International Journal of Engineering Education Vol. 41, No. 1, pp. 125–135, 2025
Printed in Great Britain

0949-149X/91 $3.00+0.00

Evaluation of a Semi-Automated Scoring Tool for Assessing Entrepreneurial Mindset Concept Maps*

EDUARDO RODRIGUEZ-MEJIA**[1], MARTHA L. CANO-MORALES[1,2], ALEXANDRA JACKSON[1], ELISE BARRELLA[3] and CHERYL A. BODNAR[1]
ORCID in order of authorship: 0009-0007-5522-0069, 0000-0003-3713-7038, 0000-0002-4662-6653, 0000-0003-0020-2035 & 0000-0002-8665-9839
[1]Experiential Engineering Education Department, Henry M Rowan College of Engineering, Rowan University, Glassboro, New Jersey, USA.
[2]Electronics Engineering Department, Faculty of Engineering, Pontificia Universidad Javeriana, Bogota, Colombia.
[3]DfX Consulting LLC, Winston Salem, North Carolina, USA.
E-mail: rodrig237@rowan.edu, canoma14@rowan.edu, jacksona4@rowan.edu, elise@dforxconsulting.com, bodnar@rowan.edu

Assessing Entrepreneurial Mindset (EM) is recognized as a challenge within the engineering education field. Concept maps (cmaps) show promise for direct assessment of EM development, but manual scoring of cmaps is time-consuming and subjective, highlighting the need for automated scoring tools to alleviate these drawbacks. Yet, automated tools for categorical scoring – an approach combining quantitative and qualitative metrics – are lacking. To address this gap, we developed the Semi-Automated Scoring Tool (S-AST) and conducted a study involving five faculty members to evaluate S-AST. Participants engaged in manual and S-AST-assisted categorical scoring sessions, followed by interviews with the research team. The research aimed to determine the extent to which S-AST facilitates categorical scoring and to identify areas for further improvement of the tool to better meet faculty needs. Results underscore the potential of S-AST in streamlining EM cmap assessments and highlights areas for improvement, such as enhancing the user interface for better navigation and clarity, improving explanations in the results file, and providing a visual representation of the cmap to increase faculty members' confidence in category assignment.

Keywords: concept maps; assessment tool; categorical scoring; educational technology

1. Introduction

Entrepreneurial Mindset (EM) consists of skills, attributes, attitudes, including multidisciplinary communication, value recognition, and innovation [1–3]. In recent years, there has been a push for more entrepreneurially-minded individuals in the engineering field as a means to encourage innovative thinking and creativity [4, 5], leading to the incorporation of EM interventions in undergraduate engineering programs [2]. To measure the impact of these interventions, it has been necessary to identify assessment methods that can capture EM development [6, 7].

One method of direct EM assessment is through concept maps (cmaps), a graphical tool used to depict an individual's knowledge surrounding a specific topic [8, 9]. Cmaps were first developed by Novak & Gowin [10] and include three main elements: concepts (terms or phrases that are related to the main topic), propositions (words or phrases that link concepts together and provide context for their connections), and hierarchies (groups of concepts and propositions that branch directly from the main topic). Assessment of EM cmaps through both quantitative and qualitative methods has shown promising results, with students' cmaps often including concepts relevant to EM, such as business knowledge, design and development, and professionalism [11, 12]. The assessment process for cmaps often involves teams of instructors or researchers who score the cmaps using quantitative parameters such as number of concepts, or qualitative measures such as map correctness [13]. Though research has focused on validation of different cmap scoring methods [11, 14], this process is often time consuming, subjective, and involves calibrations between different interpretations of map elements [15, 16]. One mixed-methods approach is called categorical scoring, which involves assigning concepts in a cmap to a set of predetermined categories [13, 17]. This method is said to be more accurate for understanding both the metrics and organization of concept maps as well as the quality and accuracy [9, 13], but it tends to be one of the more time-consuming methods and struggles to achieve high inter-rater reliability [18, 19].

One method to reduce the drawbacks of concept mapping assessment is automated scoring, which has been shown to be just as accurate and take less time than manual scoring [20]. Though automated scoring tools exist for other types of cmap assessment, there does not yet exist an automated tool for

** Corresponding author.

Martha L. Cano-Morales and Eduardo Rodriguez-Mejia should be considered joint first authors of this article.

* Accepted 23 September 2024.

categorical scoring. We developed a semi-automated scoring tool (S-AST) using an existing categorical scoring codebook specific to EM previously developed by three of the authors, which incorporates both quantitative and categorical scoring options, and a set of features that enhance the scoring process.

To evaluate the extent to which S-AST facilitates assessment of EM cmaps through categorical scoring, we seek to answer the following research questions: (1) How does S-AST alleviate difficulties associated with manual scoring of EM cmaps? and (2) What areas of improvement are still necessary for S-AST to meet the needs of faculty members?

2. Literature Review

2.1 Manual Cmap Assessment

There are many types of manual cmap assessment, all of which focus on different elements such as map quality, correctness, and similarity to other maps. Traditional scoring is the most common and simplest method, relying on counting the number of concepts, number of cross-links, and number of hierarchies to determine an overall map score [10, 21]. More qualitative measures, such as Holistic rubric scoring, evaluates cmaps on a scale from 1-3 based on organization, comprehensiveness, and correctness [22]. Some cmaps are also scored by determining their similarity to an instructor or researcher-created "criterion" or "master" cmap on the same topic, referred to as expert map comparison scoring [23, 24].

