust coding scheme also warrants a digital concept map tool that can automate the coding process and deliver further analysis of concept maps for research purposes.

Concept maps are complex tools to use and analyse (Tergan et al., 2006; van Bon-Martens et al., 2014). Since the research was currently dealing with hand-drawn concept maps by the participants, each participant was found to have their own unique way of representation, writing, drawing, connecting and expressing. Prior work has established that while the same topic is presented to participants, they can

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

create different concept maps akin to their understanding (Patel et al., 2024). This highlights that concept maps are like fingerprints and can be unique from person to person (Conceição et al., 2017; Jackson & Trochim, 2002; Lopez et al., 2014). This can also be used to represent a particular person's thought process at a particular time as well and hence it can be said that concept maps from a single person can also can evolve (D'Antoni et al., 2009; Maria Araceli Ruiz-Primo et al., 2001; van Bon-Martens et al., 2014; Van Zele et al., 2004).

The goal of this research is to use concept maps by the same participant over time as a simulated longitudinal study, which can reveal trends that can be used to evaluate the participants journey in a program or during the course of skills training or series of focussed interventions such as a semester in the context of an undergraduate engineering degree program (Reichherzer & Leake, 2006; Roberts & Johnson, 2015; Tergan et al., 2006; Zanting et al., 2003). These concept maps from the individual participants can be measured and analysed to understand the progress or evolution of the participant (Rosas & Ridings, 2017; Stoyanov et al., 2017; Trochim, 1989; Zanting et al., 2003). These concepts maps can provide the required insight into the participants understanding of what is needed for their purpose (Rosas & Ridings, 2017; Trochim, 1989). The concept maps can also be used as an augmented tool to support student evaluations beyond mastery assessment measured through grades (Jackson & Trochim, 2002; Lopez et al., 2014; Maker & Zimmerman, 2020; Talbert et al., 2020).

2. Study design and deploymemt

This section presents a detailed overview of the study design and deployment process used in this research. It outlines the instruments deployed, the target sample size, and the administration process. The subsequent section elaborates on the coding scheme deployment, describing the collaborative efforts with a team of researchers to validate its implementation.

2.1. Study design

The study deployed concept maps as a part of a larger study that also collected surveys from participants. This study is designed to measure the engineering identity of undergraduate engineering students at R1 university in the suburban area of a metropolitan city in the U.S. Two instruments are deployed in classes twice, at the beginning and the end of a semester, to pre-service engineers enrolled in classes in the first, second, third, and fourth year of the engineering curriculum. The first instrument is a survey to capture the student's engineering identity and the various influencers that affect their identity (Kumar & Summers, 2024). The instrument uses standard engineering identity items. Additionally, the survey asks the pre-service engineers to define "engineer". This definition is augmented by a concept map of the pre-service engineer's understanding of "engineer", the second research instrument. The instruments have been deployed in Fall 2024 with approximately 5,000 responses collected. The snippet of concept map template shown in Figure 1, highlights the canvas on which the in-class activity was held in person for a duration of six minutes after the engineering students received an overview about concept maps.

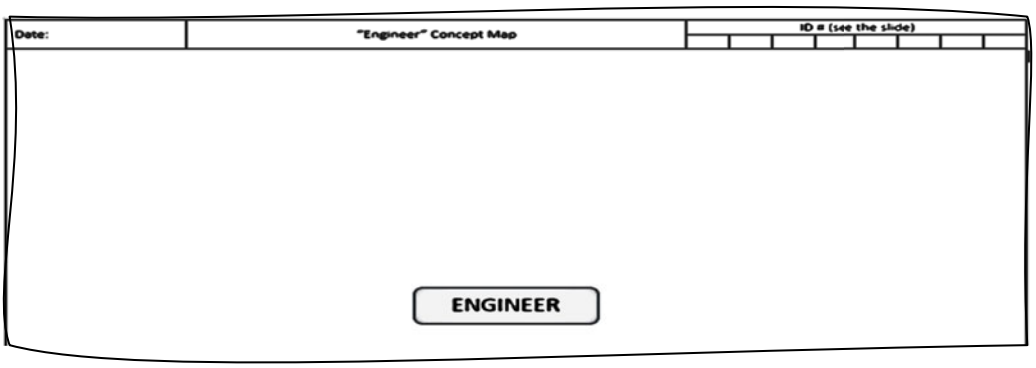


Figure 1. Concept map template

The concept maps are collected during an in-class activity. As a result, they are all hand drawn. It is these hand drawn concept maps that need to be computationally measured for their topologic complexity (J. L. Mathieson & Summers, 2010; J Mathieson & Summers, 2009). As a critical first step,

