ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #37708

Analyzing Various Scoring Methods for Fill-In Concept Maps

Ethan Cartwright (Undergraduate Student)

Undergraduate aerospace engineering student at Ohio State.

Meagan Eleanor Ita (Post-Doctoral Scholar)

Meagan is a Postdoctoral Scholar in the Department of Engineering Education at The Ohio State University. She graduated with her Bachelor's (2013) and Master's (2014) in Biomedical Engineering from The Ohio State University and completed her Ph.D. in Bioengineering at the University of Pennsylvania in 2020. Meagan conducted research on the biomechanics and physiology of chronic joint pain for her doctorate degree and has experience teaching undergraduate first-year engineering and mid-level biomechanics courses. Meagan is currently working with the Kern Entrepreneurial Engineering Network partnership at OSU, integrating her interests in STEM education, entrepreneurial partnerships, and community engagement. Long term, Meagan is passionate about leveraging biotechnological advances in physiological data collection to both 1. promote factual science communication and learning and 2. to innovate community-driven engineering solutions centered around health and wellness. She values authenticity, connection with others, & integrity and prioritizes these values as an educator, bioengineer, and scientist.

Krista M Kecskemety (Assistant Professor of Practice)

Krista Kecskemety is an Associate Professor of Practice in the Department of Engineering Education at The Ohio State University and the Director of the Fundamentals of Engineering for Honors Program. Krista received her B.S. in Aerospace Engineering at The Ohio State University in 2006 and received her M.S. from Ohio State in 2007. In 2012, Krista completed her Ph.D. in Aerospace Engineering at Ohio State. Her engineering education research interests include investigating first-year engineering student experiences, faculty experiences, and the research to practice cycle within first-year engineering.

Analyzing Various Scoring Methods for Fill-In Concept Maps

Ethan Cartwright, Meagan E. Ita, Krista M. Kecskemety

Abstract

Concept mapping is an instructional tool that has shown promise in assessing conceptual understanding, especially with deeply interconnected topics and those focused on connections, like that of the Entrepreneurial Mindset. An important step in implementing concept map assignments for assessing student subject knowledge in a classroom is scoring, of which there are many methods that have been developed. However, there are few scoring methods that have been developed for fill-in concept maps, a format where students are given the structure of the concept map along with a bank of words with which to fill it in. Fill-in concept mapping is a potentially attractive variant of concept mapping for its reduced load on both students and instructional teams, particularly in the case of large and/or low-resource settings, but has seen less extensive research efforts. This study uses adaptations to the three prominent unstructured concept map scoring methods to fit the fill-in format and compares scoring results across the methods. No statistical difference is found between each of the adapted methods, potentially indicating assessment of a similar skill, though best-use application of each method may differ. Further qualitative analysis on the implementation and scoring of the various methods is discussed. The results and the proposed alterations may be useful in further development of scoring methods for fill-in concept maps which can adequately assess conceptual understanding.

Introduction

The Kern Entrepreneurial Engineering Network (KEEN) [1] is a partnership of universities across the United States that is dedicated to infusing the Entrepreneurial Mindset (EM) into undergraduate engineering curriculum. EM is a multidisciplinary philosophy that involves problem solving, value creation, innovation, and decision making [2,3]. The Entrepreneurial Mindset is a mental framework predicated on an inclination for discovery, value creation, and seeking opportunity [4]. In particular to KEEN, EM is focused on the 3 Cs of curiosity, connections, and creating value [1].

In a larger KEEN sponsored study across multiple institutions concerned with EM infusion in undergraduate education, concept map assessments of various types have been implemented to assess student understanding of EM throughout their undergraduate experience, and the growth therein. Concept mapping was used as a tool in this setting for its ability to assess conceptual understanding and connections within complex topics like that of EM [5,6]. In addition to concept mapping being an important metacognition activity for students studying complex topics, concept mapping also directly forces students to make connections between topics which is a key component of the Entrepreneurial Mindset. As a subset of this study, fill-in concept maps have been used at a large midwestern university [7]. A fill-in concept map provides the structure of the map along with a word bank of concepts that the student is expected to use to fill it in. Along the progression of this study design, the applicability of the three prominent scoring methods for fill-in maps came into question. The three prominent scoring methods have been shown to successfully assess conceptual understanding in a given topic [8], but are formulated for concept maps that are made without a provided structure or concepts. In validating the applicability of fill-in concept mapping as an assessment tool, other studies typically use a correct/not correct scoring scale in which a score is given based on how many concepts correctly match that of the primary, complete fill-in map developed by the instructor(s) or researcher(s) [9,10].