The method we will focus on in this paper is referred to as categorical scoring, which is a mixed methods approach that involves both counting map elements and observing overall map themes and quality [17, 9]. Categorical method provides insight into the structure and content of the student's knowledge, bringing together elements of the traditional and holistic scoring methods [13]. This method was first developed by Segalas et al. [17], who created a codebook with 10 categories related to sustainability to assist in scoring sustainability-focused cmaps. These maps were qualitatively evaluated through the categorical themes, and quantitatively evaluated using a formula for complexity index (CI), found by applying the number of concepts (NC), number of categories (NCAT) and the number of interlinks between categories (NIL) to the formula presented in (1):

$$CI = NC\frac{NIL}{NCAT} \qquad (1)$$

The categorical scoring method was modified with new categories for sustainability-focused maps by multiple other studies [25, 26], eventually leading to this method expanding outside of the sustainability space to address topics such as infrastructure and engineering technology [27]. Recently, three of the authors worked on the development of the first categorical scoring codebook to assess EM concept maps, which was used for the development of the tool presented in this paper.

It has been suggested that the best choice for manual scoring method depends on the context of the cmap creation, such as the number of available scorers, the number of maps to be scored, and the time constraints on assessment [8]. Since more complicated methods such as categorical scoring often require more scorers and time [19, 26], it is necessary to develop an automated method that captures the accuracy and quality of assessment that categorical scoring provides without the common drawbacks.

2.2 Automated Cmap Scoring

Since manual cmap scoring has drawbacks, several automated scoring methods have been developed to help mitigate them. Some researchers have developed their own programs from scratch [15, 20, 28], others have used pre-existing software to develop their own approaches [23, 29], and others have modified existing programs to fit their needs [30].

Most of these automated methods use expert map scoring as the baseline and feed their programs the expert map so the student maps can be compared automatically [31–33]. Expert map scoring based programs often use similarity flooding algorithms (SFA) for comparisons [23, 30], which works by translating two different datasets into graphs and determines similarity values to compare each data point [34]. Maps have also been compared using similarity based equations for matching similar concepts and distance between similar concepts [35, 36], designing a metric based program that highlights specific map elements [37, 38], and creating a pathfinder network to compare across maps [39, 40]. In Koul et al.'s [39] program, ALA-Mapper, cmap data is converted into proximity data which determines the distance between each link and concept, then the Knowledge Network and Orientation Tool converts this data into a pathfinder network. This network is a way to depict the data without linking phrases to allow for direct measurement of relatedness between two concepts. In a more complicated equation based program, Gurupur et al. [41] use hierarchical comparison to complete Markov Chain analysis on cmaps. This program uses cmaps as XML files, parses them through a Java program for the analysis portion, and provides the number of concepts, propositions, and a score for the correctness of each concept based on the results.

Though most automated scoring tools have used Expert Comparison as their baseline, there are a few that have applied different approaches. Luckie et al.'s [42] program, RoboGrader, involves the instructor manually assigning correctness to a string of terms, and the program analyzes each map based on these correctness scores. Although this does not involve an instructor creating an expert map, this method does involve influence from an expert.

Another example of an existing tool is the Cmap-Parse tool developed by Watson et al. [20], which uses Traditional scoring as the baseline. Python programming and NetworkX are used to identify the four parameters of Traditional scoring, number of concepts (NC), number of hierarchies (NH), highest hierarchy (HH), and the number of crosslinks (NCL), and parse them into a weighted form of the Traditional Scoring equation: (NH – NC) + 5*HH + 10*NCL. Cmap data is reorganized into directed multi-graphs which can be interpreted by NetworkX algorithms, outputting each individual score.

Automated cmap scoring has helped to decrease the amount of time and effort it takes to score cmaps, but these approaches have been shown to have their own set of drawbacks that some automated methods have worked to combat. The main drawback is often accuracy, as it can be difficult to accurately depict correctness of cmaps through an algorithmic or equation based approach [20]. The similarity flooding algorithm's accuracy has been strengthened by applying WordNet to the program, which allows for words with similar meanings to retain higher similarity values when compared [33, 43]. Conlon's [42] program, the Reasonable Fallible Analyser (RFA), has also implemented an "argue" feature, where students can challenge any feedback presented by the RFA if they feel it is incorrect. A few programs have also opted for a closed-map approach, in which only a set of terms chosen by the instructor can be used in cmaps, making it easier for the program to provide correct and more useful feedback [31, 32].

Automated cmap scoring has historically offered an alternative to time consuming manual assessment [20, 33]. However, automated methods have yet to include Categorical Scoring, which has been favorable for understanding the quality of maps both qualitatively and quantitatively [13]. Pruett & Weigel [44] were interested in scoring using both Traditional (Cmap-Parse tool) and Categorical methods but performed manual Categorical Scoring using Segalas et al.'s [17] codebook due to the lack of an automated tool. Therefore, the existence of a tool that performs categorical scoring in conjunction with other scoring methods may be beneficial for comparing different types of scoring, extracting prominent themes, and identifying conceptual gaps.

3. S-AST a Semi-Automated Scoring Tool to Assess EM Cmaps

This section will provide an overview of the S-AST tool and how it can be applied to categorically score cmaps.

3.1 Tool Design and Implementation

With the aim to facilitate the process for assessing cmaps using the categorical scoring method, the research team developed S-AST to score concept maps produced (or reproduced) with the Cmap-Tools^{TM} software. The tool was developed in the programming language Python, based on the algorithm developed by Watson, Barrella & Pelkey [20] for assessing cmaps through use of traditional scoring methods. The original code for the traditional scoring method was preserved, although slightly modified to include fault tolerance and enhance the results presented in the output file. Then, a completely new code was developed to incorporate the categorical scoring approach.