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

these concept maps are reconstructed into bi-partite graphs following a systematic coding approach. The students were instructed to then create a concept map to define "engineer", with additional prompts that include: what does an engineer do? What should engineers focus on? What do you anticipate as the future for engineers? What does it mean for an engineer to be successful? and What skills or characteristics does an engineer possess? They were provided a bounded sheet with "engineer" as a concept node at the centre of the page. The time limit of six minutes was calibrated in pilot studies to ensure that there was sufficient time for students to build a concept map of sufficient complexity to be of interest (Patel et al., 2024).

2.2. Coding scheme deployment process

The concept map coding scheme established in prior work is used to code and capture the mental model of participants expressed through their concept maps finally onto the computational format. The output from this process needs to be further processed computationally. This method enables researchers to convert the tool from one format (Concept Maps-Visual Representation Models) to another format (Computational Models) using a guidebook and training manual. Hence, it is important to present all the aspects of the coding scheme for a thorough investigation to capture all potential complexities that arise through type, scenarios and cases, which enables an early mitigation or failure-modes strategy to avoid future pitfalls. To validate the objective robustness of this coding scheme, 22 undergraduate researchers were presented six concept maps and asked to translate them into the spreadsheet format to capture the bi-partite graphs (concepts are nodes, connecting edges are relations). These bipartite graphs are then analysed for the structural complexity of the graphs using a 29 item complexity metric (J. L. Mathieson & Summers, 2010; Owensby & Summers, 2014). The specific terms included in the concept maps are captured and analysed through a separate sematic coding approach.

3. Use case and validation

The method developed for the analysis of the data collected for this research was given to 22 undergraduate researchers who had not worked on concept maps or its analysis before. The guidebook, training manuals, videos and other hands-on training were provided for a duration of two weeks. The coders were tested once after the training period and once again after a couple more weeks to check if more time and effort on working on several concept maps would increase their understanding of the coding scheme and the process. All researchers were given the same material, information and time duration for the two interrater reliability (IRR) tests (D'Antoni et al., 2009; Rye & Rubba, 2002; Watson et al., 2016; West et al., 2002) conducted to check the validity of the method and tools of the coding scheme of concept maps. The consistency of interpretation and reconstruction of the concept graphs needs to be sufficient to allow for parallel processing of the concept maps without a loss of information. Thus, inter rater reliability and intra rater reliability tests are conducted to verify the consistency. As the concept maps could have multiple concepts within a single node, it is important to be able to split these consistently. Tests are runs on the objectivity of the code for vertex splitting and edge construction. Six concept maps (P1-AX05DA00, P2-AS22BL16, P3-YA25RE54, P4-GE04CO71, P5-ON25JE07, and P6-ES18TR44) are randomly selected for testing and two maps are presented in Figure 2 and Figure 3. P1 in Figure 3 is rated as a high complexity concept map of all the six maps.

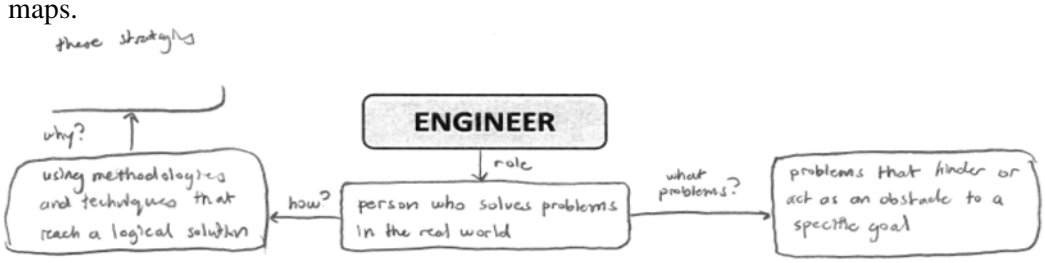


Figure 2. "6" Original concept map original (complete but template cropped)