Fill-in concept maps may be more appealing to instructors due to a perceived decrease in cognitive load for the student and workload for both the student and instructional team, but little research has been conducted looking into how a fill-in concept map should best be assessed. This lack of research prompted an adaptation of the prominent unstructured concept map scoring methods to a fill-in map. Very little analysis has been done as to how a fill-in concept map could be most adequately assessed, and whether different methods may be more applicable in various circumstances. Therefore, the primary goal of this study was to analyze these adaptations of the prominent concept map scoring methods quantitatively to see what trends arise within and across methods, and qualitatively to discuss the usage of the various methods. Although the concept maps used to accomplish this goal are centered on the topic of EM and have applications to the EM attribute of "making connections" [1], the central focus of this study is how fill-in concept map scoring methods differ in their application to EM and more broadly.

Background

A. Concept Mapping

Conceptual understanding of complex topics is predicated on recognition of interconnection between concepts within that topic, more so than the ability to recall factual information [6,8]. This is due to the importance of relationships between concepts in these topics. Typical multiple choice assessment tools fail to capture this more intricate system of connection, so many methods have been proposed to reconcile that gap [11]. Concept mapping is an assessment tool that can adequately address that failure. Concept mapping is a graphical representation of a central topic and the way related concepts connect to that central topic [8]. Concept mapping requires linking and describing connections between concepts within a central topic to demonstrate understanding, which applies more directly to how topics are structured [6]. Additionally, concept maps are generally hierarchical in nature and include links crossing hierarchies. This style of assessment is supported by semantic memory theory in that semantic memory is structured as a network of conceptual understandings and links between related concepts [11]. By assessing the structure of semantic memory, concept maps seek to better encapsulate the intricacies of conceptual understanding.

The most common presentation of a concept map assignment is without a provided structure. This means that the map creator dictates which concepts are included in the map, and how those concepts are connected because they provided with just a central topic. This is in contrast to a fill-in concept map that initially starts with little or no filled in concepts, but a provided structure and concept bank. There is an array of different methods for assessing how satisfactory a concept map encapsulates its central topic. The three most prominent methods to emerge for doing so are traditional, categorical, and holistic [8], but these three scoring methods were developed for scoring an unstructured map. While this study is concerned with fill-in concept maps, the scoring methods used are adapted from the three prominent scoring methods, and thus the usage and motivation behind each of the scoring methods is contextually important. The following describes the three prominent scoring methods for unstructured concept maps.

B. *Traditional Method*

The traditional scoring method is the most straightforward of the prominent methods. Its scoring is objective and combines three sub-scores to result in an overall score. These sub-scores assess the knowledge breadth, depth, and connectedness of a concept map, respectively. Breadth is assessed through the number of concepts present in the map, depth through the number of levels in the map—its hierarchy—and connectedness through the number of links across branches in the map. Various weightings can be given to the sub-scores to prioritize different aspects of the concept map's structure, but most prominent is the scoring developed by Novak and Gowin [12], seen in Eq. 1.

$$Score = NC + 5 * HH + 10 * NCL \tag{1}$$

NC is the number of concepts in the map, which excludes the central topic and any linking words—those not inside a bubble. *HH* is the highest hierarchy, or the highest number of concepts in a direct path from the central topic to an end-point concept. This cannot follow along a cross-link or feedback loop. *NCL* is the number of cross-links, or instances in which concepts from different branches are linked together. This particular set of weights by Novak and Gowin [13] has been shown to be highly successful in assessing expert-level understanding [14].

C. *Categorical Method*

The categorical scoring method, much like that of traditional, encapsulates the structure of a concept map but is more focused on connections between areas of the material present in the map. This method was originally developed by Segalas et al. [15] to assess sustainability concept maps, but has been successfully expanded to other topics. Of particular interest with this method is its ability to assess the complexity of a concept map. It assesses the complexity of the map through its ability to connect various categories of concepts together. The scoring formula is below in Eq. 2.

$$Complexity = NC * \frac{NIL}{N_{CAT}} \tag{2}$$

NC again is the number of concepts in the map, which excludes the central topic and any linking words. *NIL* is the number of interlinks present in the map—the number of instances in which concepts of different categories are linked together. N_{CAT} is the number of categories present in the map. Typically, categories come from learning outcomes in a topic or a comprehensive expert analysis of the topic. Unlike the traditional and holistic scoring methods, the categorical complexity has no summation of sub-scores. This is due to the method's focus on the connectedness of the map's structure. In some sense, the breadth and depth of the map could be said to be expressed through the categorical complexity, but nonetheless cannot be separately examined.

D. Holistic Method

The holistic scoring method attempts to be maximally generalizable by assessing the same parameters as the traditional method with the addition of correctness. This form of assessment is much more subjective than that of the traditional scoring technique because it has no set criteria in each of its three metrics. The particular calculation which has become most prominent was developed by Besterfield-Sacre et al. [16] using Eq. 3.

$$Score = Comprehensiveness + Organization + Correctness \qquad (3)$$

Comprehensiveness is a combination of the breadth and depth of the structure of a map. *Organization* is the level of integration between concepts on the map. *Correctness* assesses how often a map correctly connects concepts according to an expert scorer. These parameter descriptions are rules of thumb that have been adapted in various ways to alter how objective the method is and how specifically the parameters are assessed. Similar to the traditional method, a map with more concepts and longer chains of concepts will score higher in *Comprehensiveness*. A map with a webbing structure, and/or many interlinks will score higher in *Organization*.