S-AST has a graphical user interface (GUI) (Fig. 1) that provides the user with the option to select the scoring method, provide the root concept expression, select the cmap files to be assessed, define the location for the results file to be saved, and access to the EM codebook with the definitions and examples of the categories for categorical scoring. In addition, the user can access a Help Window that provides a detailed PDF help file. This Help file contains instructions for creating and exporting cmap files, as well as customizing the tool to the users' own categories and topic. By default, S-AST includes the EM codebook file with the categories' definition and an EM wordbank file with a list of concepts related to each category, nevertheless, users can change either the code book or wordbank file for their own needs. They only need to follow the format and instructions provided in a detailed help file incorporated within the tool. The tool also has the feature of batch assessment that allows the user to upload multiple files simultaneously. The tool accepts only .cxl extension files, which can be obtained from exporting a .cmap file into a .cxl using Cmap-Tools^{TM} [45].

S-AST first extracts the concepts and their links from the cmap, then it compares the concepts against the wordbank file for the selected topic, in this case EM. If there is an exact match, the concept is assigned to the corresponding category from the wordbank. For a concept that does not have a match within the wordbank file, the algorithm will

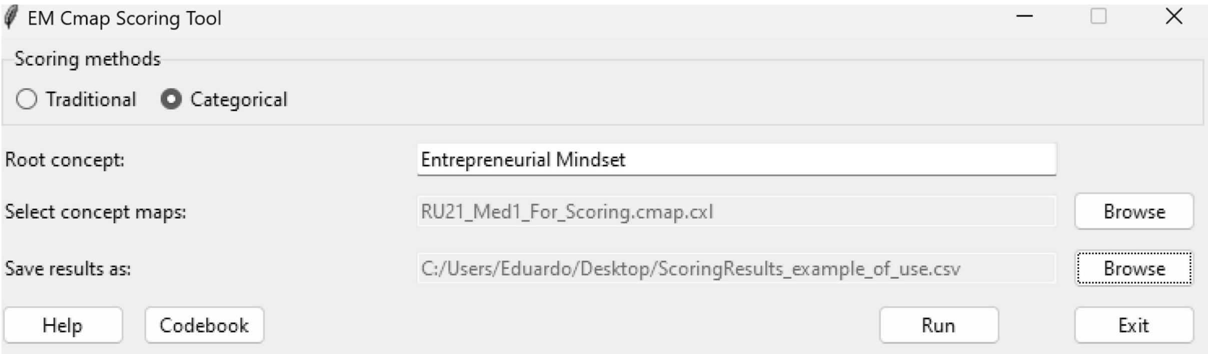

Fig. 1. Scoring Tool Graphical User Interface (GUI).

assign it to the category of the first concept of the hierarchy to which it belongs. This is called a pre-assignment. After all concepts are assigned or pre-assigned to a category, the concepts with a pre-assigned category are presented to the user for review allowing the user to either accept the categories or manually change them to a new category based on their own line of reasoning. After the pre-assignment review, the algorithm calculates the scoring metrics and generates the results file with the complexity index along with detailed cmap metrics such as the number of concepts, the number of categories, the list of concepts within each category, the number of crosslinks, and a matrix with the crosslinks between categories.

S-AST includes fault tolerance codification that prevents the program from crashing unexpectedly. For instance, it verifies that the wordbank exists before trying to access it. It also verifies that the cmap file has the correct file extension, that it has at least two concepts and one link, and that at least one concept node matches the Root concept written in the main GUI. Additionally, it checks that the results output file is not open. In any of those cases the program safely ends and informs the user about the fault.

The program was designed to facilitate assessing concepts with unexpected format, for example it eliminates double spaces or new lines within concepts. It eliminates concepts with no links to other concepts and empty nodes. Additionally, it identifies duplicated concepts and merges them into one.

S-AST is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License and can be downloaded from GitHub under the name EM-Cmap_Scoring_Tool [46].

4. Study Method

To evaluate the use of S-AST, a study was conducted with faculty members who had prior experience teaching design courses and a background with the concept of EM. Participants took part in two one hour sessions, one for manual categorical scoring and one with the use of S-AST. Each session consisted of a think aloud cognitive interview where the faculty member talked through their approach to scoring the provided cmaps and a follow up interview. Institutional Review Board (IRB) approval and informed consent for each participant was obtained before conducting the study.

4.1 Participants

Faculty members who met the criteria of having prior experience with teaching design courses and a background with the concept of EM were contacted via email to participate in the study. From those who responded to the invitation, five were selected. Participants demographics included two female and three male faculty members. Faculty members' background with EM was developed through attending a KEEN [47] conference or workshop (four participants), being a Co-PI on a KEEN grant (two participants), or participating in campus-based efforts to integrate EM into their classes (two participants). All faculty members had experience with teaching Junior or Senior students' design projects, and were from Biomedical Engineering (one participant), Chemical Engineering (one participant), Experiential Engineering Education (two participants), and Mechanical Engineering (one participant). Participants were offered a $50 dollar prepaid gift card for their participation in the study.

4.2 Data Collection

Participants were instructed to complete a self-guided virtual training [48] on how to score a cmap using the categorical scoring method, both manually and with the aid of S-AST prior to the first meeting with the research team. This virtual training allowed faculty members to familiarize themselves with the scoring process and how to complete categorical scoring with both a simple and more complex cmap. In the first think aloud session, participants were asked to manually score four cmaps using the categorical approach, while articulating their actions, thought processes, and any difficulties encountered. After the think aloud process, participants were asked follow up questions regarding their experiences with the manual categorical scoring, which included challenges while scoring the cmaps, interpreting the scoring metrics, and perceptions of the scoring method. During the second session, participants were asked to score the same four cmaps but using S-AST, while articulating their actions, thought processes and difficulties encountered. Similarly to the first session, participants responded to follow-up questions regarding their experience of scoring with S-AST, which included challenges, ease of use, and perceptions of S-AST. Additionally, participants were asked about possible changes or improvements to the tool as well as positive features they liked while scoring the cmaps. Interviews were audio recorded with Zoom and subsequently, the transcripts generated by Zoom were reviewed and cleaned by the research team.