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

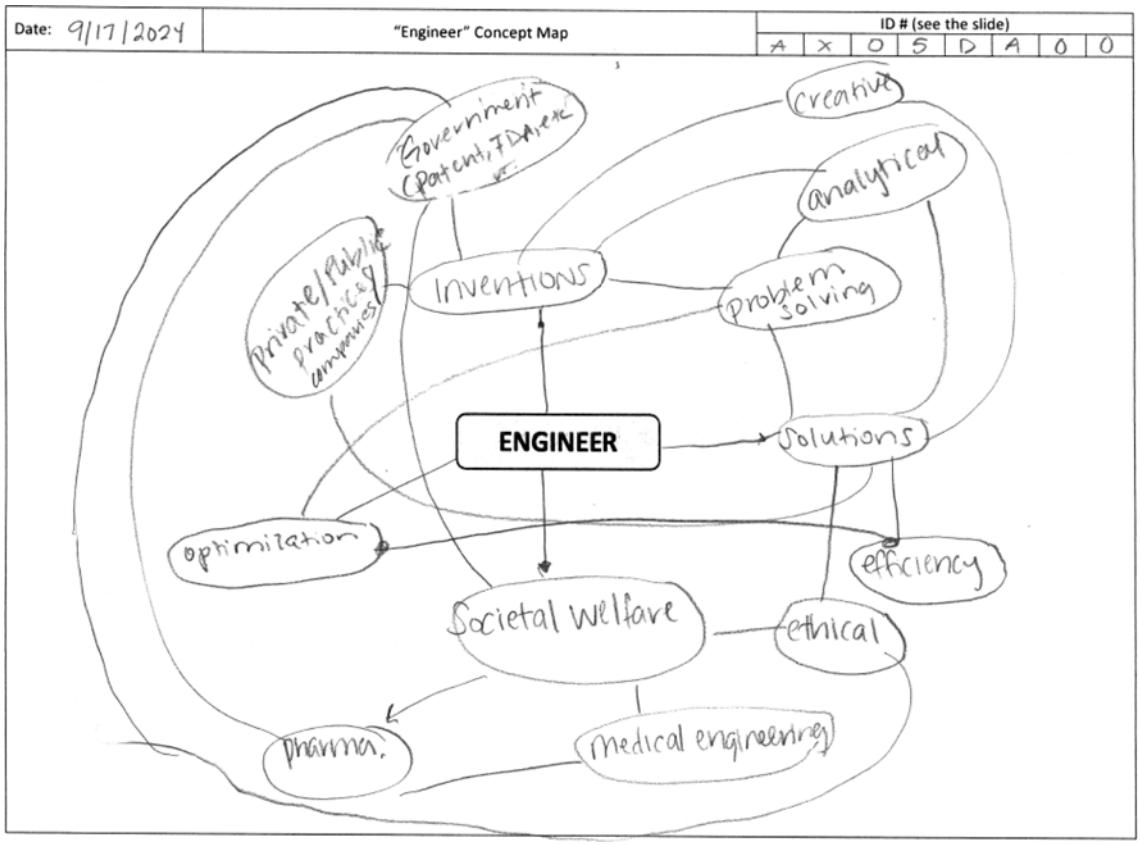


Figure 3. "1" Original concept map

3.1. Validation of coding scheme

This section highlights the analysis of the coding scheme with respect to the vertices and the edges. Three concept maps were selected from the first test (P1, P2, and P3) from a set of seven total concept maps and three concept maps selected from the second test (P4, P5, and P6) from a total of fifteen total concept maps. These 6 concept maps were chosen randomly using the random excel function. The results of the analysis are presented in this section using the averages of the vertices and edges coded by the participants using the coding scheme for all these six concept maps as part of this use case. The analysis was done to find the standard deviation and coding consistency amongst the 22 coders looking specifically the number of vertices and edges chosen to be coded by them by applying the specific types, scenarios and cases in the coding scheme.

3.1.1. Analysis of vertices

Figure 4 highlights the standard deviation on the average number of vertices. Here one can see that P1 (Figure 2) and P6 (Figure 3) were the most complex in nature and hence they had the highest deviation compared to the other four with P5 being the least.