Methods

A. Data Collection

Within this study, a fill-in concept map regarding EM was completed by first year engineering students at a large midwestern university after the completion of an introductory engineering course in fall 2020. This fill-in map was developed based off of an expert concept map [17]. The concept map was then administered to students near the end of the semester as a completion grade. This student population consisted of first-year pre-engineering students with little exposure to EM or concept mapping. The fill-in maps were administered through Cmap Tools [18], an electronic format, for ease of scoring.

The collected concept maps were then scored using each of the adapted scoring methods by two separate graders, according to general best scoring practices [19], and later compared for accuracy in scoring. First, the adapted traditional scoring method was used and discussion was held between scorers to sort out discrepancies in scoring. As will be discussed later, the adapted traditional scoring method should yield a result independent of scorer, so all maps were scored and compared without initial calibration, and discussion was held until there was perfect agreement in scoring result. The set of maps was scored using the adapted categorical method after an initial discussion between scorers on what categories the concepts of the primary concept map fall into and how the category a concept fits under may differ depending on the connections made to it. After this discussion, the maps were scored and discussion of each map’s score was held until there was perfect agreement. Similar to the adapted traditional scoring method, the adapted holistic scoring method should give a result independent of scorer, aside from the breadth component which was scored as a part of scoring the maps using the adapted categorical method. With this, no calibration was done, and all maps were scored and discussed by the two scorers until perfect agreement in scores was reached.

B. Primary Concept Map

To carry out the study, a concept map for the topic of interest was generated to use as both the model concept map for scoring, and as a template from which concepts were removed to construct the participant assignment. The topic of interest in this study, as discussed prior, is the Entrepreneurial Mindset. The concept map used draws from prior work by Bodnar et al. as a simplified version of that complete EM concept map [17].

C. Adapted Scoring Methods

The three prominent scoring methods used in unstructured maps were altered in a way that attempted to preserve the intent of the original method. The important aspects of these methods that were prioritized in their adaptations are as follows:

Traditional: Objectivity in and ease of scoring.

Categorical: Capturing connections between various aspects of a topic.

Holistic: Correctness in conjunction with concept map structure.

The traditional method translates poorly to the fill-in format due to it solely capturing of the structure of a map. For a set of fill-in maps, the structure is provided, so will be identical provided no concepts are empty. Because of this, the adapted method is quite different from the traditional method in terms of its specific attributes, but retains its intent. The adapted method is similar to a correct/not correct scoring scheme as it is the most objective and rigid structure. This adapted method is occasionally referred to as link similarity, but will be referred to as the adapted traditional method in this study much like the other adapted methods. Eq. 4 shows the scoring formula where *NML* is the number of matching links, *NTL* is the total number of links, and *NNL* is the number of non-matching links. A matching link is an instance in which two concepts were correctly connected together with a linking word when compared to the primary map. Since the maximum score is a map in which no non-matching links are present, the maximum score is 1. The total number of links in a map is 22 due to the maps being fill-in. This method is self-normalizing and results in a link similarity quotient.

$$Score = \frac{NML}{NTL+NNL} \quad (4)$$

The categorical scoring method was better fit to transfer to the fill-in format. Although it also captures the structure of a map, it does this in the context of the concepts that were placed by the participant completing the map. As in the typical categorical method, concepts were mapped to a category in which they fit, and points were awarded based on the number of instances that concepts of different categories were connected. Notably, scorers attempted to use context within the map to categorize concepts which allowed for identical concepts to fit into different categories in different maps. This method retains the same formula used for unstructured maps. Although the primary map is considered the “correct” map for the other methods, the categorical method doesn’t consider correctness. Eq. 2 is repeated here for clarity, showing the scoring formula, where *NC* is the number of concepts used, *NIL* is the number of interlinks, or instances of concepts of different categories being connected, and N_{CAT} is the number of categories present in the map.

$$Score = NC * \frac{NIL}{N_{CAT}} \tag{2}$$

The holistic method assesses the breadth, depth, organization and correctness of a map, some of which are transferable to a fill-in map. The correctness metric can be transferred with the usage of a primary map. The organization metric assesses the structure of the map, and was dropped because the structure is identical across maps. The adapted method uses the same three-point total for the remaining two metrics. A half point is removed from the correctness total for each instance in which a concept does not correctly connect to another concept when compared to the primary map. The comprehensiveness metric was given points through the categories that were assigned during the categorical scoring of the maps. This was done because an inclusion of more categories indicates that the map demonstrates a higher breadth of information. The breakdown of points awarded in the comprehensiveness metric can be found below in Table 1, along with the formula for the method generally. Eq. 5 shows the scoring formula, where *Comprehensiveness* is the sub score for the comprehensiveness metric, and *Correctness* is the sub score for the correctness metric.