The four cmaps used during the think aloud interviews, were developed by the research team based on their experience with using and assessing EM cmaps in the classroom. They were identified as "Easy", "Med1, "Med2", and "Difficult" according to the research team criteria based on the number of concepts included, the listing of the concepts within the Wordbank file, and the number of interconnections. The "Easy" cmap only included concepts found in the Wordbank file, while "Med1" and "Med2" included a mixture of concepts within and outside of the Wordbank file. On the "Difficult" cmap, nearly half of the concepts were not found in the Wordbank, and the structure of the connections was more elaborate. Figs. 2 and 3 present a visual representation of the easy and difficult maps respectively.

4.3 Data Analysis

With the lens of the research questions, inductive thematic analysis was conducted to identify and organize patterns of meaning from the transcripts data set [49]. The transcripts were divided equally among two of the research team members for the

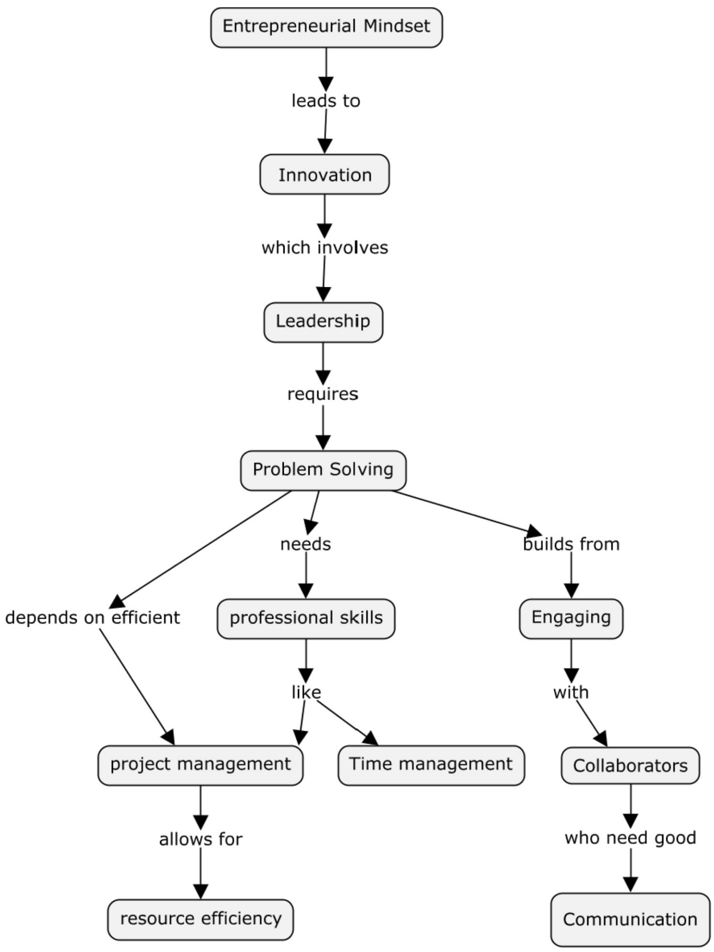

Fig. 2. "Easy" concept map.

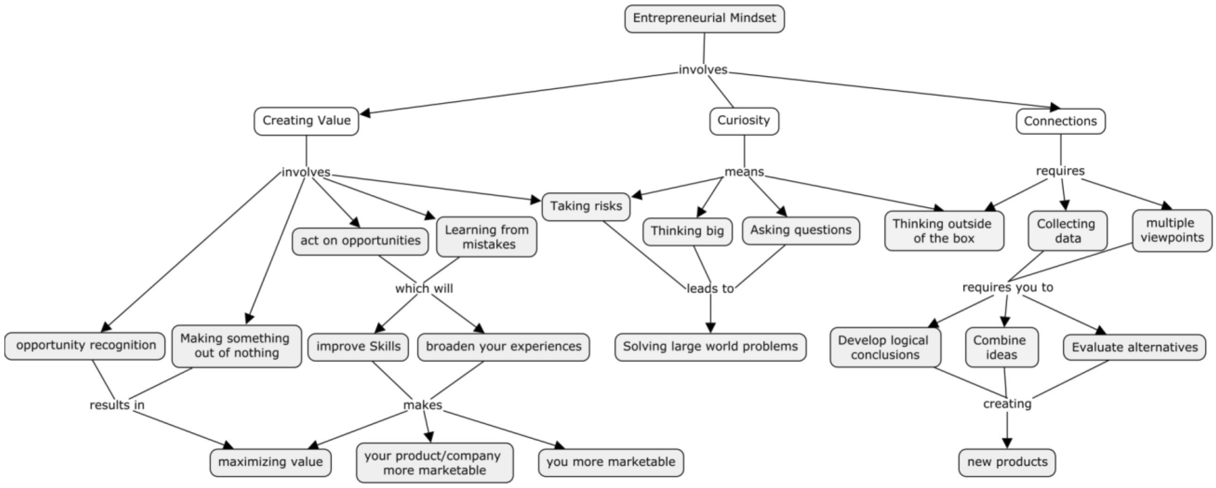

Fig. 3. "Difficult" concept map.

generation of initial codes, both semantic and latent. The two researchers then identified overarching themes present within the initial codes and met with the third researcher to confirm their approach and that the themes were relevant to the study's research questions. After consolidating the themes, two codebooks were created, with the definition of themes and subthemes. Subsequently, all three of the researchers went through one of the transcripts and coded it using both codebooks. A discussion followed to reconcile the themes and interpretation of the coded segments, thus verifying the codebooks' accuracy.