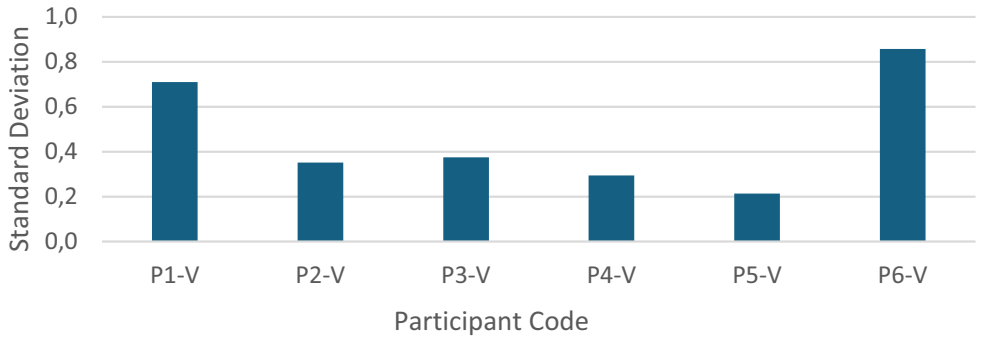


Figure 4. Standard deviation of average of vertices for various concept maps

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

It shows that except for P6, the rest of the maps were coded to be around the same number of vertices. From Figure 5 one can see that except for G, T, and U, rest of the coders were all in the tight grouping of the histogram. The 22 coders are represented as letters from A to V sorted in an alphabetical order of their first name. Coders G, T, and U on an average added an additional vertex, while the remaining nineteen coders did not. The nineteen coders were approximately 0.8 vertex between each other, thus showing that they were almost around the same number of vertices when applying the coding scheme. There is a slight positive slope with the trendline for the 22 coders presented in this format.



Figure 5. Coding consistency of various coders through average of vertices

3.1.2. Analysis of edges

Like the analysis of the vertices, Figure 6 shows that edges on an average were the more difficult element on the concept maps to code and hence there is a larger magnitude of standard deviation recorded in their coding. The largest deviation is P2 which was coded with a deviation of 1.6 edges compared to its counterparts. Figure 7 shows the coding consistency related to edges for the 22 coders and from the trendline there is no relevance or order to this coding consistency with respect to the coders.

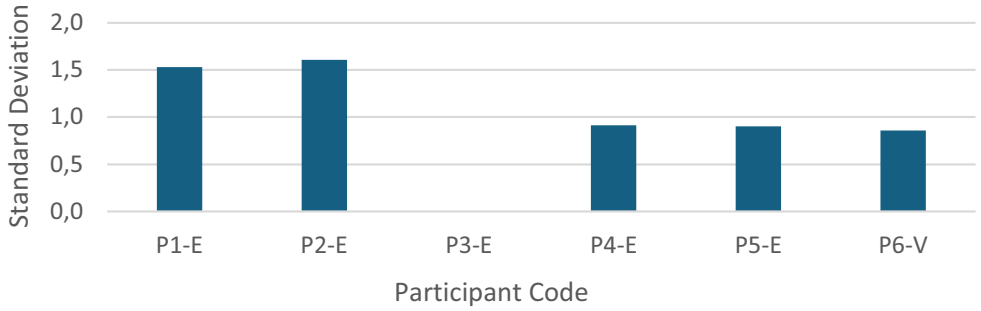


Figure 6. Standard Deviation of Average of Edges for Various Concept Maps



Figure 7. Coding consistency of various coders through average of edges

Another interesting observation in Figure 6 here would be that the set two of concept maps taken for this analysis reveals that the standard deviation is the same for all the three concept maps (P4, P5, and P6) highlighting that the coders might have gotten used to the coding scheme and were able to be in a particular deviation range. This also highlights consistency of the coding scheme and the coders. It is also seen that P3 was zero deviation on the coding of the edges. In Figure 7 we can wee that except for coder

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

M and U, the rest of the twenty coders are all consistent within an average of 15.67 to 16.50 showing a deviation less than one edge. Thus, highlighting that the coding scheme is consistent and sufficient.

3.2. Inter-Rater Reliability (IRR) - Fleiss' Kappa

The final validation of the coding scheme developed for concept maps can be shown with IRR testing. In this section we focus the analysis, validation and testing the hypotheses that the coding scheme is robust with reasonable agreement between the coders, and this is not based on chance as suggested by the IRR's null hypothesis. The testing includes IRR for coding the vertices and edges. Also, as this research had 22 coders, the research will be using Fleiss' Kappa (FKappa) (Falotico & Quatto, 2015; Moons & Vandervieren, 2023) which is a statistical measure used to evaluate the agreement between multiple coders (more than two) when assigning categorical codings to a fixed set of items. It is an extension of Cohen's Kappa. FKappa is the main statistic that shows the agreement between coders and will have a value between zero and one, where zeo would be no agreement and one would be in complete agreement eliminating the probability of agreement by chance and other bias. FKappa is widely used in research fields where IRR among multiple observers or evaluators needs to be quantified, such as healthcare, education, and social sciences.