Table 1: Comprehensiveness Sub score Breakdown by Number of Included Categories

Number of Categories	**Comprehensiveness Sub Score**
9	3
7 to 8	2.5
5 to 6	2
3 to 4	1.5
1 to 2	1

$$Score = Comprehensiveness + Correctness \tag{5}$$

D. Correlation between methods

Scores for all three adapted methods were normalized by maximum score (traditional: 1; holistic: 6; categorical: highest score achieved by any participant) to allow for comparisons across methods. The normality of normalized scores was assessed with a goodness-of-fit test on the residuals of all outcomes; all normalized scores were found to be non-normally distributed. To test the effect of the different scoring methods on student scores, matched pairs Wilcoxon signed rank comparisons (for non-normal distributions; $\alpha=0.05$) were performed between the different scoring methods on the normalized scores (traditional vs. holistic; traditional vs. categorical, and categorical vs. holistic).

From these normalized scores, the *differences* in normalized score were calculated for both categorical and holistic relative to their respective traditional score. This *difference* score was computed to quantitatively assess how a student performed with the categorical scoring compared to traditional and the holistic scoring compared to traditional. An example of this calculation for one student map is shown in Table 2. Traditional was used as the baseline for comparison because of its similarity to the correct/not correct scoring method which is prevalent in the literature [9,10].

Table 2: Sample Percentage Difference from Traditional. This example shows that this student received a higher percentage score with categorical than traditional (+9.53%) yet a lower percentage score with holistic than traditional (-7.14%).

traditional	categorical	holistic	categorical difference	holistic difference
57.14	66.67	50	9.53	-7.14

Results

Each of the scoring methods was recorded in terms of its component sub scores, the total based on the sub scores, and the converted percentile equivalent. Tables 3, 4, and 5 are a sample of these results for 5 concept maps.

Table 3: Traditional Scoring Method Sample Data

Participant	NML	NNL	Score	Percentage
1	22	0	1	100
2	17	5	0.62963	62.963
3	12	10	0.375	37.5
4	10	12	0.29412	29.412
5	10	12	0.29412	29.412

Table 4: Categorical Scoring Method Sample Data

Participant	NC	NIL	NCAT	Score	Percentage
1	17	7	9	13.222	58.333
2	15	8	9	13.333	58.824
3	17	7	9	13.222	58.333
4	17	9	9	17	75
5	15	8	8	15	66.176

Table 5: Holistic Scoring Method Sample Data

Participant	Correctness	Comprehensiveness	Score	Percentage
1	3	3	6	100
2	0.5	3	3.5	58.333
3	0	3	3	50
4	0	3	3	50
5	0	2.5	2.5	41.667

Table 6 shows the p-values computed from matched paired Wilcoxon comparisons. There are no significant differences between any of the datasets, demonstrating that students do not collectively perform better or worse by one scoring method relative to another ($p > 0.12$).

Table 6: Wilcoxon Nonparametric Comparison

dataset 1	dataset 2	p-value
Trad	Hol	0.223
Trad	Cat	0.1297
Hol	Cat	0.3128

Figures 1 and 2 represent the difference in score for each map between categorical and traditional in grey and holistic and traditional in red. Dots on the figures represent individual points within the larger dataset. Figure 1 compares the differences using the same "percent difference" axis as a comparison of the two datasets, while Figure 2 has an expanded axis for the holistic dataset for better clarity in the spread.