In the second coding cycle, the transcripts were again split between two of the researchers who deductively coded each transcript with the verified codebooks [50]. Each of the transcripts were interchanged between the researchers for verification of code application. Any discrepancies were reconciled through discussion with the third researcher on the project.

4.4 Research Quality

The study ensured research quality through the guidance of validation measures outlined in the Q3 framework [51]. Procedural validation, defined by mitigating threats to overall validation on the data collection and analysis, was achieved by using the think aloud cognitive interviews to elicit participant's perceptions in a context that resembled a natural setting for assessing cmap complexity, either manually or with the S-AST. Detailed immersion during the data analysis phase allowed the research team to better understand the participants' experiences for theme identification from within the data. Process reliability (collecting and recording data in a dependable and accurate way) was achieved by developing the session protocols for the think aloud cognitive interview and follow up interview and keeping an audit trail while going through the data analysis with reflections on the data and the analysis process being conducted.

4.5 Limitations

Limitations of this study are the small sample sizes of both number of participants and number of cmaps for scoring. Other limitations are the requested criteria of participants having experience with cmaps and EM, and all participants being faculty at a single institution, resulting in a potential bias in the findings and the results not necessarily being transferable to other higher educational contexts.

5. Results

This section addresses the overarching research question: *To what extent does S-AST facilitate the assessment of EM cmaps?* by addressing the difficulties alleviated through use of S-AST and areas still in need of further improvement.

5.1 How does the S-AST Alleviate Difficulties Associated with Manual Scoring of EM Cmaps?

There were three themes and corresponding sub themes identified in response to the research question: *How does the S-AST alleviate difficulties associated with manual scoring of EM Cmaps?* (Table 1)

The two most prevalent sub themes identified across participants were the support of students' learning through score comparison and feedback, and efficiency through time savings compared to manual scoring. Four out of five participants commented on the way they perceived S-AST *"helps you get your students the feedback very quickly"* (faculty 1), or *"see (...) which areas are students thinking about when they're thinking about entrepreneurial*

Table 1. Difficulties alleviated when using S-AST

Theme	Sub-theme	Definition	Examples
Consistency in scoring	–	Being sure that the categorical metrics were identified and counted appropriately across all Cmap files in the same way, thus ensuring fairness in the final complexity index by reducing human mistakes.	"There are likely many things that I would have made mistakes with repeated reviews, so that eliminates kind of potential . . . error or mistakes. So it makes it more consistent. . ." (faculty 1)
Supporting students' learning	Through score comparison and feedback	Ability to compare students' results and quickly share them with the students.	"So you could theoretically upload a whole classes worth of students, and run them all at once and do all the categorization, and then compare across... You know... this one is much more interconnected, but it doesn't have as many nodes, and this one has more nodes, and some of them may be more similar." (faculty 2)
	By measuring their progression	Identify student's progression by seeing how their scores change over time.	"use this [tool] instead of doing a traditional or doing it by, you know, [scoring] by hand (. . .) looking at whether students are gaining complexity. In that case I would say, Oh, yeah, absolutely." (faculty 3)
Efficiency	Through time savings	Reduction in time and effort investment in obtaining scores for the cmaps and sharing it with the students.	"I prefer using the automated, because I think it's a more efficient way to get to the same place." (faculty 3) "This works really well in doing it a much faster way and getting nearly the same score" (faculty 4)
	Due to batch assessment	Ability to score more than one cmap at the same time and obtain the results in a single file.	"I really love that you can select more than one [cmap] at once. That's a huge like feature" (faculty 4)

mindset and which areas are they missing, so you can sort of push harder on the things that they are not getting" (faculty 5).

As for efficiency, participants' comments were often tied to the time it would take for grading and assessing batches of cmaps. All participants commented something related to *"the fact that you know* [the S-AST] *it's a time saver"* (faculty 1), the way the S-AST is *"streamlining the process"* (faculty 1) as it is *"less time intensive"* (faculty 2) and *"obviously it's way faster than the manual scoring"* (faculty 5).

5.2 What Areas of Improvement are still Necessary for the S-AST to Meet the Needs of Faculty Members?

Table 2 presents the four identified themes that address the research question: *What areas of improvement are still necessary for the S-AST to meet the needs of faculty members?*

The four most common sub themes were trust issues from the lack of information regarding the metrics, navigating the GUI, not being able to see the original cmaps, and expectations about the way the results information was to be presented in the output file.

Three out of five participants commented on the possible trust issues as *"with a* [sic] *automated tool, I actually wouldn't understand how the scoring works, it would just be completely embedded in the algorithm. I'm not entirely sure if it's a pro and a con, but it's just an observation."* (faculty 1), while another participant mentioned that it would be *"useful to like have to do some manual scoring before using the automated tool. Just so you like, have a clear sense of what's happening in the background of the tool"* (faculty 5). Additionally, one participant wondered if:

> *"is there a way to go back and physically see where it said there were 22 interlinks? Is there a way to go back and physically see those? (...) cause the presentation looks very clean. But then, if there's a discrepancy between what the tool did and what I* [manually] *did, and I wanted to go in and see the reason for discrepancy"* (faculty 3).