It is also noted that unlike ANOVA or Z-test, the p-value is preferred to be zero in FKappa (or very close to zero), indicating that the null hypothesis (which assumes agreement by chance) can be rejected with very high confidence. In other words, the observed level of agreement among coders is statistically significant and not due to random chance. The research also acknowledges that a p-value of zero doesn't necessarily mean perfect agreement; it just means that the observed agreement is statistically significant. The strength of the agreement is still determined by the kappa value itself. If kappa value is between 0.75 and One, it means excellent agreement. And between 0.4 to 0.75 it is moderate to substantial agreement and Zero to 0.4 would be slight to fair agreement. Thus, a p-value of Zero means the level of agreement among coders is real and not due to chance, but the extent of this agreement still depends on the Kappa value.

3.2.1. Vertices IRR

This section presents the IRR related to vertices to be checked between all coders on the vertices of the concept maps. The vertices are the nodes that are presented in the concept maps that need to be coded in the excel file as close to the original concept map, but also ensuring accuracy and consistency by applying the rules in the guidebook and the training manual. With the use of FKappa it can be seen that from Table 1 and Table 2, FKappa value for the coding of vertices is 0.6787, thus showing that most of the coders are in substantial agreement with each other, hence proving the effectiveness of the coding scheme and process (method and tool) for vertices.

Table 1. FKappa IRR calculation for coding of vertices

		Number of Vertices Chosen to be Coded by Coders as Per Coding Scheme										
		4	5	6	7	13	14	15	16	17	26	27
Participant Code	***P1***					1	19	1		1		
	P2							3	19			
	P3							1	19	2		
	P4					20	2					
	P5										1	21
	P6	1	14	3	4							

It can be observed from the coding as well that P1 and P6 can be seen to have more spread of values. If we consider P1 in the first row, we see that out of the 22 coders, 1 coded for 13 vertices, 19 coders coded 14 vertices, 1 coder coded it to have fifteen vertices and one more coder coded it to have seventeen vertices. Thus, showing that each row adding up to 22 coders to complete the calculation section of this table. Table 2 highlights several final parameters pertaining to Fkappa test, such as m - number of coders/ coders, n- number of concept maps selected for the IRR test, pa - the average proportion of coders that agree for each item, pe - the expected agreement by chance, s.e - standard error which measures the

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

variability of the Kappa statistic, z - a statistic that tests the null hypothesis, p - the probability of observing the computed z-value under the null hypothesis and it has to be less than 'alpha' - which is the significance level for confidence intervals, often set to 0.05, and finally the lower and upper which are the bounds of the confidence interval.

Table 2. FKappa final IRR results for coding of vertices

Important Final FKappa (IRR) Parameters for Vertices										
m	**n**	**pa**	**pe**	**kappa**	**s.e**	**z**	**p-value**	**alpha**	**lower**	**upper**
22	6	0.7345	0.1736	0.6787	0.0117	58.0980	0	0.05	0.6558	0.7016

3.2.2. Edges IRR

Like the IRR of vertices, the research obtained a Fkappa value of 0.4562 with an alpha of 0.05 and other important values pertaining to the FKappa test as shown in Table 3. Thus, indicating that there was substantial agreement on the coding scheme applied to the edges as well. The research acknowledges that as the edges are the elements that create the complexity in concept maps which ultimately understands the flow of information by connecting the facts or data points. These edges also get further complicated when the writing utensil or the scan of the concept maps have issues. Thus, justifying the need for a digital tool that can eliminate such human-induced errors. Again, this Fkappa value for coding of edges according to the process and tool, proves and validates the coding scheme developed for concept maps.

Table 3. FKappa final IRR results for coding of edges

Important Final FKappa (IRR) Parameters for Edges										
m	**n**	**pa**	**pe**	**kappa**	**s.e**	**z**	**p-value**	**alpha**	**lower**	**upper**
22	6	0.5094	0.0978	0.4562	0.0085	53.3696	0	0.05	0.4394	0.4730

4. Conclusion and limitations

The final analysis and IRR validation demonstrates the effectiveness and the consistency of the coding scheme. An IRR with Fleiss' Kappa value of 0.67 for vertices and 0.45 for edges shows that there is substantial agreement about the use of the coding scheme to effectively code complex concept maps that were hand-drawn. This paper presents an evidence-based approach to validate the coding scheme for concept maps. The coding scheme had also deployed another set of parameters from the concept maps to be coded as part of the 'Summary' that can help researchers to quickly bracket the concept maps into excellent, good, moderate, and bad. This can also aid in the further analysis of the maps and doubles itself as a filter to catch maps that may be overly complex. This classification could also be useful to connect with the larger research connected with these concept maps that is being investigated as a surrogate to measuring Engineering Identity.