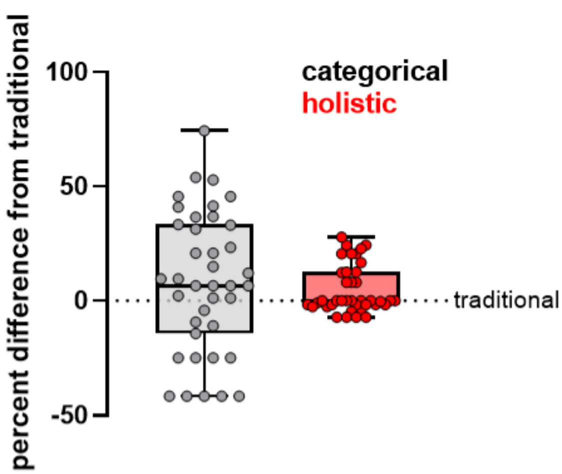

Figure 1: Percentage Difference from Traditional Comparison Same Scale

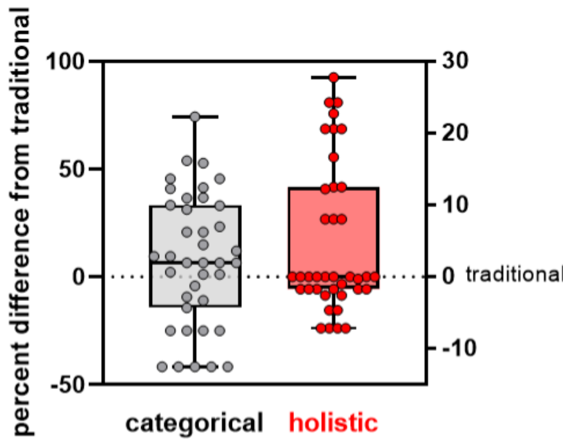

Figure 2: Percentage Difference from Traditional Comparison Different Scale

The visual representation of the percentage difference from the adapted traditional method reveals how tightly spread individual scores are across the scoring methods. In particular, the tighter spread between holistic and traditional may point to the structural similarities of the methods that is otherwise unseen with the paired t-tests.

Discussion

A. Inter-method scoring

The statistical analysis performed in this study indicates that each dataset was found to have a non-normal distribution, and the datasets were found to not be significantly different from each other. The non-normality of the data is potentially an indication of the subject matter understanding of the sampled population, or the difficulty of the assignment. The fill-in concept map assignment was administered in a first-year engineering course, a student population that may have had a similar level of exposure, or lack thereof, to the EM subject matter. Additionally, a high density of scores was found near the top of the range, indicating that the assignment could have been easy for the participants leading to a ceiling effect.

The lack of statistical difference between scoring methods indicates that the scoring methods may be assessing a similar skill. Although this does not prove that these scoring methods assess conceptual understanding, this is a promising step toward generating scoring methods for fill-in concept maps.

B. Traditional

The adapted traditional scoring method closely resembles the correct/not correct method that is typically employed for fill-in maps. While little analysis of the efficacy of the method in comparison with other methods has been conducted, much research has demonstrated that fill-in concept maps can successfully predict high-level conceptual understanding when graded with the correct/not correct method [9,10]. With this context, it could be considered the most consistent and grounded scoring method when applied to fill-in maps. An instructor can be confident that a high scoring map generally describes the central topic better than a low scoring map.

Additionally, with the objectivity in this method, little expertise is necessary in the area of the concept map topic aside from constructing the primary map; few discrepancies in results should occur, and scoring a large set of maps should be fairly efficient with the adapted traditional method compared to other methods.

However, as indicated, this method assumes objectivity—that a concept is either used correctly or incorrectly. This assumption can be problematic when the interpretation of concepts is unclear or when a connection can be made between two concepts that are not represented in the primary map. There is no gradient of correctness in this method, so regardless of whether the inclusion of a concept could be considered valid or not—it will receive no points if it does not match the primary map. Additionally, a cascade effect can be exhibited in incorrectly placed concepts. For concepts that connect to multiple other concepts, a singular incorrect placement will cause all connected concepts to also be marked incorrect. Should two separate participants misplace the same number of concepts, they may receive different scores depending on which concepts those are. This may be sought after in the case of fundamental concepts being misplaced, but potentially runs counter to the spirit of a traditional map in its objectivity.

Figure 3 shows an example of this cascading effect. The concepts highlighted in blue are concepts that were incorrectly placed, and the linking words circled in red are locations where an interlink is not correctly matching. In the case of this map, two incorrectly placed concepts caused the map to have 15 of 22 correctly matching links, an adapted traditional score of 52%. If adapted traditional score were proportional to the number of correctly placed concepts, this map would receive an adapted traditional score of 79%.

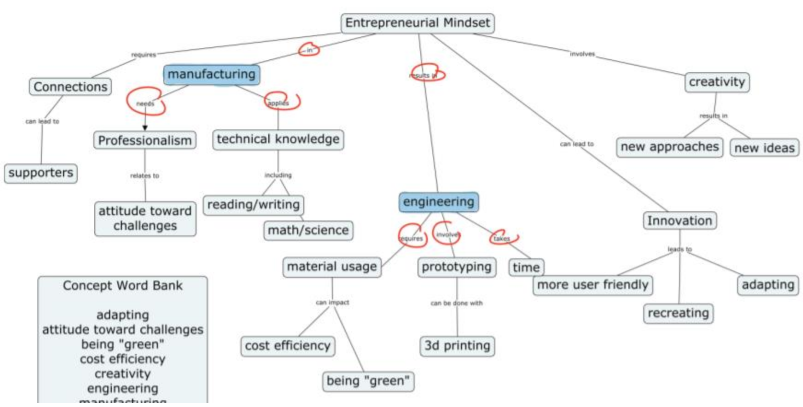

Figure 3: Traditional Scoring Cascade Effect

There is a plethora of benefits of the traditional scoring method, but it requires careful execution when developing the primary map. An instructor should pay close attention to whether concepts could validly fit in other locations than is reflected in a primary map. Should a primary map control for this, this adapted traditional scoring method would likely be an effective scoring method.