Four of the participants commented on the difficulties navigating the S-AST GUI, including not *"know*[ing] *that each screen was going to be one concept map at a time."* (faculty 4), missing the scroll bar on the manual categorization GUI, and getting confused by the presentation of the subcategories in the dropdown menu for the manual categorization assignment.

Four of the participants mentioned their expectation regarding how the information was presented in the output file, like expecting that *"the number of concepts would be the total of all of the concepts in* [each category] *section* [including the sub categories and the main category]*"* (faculty 5), not being sure what to do with the information if not having previously went through the training videos, *"some sort of chart that allows the instructors to see what are the range of complexity scores"* (faculty 4), and expecting each cmap file information in different tabs/sheets of the spreadsheet file.

Lastly, all five participants commented that not having access to the original visual representation of the cmap file can make it difficult to manually categorize concepts because *"sometimes you can read between the lines when you see more of it*

Table 2. Areas of improvement to be addressed relate to

Theme	Sub-theme	Definition	Examples
Trust issues	From lack of information regarding metrics	Uncertainty of the process followed for identification of the categorical metrics by the algorithm.	"There's this nagging feeling in my head that, you know. Did it . . . Did it work well? and did it tag Well?" (faculty 1)
Interface	Navigation	Hesitation while using the main window user interface and the manual categorization window user interface.	" (...)This is where I was a little confused. I'm not sure whether I should reject all the assignments and leave them in the no category. But that reads to me like that will undo everything I just did, or to collect accept all assignments, and that will do the ones that I have done in the new category." (faculty 2)
	Access to original Cmap	Lack of visual context for manually assigning concepts not found in the wordbank and potentially losing student's intent.	"You do lose some of that nuance from like the connecting words and sort of how the concepts fit in with each other in the big picture (...) none of the connecting words are there. So like, even though I can see all the concepts, you know, I don't know. It's harder to see what the students thought process is" (faculty 5)
	Customized feedback to students	The tool doesn't allow the user to include additional feedback in a concept map assessment.	"I mean, indeed, I think the way I see this tool is, I think you're slowly developing it. It all this thing that I did here, would completely be behind a visual, you know a GUI, where it pops up, you press a button, it starts to label and highlight and color. And then you can actually not only comment specifically . . . or comment generally, but you can comment specifically by space, meaning you can say, this category or this group of a cluster of information was fantastic. And that's how we do it with essays, we wanna circle things." (faculty 1)
	Output file format	Expectations of the results presentations or navigating the information in the file.	"I'm expecting each cmap to be in a different tab. I don't know if that's what's gonna happen." (faculty 5)
Other OS	Availability for iOS	The initial version of the tool is only available for Windows OS. Thus, participants with iOS machines couldn't use the tool on their personal computers.	"Mac users are gonna be troublemakers, it seems like (. . .) as a Mac user, this was not super easy." (faculty 5)
Word identification	–	The tool looks for literal concepts in the wordbank, losing its time efficiency when there are too many concepts not found in the wordbank. The tool could go beyond the literal search, to be able to assign concepts that are not in the word bank, and not just pre-assign them to the hierarchy's category.	"[cmaps] where the words are not often found [in the wordbank]. So you would have a lot of categorization to do on that second screen of this semi automated method. I don't know whether I would prefer to do that with the semi automated tool or with the manual scoring, and it is possible that I would . . . I would actually prefer a hybrid version of that where, in addition to the code book, I would also look at the Data Bank of of words or the word bank to be able to help classify in that and I didn't do that today, but that's what I would see as my ultimate usage of do everything through the automated tool, but merge those classifications with both resources." (faculty 2)

rather than just the entry in isolation" (faculty 2). This lack of access to the original cmap impacts the confidence level of faculty participants in *"make*[-ing] *a clear decision"* (faculty 5) as to where to assign a concept as *"the context being from the link, words or categories before it*, [they] *might be inclined to put it somewhere else"* (faculty 4).

6. Discussion

S-AST was developed in response to the need of an automated approach for categorical scoring of cmaps. Building on a traditional automated scoring tool developed by Watson, Barrella & Pelkey [20], S-AST created a means for automating the categorical scoring process by modifying the GUI to allow for this additional scoring approach, adding fault tolerance, and creating a results output file that captures the categorical scoring results.

6.1 Difficulties Alleviated by the S-AST

During the manual scoring phase, the research team findings aligned with previous literature on the difficulties associated with manual categorical scoring including that it can be time consuming, cumbersome, and tedious because of the identification of the different metrics [13, 20]. The identified themes for difficulties alleviated are relevant as they show that the use of S-AST can help reduce the time investment and the tediousness in identifying the metrics consistently across all cmaps, as well as give the faculty members enough information on the content to support formative and summative assessment.

Findings related to the ability of the tool to support students' learning are aligned with previous identifications of the categorical scoring method as the only cmap scoring method that allows insight into content and structure of the knowledge gained by students [13]. Because of its use of nominal data (categories) and ordinal data (scoring metrics), the results obtained with S-AST help faculty see the content that the student provided and better interpret the summative assessment, while being able to provide formative assessment as well [13, 20]. Lastly, reducing the time in assessing cmaps with the categorical scoring method can help make more faculty adopt this scoring method, as this is one of the limitations of the scoring method because of the dependence on the number of concepts present in the cmap [13]. It can also further help support students' learning, as teachers can give feedback more quickly and allows faculty to invest their time in other classroom activities [43].