The important identity metrics from the survey and the definitions (also obtained from the survey) can be mapped against the data, information, and knowledge from these concept maps, thus completing the loop of using concept maps as an augmented tool for engineering identity. Furthermore, inspection on excellent concept maps can also help us understand why it may have scored high and what it relates to in the survey. 'Is this excellent map from a high performing student' or 'moderate' or 'low'. Thus, giving the evaluators an opportunity to dive deep into the root-cause analysis and provide the required support and systems for improvement for the student.

The use of such concept maps has been established by earlier research as a tool to measure learning and to be used as an augmented evaluation or assessment tool in academia. This research establishes the validity of the coding scheme barring which the concept maps are prone to coding errors as the complexity of concept maps can be compounded when it is coded incorrectly by multiple researchers. Concept maps can now be processed with an acceptable confidence level on the research, hence aiding the downstream analysis methods, such as the 29-complexity metrics that provides the topology analysis, vocabulary/keywords providing the sentiment and trends, and finally comprehensive analysis to subsequently rate or score the concept maps for meaning and purpose.

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

The results from these various types of analysis on the concept maps reveal micro and macro trends, which can then guide decision making for its stakeholders by creating a science for cohorts that can be relied upon. This can effectively form the basis on cohort science that can help investigate effectiveness and impact of programs and interventions. The established concept map coding scheme can be now used (with some further refinement) at scale, to process the larger data set of thousands of concept maps. The analysis from these maps can now augment the original study looking at undergraduate student's engineering identity. Future work can also investigate the use of concept maps as an ideal replacement candidate or surrogate for the objective measurement of engineering identity; thus, eliminating the subjective nature of self-efficacy-based measurement as established by prior work.

The absolute value of using concept maps and can be extracted when we are clearly able to extract the data, information and knowledge trails from participants and map them longitudinally. This can enable a wide variety of use cases for concept maps beyond just a surrogate for evaluation. This can further be used in conjunction with artificial neural networks or other AI tools to also create a predictive tool, method, and practice to understand existing trends and map future trends that can help in planning and resource allocation to meet the projected demands or forecast. Measuring recall, retention, evaluation and projection which are important metrics for several sectors and domains, thus making concept maps a potential tool for measurement and design.

References

Besterfield-Sacre, M. E., Gerchak, J., Lyons, M., Shuman, L. J., & Wolfe, H. (2004). Scoring concept maps: An integrated. *Journal of Engineering Education*, April, 105–115.

Chang, S. N. (2007). Externalising students' mental models through concept maps. *Journal of Biological Education*, 41(3), 107–112. https://doi.org/10.1080/00219266.2007.9656078

Coffey, J. W., Hoffman, R., & Cañas, A. (2006). Concept map-based knowledge modeling: Perspectives from information and knowledge visualization. *Information Visualization*, 5(3), 192–201. https://doi.org/10.1057/palgrave.ivs.9500129

Conceição, S. C. O., Samuel, A., & Yelich Biniecki, S. M. (2017). Using concept mapping as a tool for conducting research: An analysis of three approaches. *Cogent Social Sciences*, 3(1). https://doi.org/10.1080/23311886.2017.1404753

Cornwell, P. J. (1996). Concept maps in the mechanical engineering curriculum. *ASEE Annual Conference Proceedings*, 835–841. https://doi.org/10.18260/1-2–5933

D'Antoni, A. V., Zipp, G. P., & Olson, V. G. (2009). Interrater reliability of the mind map assessment rubric in a cohort of medical students. *BMC Medical Education*, 9(1), 1–8. https://doi.org/10.1186/1472-6920-9-19

Dewey, J., & Bentley, A. F. (1949). Knowing and the known. Boston: Beacon Press.

Falotico, R., & Quatto, P. (2015). Fleiss' kappa statistic without paradoxes. *Quality and Quantity*, 49(2), 463–470. https://doi.org/10.1007/s11135-014-0003-1

Gamble, P. R., & Blackwell, J. (2001). Knowledge management: A state of the art guide.