C. Categorical

The adapted categorical scoring method has little differences from the original categorical scoring method which comes with some concerns in the fill-in format. This method was time consuming both in categorizing the concepts generally, and in interpreting participant intent for categorization in the context of each specific map. A scorer using this method should have clear understanding of the topic and its underlying concepts.

On top of the more rigor required to assess a map categorically, some concerns were present. Because of the inverse relationship with number of categories inherent to the scoring method, without increasing the number of interlinks between concepts, the score of a map is diminished for increasing the number of categories they use in their map. This disincentivizes emphasizing breadth of the map which is generally not a desired outcome in concept mapping assessments.

Occasionally the feature described above had an interesting effect wherein the primary map would receive a lower score than a traditionally incorrect map because that incorrect map used fewer categories. Figure 4, below, was found to only include 15 of the 17 concepts in the concept bank, thus repeating words. With this choice by the participant, the map used 8 categories and had 8 interlinks, compared with 9 and 7 for the primary map, respectively. Consequently, this map, and others, received a higher score than the primary map in the adapted categorical scoring method.

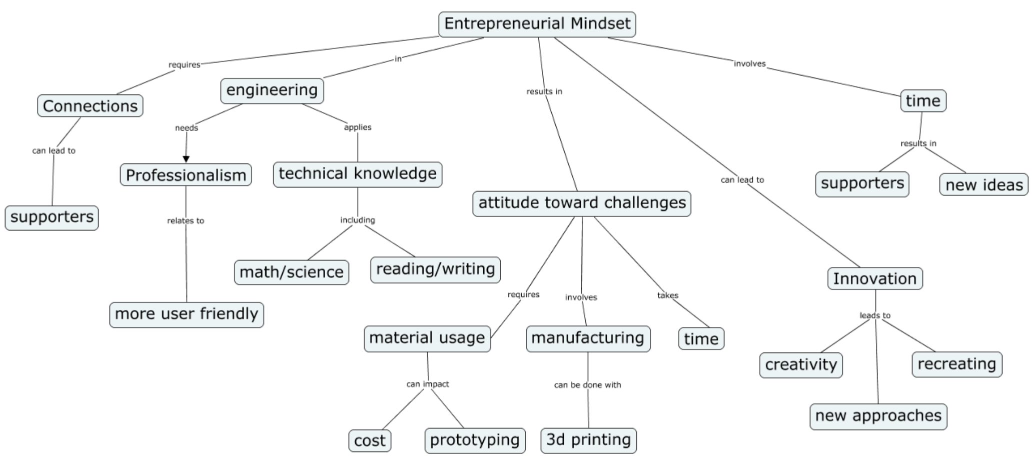

Figure 4: Categorical Primary Map Outscoring

It is important to note the flexibility that is offered to the student through the adapted categorical scoring method. Because of the fluidity of the category that a concept can fit under, participants were allowed, without having been told, to change the meaning of a concept away from what an instructor may have thought is ideal. This can be beneficial in that a participant's difference from a primary map could indicate nuanced understanding of the central topic, but may also incentivize truly incorrect concept placement.

The example map in Figure 5 shows this flexibility. The concepts highlighted in blue are concepts that are placed in a different location than that of the primary map. This map connects concepts of different categories much more regularly than the primary map. The differently placed concepts in the other two scoring methods would be considered incorrect and lead to point deductions, but in categorical are only considered by their category. From the perspective of the topic in question, some of these concepts should be considered incorrect, but some could be argued to fit their different location. This shows both the promise and drawback of the categorical method's flexibility. In this case, the map scored much more highly than the primary map with 22.7 points, compared with 13.2 points for the primary map.