6.2 Areas of Improvement for the S-AST

Further improvements of S-AST can include a window or message prior to the presentation of the manual categorization GUI to indicate that results will be presented one window at a time, highlighting or using bold text for the file name of the cmap being assessed, and including a confirmation message that prompts the user to scroll down, use arrows to indicate points of interest, or having the "Continue" button appear at the end after the user has scrolled down on the manual categorization window. The inclusion of bold text and arrows has been found to decrease task time and reduce software use errors [52]. For the total number of concepts, S-AST could be updated to count all the concepts under each category so faculty can easily perform a formative assessment. As for the chart, the output file already presents a table with the files assessed, the metrics and complexity index that can be imported to SPSS, MATLAB, or other software for statistical analysis that supports csv files. Lastly, for lowering the dependence on the training videos, the help file of S-AST can be updated to give the user a quick introduction to the information that will be saved in the output file after running the program.

As the categorical scoring method relies on nominal and ordinal data types, meaning that in the absence of a visual context the classification of nominal data will be compromised, it could be beneficial to include access to visual copies of the cmaps. This approach can also help ensure accuracy in the method, since categorical scoring can be considered to be both quantitative and qualitative, as it counts components to obtain the metric values for the complexity index and categorizes concepts [20]. The loss in the qualitative component, access to the visual representation of the cmaps, is the compromise for the time savings and consistency of identifying the components of the categorical metrics.

The faculty testing provided valuable insight into the strengths and weaknesses of the S-AST, and future versions of the tool will address the concerns raised by faculty to improve the user experience and value of the scoring results for EM assessment. Future work could explore ways to more fully automate the scoring process, perhaps through the use of machine learning and artificial intelligence.

7. Conclusions

S-AST was developed in response to the need of an automated approach for categorical scoring of cmaps. The tool was evaluated with faculty members to determine the difficulties it alleviated in comparison to manual categorical scoring and areas still in need of improvement. The main benefits of using S-AST were consistency in identifying categorical metrics, time savings, batch assessment capability, and support for student learning through score comparisons. Despite these advantages, S-AST raised trust issues among participants. They expressed concerns about losing access to the visual representation of the cmap and the student's intentions, and about not being able to see how S-AST computed the metrics. Additionally, participants identified areas for improvement in the navigation of the GUI. Faculty testing and feedback will be used to improve the tool and further ease the scoring burden of EM cmaps for assessment.

Acknowledgments – The authors would like to acknowledge a Kern Family Foundation grant entitled "Building an EM Concept Map Toolbox for Course and Program Evaluation" for funding the work associated with this study.

References

1. C. Naumann, Entrepreneurial mindset: A synthetic literature review, *Entrepreneurial Business and Economics Review*, **5**(3), pp. 149–172, 2017.
2. L. B. Bosman and M. Phillips, Integrating the Entrepreneurial Mindset Into the Engineering Classroom, *IEEE Transactions on Education*, **65**(2), pp. 150–155, 2021.