Grundspenkis, J., & Strautmane, M. (2010). Usage of graph patterns for knowledge assessment based on concept maps. *Scientific Journal of Riga Technical University. Computer Sciences*, 38(38), 60–71. https://doi.org/10.2478/v10143-009-0005-y

Gurupur, V. P., Pankaj Jain, G., & Rudraraju, R. (2015). Evaluating student learning using concept maps and Markov chains. *Expert Systems with Applications*, 42(7), 3306–3314. https://doi.org/10.1016/j.eswa.2014.12.016

Henry, N. L. (1974). Knowledge management: A new concern for public administration. *Public Administration Review*, 34(3), 189. https://doi.org/10.2307/974902

Jackson, K. M., & Trochim, W. M. K. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods*, 5(4), 307–336. https://doi.org/10.1177/109442802237114

Khajeloo, M., & Siegel, M. A. (2022). Concept map as a tool to assess and enhance students' system thinking skills. *Instructional Science*, 50(4), 571–597. https://doi.org/10.1007/s11251-022-09586-5

Kinchin, I. M., Möllits, A., & Reiska, P. (2019). Uncovering types of knowledge in concept maps. *Education Sciences*, 9(2). https://doi.org/10.3390/educsci9020131

Kumar, P., & Summers, J. D. (2024). Refining engineering identity framework for implementation in engineering education – A literature-based analysis. Volume 4: *21st International Conference on Design Education (DEC)*. https://doi.org/10.1115/DETC2024-143188

Lachner, A., Backfisch, I., & Nückles, M. (2018). Does the accuracy matter? Accurate concept map feedback helps students improve the cohesion of their explanations. *Educational Technology Research and Development*, 66(5), 1051–1067. https://doi.org/10.1007/s11423-018-9571-4

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

Lopez, E. J., Shavelson, R. J., Nandagopal, K., Szu, E., & Penn, J. (2014). Ethnically diverse students' knowledge structures in first-semester organic chemistry. *Journal of Research in Science Teaching*, 51(6), 741–758. https://doi.org/10.1002/tea.21160
Maker, C. J., & Zimmerman, R. H. (2020). Concept maps as assessments of expertise: Understanding of the complexity and interrelationships of concepts in science. *Journal of Advanced Academics*, 31(3), 254–297. https://doi.org/10.1177/1932202X20921770
Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching*, 31(1), 91–101. https://doi.org/10.1002/tea.36603101091
Mathieson, J. L., & Summers, J. D. (2010). Complexity metrics for directional node-link system representations: Theory and applications. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 44137, 13–24.
Mathieson, J., & Summers, J. (2009). Relational DSMs in connectivity complexity measurement. *Proceedings of 11th International DSM Conference*, 15–26.
Mathieson, J., & Summers, J. D. (2017). A protocol for modeling and tracking engineering design process through structural complexity metrics applied against communication networks. *Concurrent Engineering Research and Applications*, 25(2), 108–122. https://doi.org/10.1177/1063293X16666936
McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475–492. https://doi.org/10.1002/(SICI)1098-2736(199904)36:4<475::AID-TEA5>
Milne, A., & Leifer, L. (1999). The ecology of innovation in engineering design. *International Conference on Engineering Design*.
Moon, B., Johnston, C., & Moon, S. (2018). A case for the superiority of concept mapping-based assessments for assessing mental models. *International Conference on Concept Mapping*, 2008, 1–10.
Moons, F., & Vandervieren, E. (2023). Measuring agreement among several coders classifying subjects into one-or-more (hierarchical) nominal categories: A generalisation of Fleiss' kappa. *arXiv*.
Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them. *Technical Report IHMC CmapTools 2006-01*, 1–31.
Owensby, J. E., & Summers, J. D. (2014). Assembly time estimation: Assembly mate based structural complexity metric predictive modeling. *Journal of Computing and Information Science in Engineering*, 14(1).
Patel, A., Summers, J. D., Kumar, P. P., & Edwards, S. L. (2024). Investigating the use of concept maps and graph-based analysis to evaluate learning. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://doi.org/10.18260/1-2–47702
Reichherzer, T., & Leake, D. (2006). Understanding the role of structure in concept maps. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2004–2009.
Roberts, R., & Johnson, P. (2015). Understanding the quality of data: A concept map for 'the thinking behind the doing' in scientific practice. *Curriculum Journal*, 26(3), 345–369. https://doi.org/10.1080/09585176.2015.1044459
Rosas, S. R., & Ridings, J. W. (2017). The use of concept mapping in measurement development and evaluation: Application and future directions. *Evaluation and Program Planning*, 60, 265–276. https://doi.org/10.1016/j.evalprogplan.2016.08.016
Ruiz-Primo, M. A. (2004). Examining concept maps as an assessment tool. Concept Maps: Theory, Methodology, Technology. *Proc. of the First Int. Conference on Concept Mapping*, 1, 555–563.
Ruiz-Primo, M. A., Shavelson, R. J., Li, M., & Schultz, S. E. (2001). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *International Journal of Phytoremediation*, 21(1), 99–141. https://doi.org/10.1207/S15326977EA0702_2
Rye, J. A., & Rubba, P. A. (2002). Scoring concept maps: An expert map-based scheme weighted for relationships. *School Science and Mathematics*, 102(1), 33–44. https://doi.org/10.1111/j.1949-8594.2002.tb18194.x
Schau, C., Mattern, N., Zeilik, M., Teague, K. W., & Weber, R. J. (2001). Select-and-fill-in concept map scores as a measure of students' connected understanding of science. *Educational and Psychological Measurement*, 61(1), 136–158. https://doi.org/10.1177/00131640121971112
Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review*, 30(2), 431–455. https://doi.org/10.1007/s10648-017-9403-9
Stoyanov, S., Jablokow, K., Rosas, S. R., Wopereis, I. G. J. H., & Kirschner, P. A. (2017). Concept mapping—An effective method for identifying diversity and congruity in cognitive style. *Evaluation and Program Planning*, 60, 238–244. https://doi.org/10.1016/j.evalprogplan.2016.08.015
Strautmane, M. (2012). Concept map-based knowledge assessment tasks and their scoring criteria: An overview. *Concept Maps: Theory, Methodology, Technology. Proc. of the Fifth Int. Conference on Concept Mapping*, 1, 80–88.
Summers, J. D., & Shah, J. J. (2010). Mechanical engineering design complexity metrics: Size, coupling, and solvability. *Journal of Mechanical Design*, 132(2).