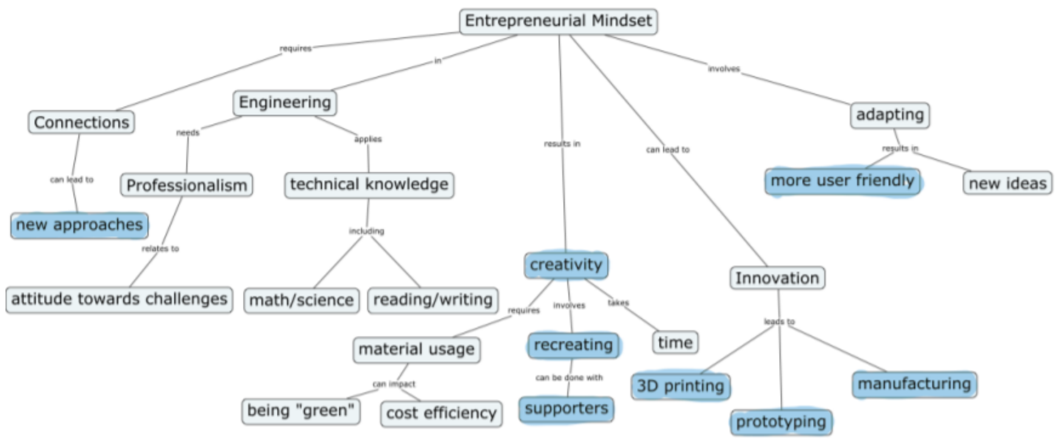

Figure 5: Categorical Scoring Flexibility

The categorical method has complex interactions when adapted to fill-in maps, but is a promising method, particularly with additional adaptations and considerations during usage. This method attempts to capture the complexity of a concept map through its many connections between the categories of the central topic. An adaptation of the method that refrains from reducing a map's score for utilizing more categories of the central topic may better fulfil the goal of assessing breadth of conceptual understanding. Additionally, an adapted categorical scoring method would likely be best employed for a topic that is deeply connected and particularly complex in its connections between concepts of the topic.

D. Holistic

Much like the adapted categorical scoring method, the adapted holistic scoring method is similar to its original, but presents some concerns, albeit more subtle than that of the adapted categorical method. Unlike the holistic method when applied to unstructured maps, this adapted method is not substantially more time consuming than the categorical method. This method requires categorizing concepts as well, but there is much less subjectivity overall. This manifests in the rigid scoring for comprehensiveness and correctness based on the categories and the primary map, respectively.

This adapted holistic method did not distinguish between maps in a realistic pattern. In comprehensiveness, very few maps did not receive a perfect score, whereas in correctness, a majority of maps received a score of zero. Because of the structure provided to the student, missing categories generally requires not including concepts, or significantly reassigning the category in which concepts fit which is generally uncommon. Additionally, because each incorrect concept lost the student half a point of the three correctness, a moderately correct map (11 of 17 correct concepts) would still receive a score of zero. Much like the traditional scoring method, the holistic scoring method requires that a map matches that of a primary map without concern for a gradient of how correct a student response may be, but because this correctness is only one component of the overall score, this circumstance is less prominent in the holistic case.

The pie chart in Figure 6 shows the distribution of scores in the correctness metric across the dataset. A score of zero is the overwhelming mode for this metric. The adapted holistic scoring method is unable to distinguish levels of correctness in the concept maps beyond the top half of performing scores in the metric, and could be adjusted to better do so.

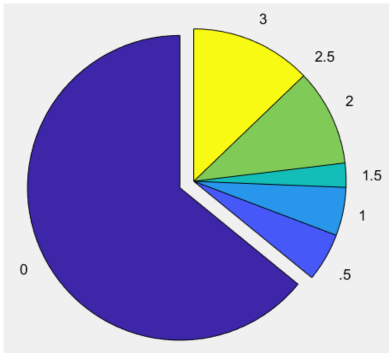

Figure 6: Holistic Scoring Method Correctness Distribution

Although there are difficulties in adapting the holistic scoring method to a fill-in map while retaining its original spirit of comprehensive analysis, it could be the most flexible method given some alterations. A potential solution to the correctness problem is a correctness scale as a percentage of correctly placed concepts. This change allows for more differentiation in assessing a map's concept placement. A restructuring of the scoring scales for this adapted holistic method may yield a scoring method that is more generalizable in many circumstances. This is because it would have a diminished presence of the traditional scoring downside of forced objectivity, with the caveat of taking significantly more resources to score the maps.

Conclusion

The adapted fill-in concept map scoring methods used in this study were found to have no significant difference in normalized scores across the dataset. This potentially indicates that the scoring methods assess a similar skill and can be generally used to assess fill-in concept maps. However, instructors should still consider the context in which these scoring methods are used to avoid potential flaws in assessment, as these scoring methods are derived from scoring methods with particular goals in mind. In particular for each adapted method, it should be noted that the traditional method is the most grounded in prior research relating to fill-in maps, that the categorical method is best suited to assess complex and highly integrated topics, and that the holistic method may be more applicable in a range of circumstances. With fill-in maps more generally, an instructor should either use a scoring method that avoids concept ambiguity, or construct a primary map in a way that minimizes concept ambiguity.

In conjunction with the quantitative statistical data, the qualitative information from scorers informs the various differences in the assessment of each scoring method and potential alterations to the methods to better suit their goal of assessing conceptual understanding from fill-in concept maps. Each method may have best use circumstances due to their particular advantages and disadvantages. The traditional method failed to allow for a gradient of correctness, but was particularly easy to implement; the categorical method introduced a point reduction for breadth of categorization, but is well suited for assessing highly interconnected topics; and the holistic method was resource intensive, but may be a more generalizable scoring method. These findings track with the formulation goals of the unstructured scoring methods [8].

Future work should analyze the ability of the adapted scoring methods to assess conceptual understanding of a particular topic. This can be done to either the adapted methods used in this study or to the adapted methods after undergoing the alterations proposed in this study, those being the removal of the inverse relationship between category inclusion and score for categorical and a restructured point distribution for the holistic correctness metric. Prior work analyzing instructional tools in engineering entrepreneurship found that concept mapping is advantaged in its ability to assess conceptual connections and depth [5]. Other studies, however, have used concept maps focused on particular topics within a curriculum [10,15,20]. Collectively, these studies demonstrate a broad set of applications for concept maps from informing on student mindset attributes to guiding focused course redesign. Focused studies like these can be conducted with these adapted methods to assess the efficacy of the methods in similar circumstances.

Additionally, a broader set of perspectives in scoring with the various methods should be considered. This can be done with similar studies analyzing applications of the adapted methods and the perception of scoring method usage by instructional teams. For example, Martine et al. used the unstructured traditional and holistic methods to assess student perception of EM, finding that evaluation method should be selected following consideration of research goals [3]. This study and others can be similarly used for analysis of the adapted fill-in scoring methods.

References

1. "Engineering Unleashed," Online, 2021. URL https://engineeringunleashed.com/
2. Byers, T., Seelig, T., Sheppard, S., and Weilerstein, P., "Entrepreneurship, Its Role in Engineering Education," *The Bridge* Summer 2013.
3. Martine, M.M., et al, "Concept Maps as an Assessment Tool for Evaluating Students' Perception of Entrepreneurial Mind-set," *126th American Society for Engineering Education Annual Conference and Exposition*, June 15-19, Columbus, OH, 2019.
4. Bosman, L., and Fernhaber, S., "Teaching the Entrepreneurial Mindset to Engineers," *Springer International Publishing* 2018.
5. Purzer, S., Fila, N., and Nataraja, K., "Evaluation of Current Assessment Methods in Engineering Entrepreneurship Education," *American Society for Engineering Education* Winter 2016.
6. Novak, J., "Concept Mapping: A Useful Tool for Science Education," *Journal of Research in Science Teaching* 27.10 (1990): 937-949.
7. Jackson, A., et al, "Refining an Entrepreneurial Mindset Primary Concept Map through Multi-Institutional Collaboration," *Research in Engineering Education Symposium & Australasian Association for Engineering Education Conference*, December 5-8, Perth, WA, 2021.
8. Watson, M.K., Pelkey, J., Noyes, C.R., and Rodgers, M.O., "Assessing Conceptual Knowledge Using Three Concept Map Scoring Methods," *The Research Journal for Engineering Education* 105.1 (2016): 118-146.
9. Soleimani, H., and Nabizadeh, F., "The Effect of Learner Constructed, Fill in the Map Concept Map Technique, and Summarizing Strategy on Iranian Pre-university Students' Reading Comprehension," *English Language Teaching* 5.9 (2012): 78-87.
10. Schau, C., et al, "Select-and-Fill-in Concept Map Scores as a Measure of Students' Connected Understanding of Science," *Educational and Psychological Measurement*, 61.1 (2001): 136-158.
11. Turns, J., Atman, C.J., and Adams, R., "Concept Maps for Engineering Education: A Cognitively Motivated Tool Supporting Varied Assessment Function," *IEEE Transactions on Education* 43.2 (2000): 164-173.
12. Montfort, D., Brown, S., and Pollock, D., "An Investigation of Students' Conceptual Understanding in Related Sophomore to Graduate-Level Engineering and Mechanics Courses," *The Research Journal for Engineering Education* 98.2 (2009): 111-129.
13. Novak, J. D., and Gowin, B., "Learning How to Learn," *Cambridge University Press, New York* NY 36.1 (1984): 93-108.
14. Ruíz-Primo, M., "On the use of concept maps as an assessment tool in science: What we have learned so far," *Revista Electrónica de Investigación Educativa* 2.1 (2000).
15. Segalàs, J., Ferrer-Balas, D., and Mulder, K. F., "Conceptual maps: measuring learning processes of engineering students concerning sustainable development," *European Journal of Engineering Education* 33.3 (2008): 297-306.
16. Besterfield-Sacre, M., et al, "Scoring Concept Maps: An Integrated Rubric for Assessing Engineering Education," *The Research Journal for Engineering Education* 93.2 (2004): 105-115.
17. Bodnar, C.A., Jadeja, S., Barrella, E., "Creating a Primary 'Entrepreneurial Mindset' Concept Map," *American Society for Engineering Education Virtual Conference*, June 22-26, 2020.
18. "Cmap Tools," Online, 2021. URL https://cmap.ihmc.us/cmaptools/
19. Richmond, S.S, "A Set of Guidelines for the Consistent Assessment of Concept Maps," *International Journal of Engineering Education*, 30.5 (2014): 1072-1082.

20. Kecskemety, K.M., et al, "Using Concept Maps for Assessment in an Aerospace Structures Course: A Comparison of Scoring Techniques," *AIAA Scitech 2022 Forum*, January 3-7, San Diego, CA, 2022.