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press

Sun, Y. (2004). Methods for automated concept mapping between medical databases. *Journal of Biomedical Informatics*, 37(3), 162–178. https://doi.org/10.1016/j.jbi.2004.03.003

Talbert, L. E., Bonner, J., Mortezaei, K., Guregyan, C., Henbest, G., & Eichler, J. F. (2020). Revisiting the use of concept maps in a large enrollment general chemistry course: Implementation and assessment. *Chemistry Education Research and Practice*, 21(1), 37–50. https://doi.org/10.1039/c9rp00059c

Tergan, S. O., Keller, T., & Burkhard, R. A. (2006). Integrating knowledge and information: Digital concept maps as a bridging technology. *Information Visualization*, 5(3), 167–174. https://doi.org/10.1057/palgrave.ivs.9500132

Torre, D. M., Daley, B., Stark-Schweitzer, T., Siddartha, S., Petkova, J., & Ziebert, M. (2007). A qualitative evaluation of medical student learning with concept maps. *Medical Teacher*, 29(9–10), 949–955. https://doi.org/10.1080/01421590701689506

Trochim, W. M. K. (1989). Concept mapping: Soft science or hard art? *Evaluation and Program Planning*, 12(1), 87–110. https://doi.org/10.1016/0149-7189(89)90027-X

van Bon-Martens, M. J. H., van de Goor, L. A. M., Holsappel, J. C., Kuunders, T. J. M., Jacobs-van der Bruggen, M. A. M., te Brake, J. H. M., & Van Oers, J. A. M. (2014). Concept mapping as a promising method to bring practice into science. *Public Health*, 128(6), 504–514. https://doi.org/10.1016/j.puhe.2014.04.002

Van Zele, E., Lenaerts, J., & Wieme, W. (2004). Improving the usefulness of concept maps as a research tool for science education. *International Journal of Science Education*, 26(9), 1043–1064. https://doi.org/10.1080/1468181032000158336

Watson, M. K., Pelkey, J., Noyes, C. R., & Rodgers, M. O. (2016). Assessing conceptual knowledge using three concept map scoring methods. *Journal of Engineering Education*, 105(1), 118–146. https://doi.org/10.1002/jee.20111

West, D. C., Park, J. K., Pomeroy, J. R., & Sandoval, J. (2002). Concept mapping assessment in medical education: A comparison of two scoring systems. *Medical Education*, 36(9), 820–826. https://doi.org/10.1046/j.1365-2923.2002.01292.x

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2), 166–184. https://doi.org/10.1002/tea.20049

Zanting, A., Verloop, N., & Vermunt, J. D. (2003). Using interviews and concept maps to access mentor teachers' practical knowledge. *Higher Education*, 46(2), 195–214. https://doi.org/10.1023/A:1024719816657

https://doi.org/10.1017/pds.2025.10036 Published online by Cambridge University Press