3. D. F. Kuratko, G. Fisher and D. B. Audretsch, Unraveling the entrepreneurial mindset, *Small Business Economics*, **57**(4), pp. 1681–1691, 2021.
4. J. M. Bekki, M. Huerta, J. S. London, D. Melton, M. Vigeant and J. M. Williams, Opinion: Why EM? The Potential Benefits of Instilling an Entrepreneurial Mindset, *Advances in Engineering Education*, **7**(1), pp. 1–11, 2018.
5. A. Huang-Saad, C. Bodnar and A. Carberry, Examining Current Practice in Engineering Entrepreneurship Education, *Entrepreneurship Education and Pedagogy*, **3**(1), pp. 4–13, 2020.
6. A. Huang-Saad, C. S. Morton and J. C. Libarkin, Entrepreneurship Assessment in Higher Education: A Research Review for Engineering Education Researchers, *Journal of Engineering Education*, **107**(2), pp. 263–290, 2018.
7. S. E. Zappe, S. L. Cutler and L. Gase, A Systematic Review of the Impacts of Entrepreneurial Support Programs in Science, Technology, Engineering, and Math Fields, *Entrepreneurship Education and Pedagogy*, **6**(1), pp. 3–31, 2023.
8. J. D. Novak, Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning, *Instructional Science*, **19**, pp. 29–52, 1990.
9. A. Jackson, E. Barrella and C. Bodnar, Application of concept maps as an assessment tool in engineering education: Systematic literature review, *Journal of Engineering Education*, pp. 1–15, 2023.
10. J. D. Novak and D. B. Gowin, *Learning How to Learn*, Cambridge University Press, 1984.
11. M. M. Martine, L. X., Sunbury, C. M., Schneider, J. A., Hixson and C. Bodnar, Concept maps as an assessment tool for evaluating students' perception of entrepreneurial mind-set, *ASEE Annual Conference and Exposition, Conference Proceedings*, Tampa, FL, June 16–19, 2019.
12. P. Weber, S. Lee and H. Dillon, Benefits of Statics Concept Mapping in Career Cognition, Conference Proceedings, *ASEE Annual Conference & Exposition*, Minneapolis, MN, June 16–29, 2022.
13. M. K. Watson, J. Pelkey, C. R. Noyes and M. O. Rodgers, Assessing Conceptual Knowledge Using Three Concept Map Scoring Methods, *Journal of Engineering Education*, **105**(1), pp. 118–146, 2016.
14. E. Cartwright, M. Ita and K. Kecskemety, Analyzing Various Scoring Methods for Fill-In Concept Maps, 2022 *ASEE Annual Conference & Exposition*, Minneapolis, MN, June 16–29, 2022.
15. A. Franklin, T. Li, P. Jamieson, J. Semlak and W. Vanderbush, Evaluating metrics for automatic mind map assessment in various classes, *Proceedings – Frontiers in Education Conference, FIE*, El Paso, TX, October 21–24, 2015.
16. R. Castles and V. Lohani, A paradigm for comprehensive concept map-based modeling of student knowledge, *ASEE Annual Conference and Exposition, Conference Proceedings*, Austin, TX, June 14–17, 2009.
17. J. Segalas, D. Ferrer-Balas and K. F. Mulder, Conceptual maps: Measuring learning processes of engineering students concerning sustainable development, *European Journal of Engineering Education*, **33**(3), pp. 297–306, 2008.
18. M. Hu, T. Shealy, J. Grohs and R. Panneton, Empirical evidence that concept mapping reduces neurocognitive effort during concept generation for sustainability, *Journal of Cleaner Production*, **238**, pp. 1–12, 2019.
19. A. R. Bielefeldt, First-year students' conceptions of sustainability as revealed through concept maps. *ASEE Annual Conference and Exposition, Conference Proceedings*, New Orleans, LA, June 26–29, 2016.
20. M. K. Watson, E. Barrella and J. Pelkey, Assessment of conceptual knowledge using a component-based concept map scoring program, *International Journal of Engineering Education*, **34**(3), pp. 1025–1037, 2018.
21. J. D. Novak and A. J. Canas, The theory underlying concept maps and how to construct and use them. Florida Institute for Human and Machine Cognition, 2008.
22. M. E. Besterfield-Sacre, J. Gerchak, M. Lyons, L. J. Shuman and H. Wolfe, Scoring Concept Maps: An Integrated, *Journal of Engineering Education*, **93**(2), pp. 105–115, 2004.
23. T. W. Foster, M. A. Shahhosseini and G. Maughan, Advancing diagnostic skills for technology and engineering undergraduates: A summary of the validation data, *Journal of Technology Education*, **28**(1), pp. 2–18, 2016.
24. J. M. T. Walker and P. H. King, Concept mapping as a form of student assessment and instruction, *ASEE Annual Conference Proceedings*, Montreal, June 16–19, pp. 7151–7166, 2002.
25. S. Ferguson and R. W. Foley, Assessment of student learning outcomes and ABET accreditation: A pilot study of fourth-year engineering students using longitudinal concept maps, *ASEE Annual Conference and Exposition, Conference Proceedings*, Columbus, OH, June 25–28, 2017.
26. D. C. Shallcross, Concept Maps for Evaluating Learning of Sustainable Development, *Journal of Education for Sustainable Development*, **10**(1), pp. 160–177, 2016.
27. M. W. Roberts and C. Haden, Assessing student learning of civil engineering infrastructure, *ASEE Annual Conference and Exposition, Conference Proceedings*, New Orleans, LA, June 26–29, 2016.
28. P. Jamieson and J. Eaton, Towards a better graphlet-based mind map metric for automating student feedback, *ASEE Annual Conference and Exposition, Conference Proceedings*, Seattle, WA, June 14–17, 2015.
29. T. Conlon, "Please Argue, I Could Be Wrong": A Reasonable Fallible Analyser for Student Concept Maps. In L. Cantoni & C. McLoughlin (Eds.), *Proceedings of EdMedia + Innovate Learning*, pp. 1299–1306, 2004a.
30. T. Coma-Rosello, A. Aguelo-Arguis, P. Alvarez, C. Sanz and S. Baldassarri, Analysis of Innovative Approaches in the Class Using Conceptual Maps and Considering the Learning Styles of Students, *Revista Iberoamericana de Tecnologias Del Aprendizaje*, **13**(4), pp. 120–129, 2018.
31. H. E. Herl, H. F. O'Neil, G. K. W. K. Chung, R. A. Dennis, J. J. Lee and H. F. O'Neil Jr., Feasibility of an on-line concept mapping construction and scoring system, *American Educational Research Association*, 1997.
32. B. E. Cline, C. C. Brewster and R. D. Fell, A rule-based system for automatically evaluating student concept maps, *Expert Systems with Applications*, **37**(3), pp. 2282–2291, 2010.
33. Y. Lin, A. M. Shahhosseini, M. A. Badar, W. T. Foster and J. C. Dean, Using conceptual mapping to help retain tribal knowledge, *ASEE Annual Conference and Exposition, Conference Proceedings*, New Orleans, LA, June 26–29, 2016.
34. S. Melnik, H. Garcia-Molina and E. Rahm, Similarity Flooding: A Versatile Graph Matching Algorithm, *Data Engineering*, pp. 117–128, 2002.
35. P. Jamieson, Using modern graph analysis techniques on mind maps to help quantify learning, *Proceedings – Frontiers in Education Conference, FIE*, Seattle WA, October 3–6, 2012.

36. A. Franklin, R. Sunderhaus, C. Bell and P. Jamieson, Improved method for creating criterion maps for automatic mind map analysis, *Proceedings – Frontiers in Education Conference, FIE*, Eire, PA, October 12–15, 2016.
37. J. X. Hao, R. C. W. Kwok, R. Y. K. Lau and A. Yan Yu, Predicting problem-solving performance with concept maps: An information-theoretic approach, *Decision Support Systems*, **48**(4), pp. 613–621, 2010.
38. G. P. Jain, V. P. Gurupur and E. D. Faulkenberry, Artificial Intelligence Based Student Learning Evaluation Tool, *2013 IEEE Global Engineering Education Conference (EDUCON)*, Abu Dhabi, UAE, April 10–13, pp. 751–756, 2013.
39. R. Koul, R. B. Clariana and R. Salehi, Comparing several human and computer-based methods for scoring concept maps and essays, *Journal of Educational Computing Research*, **32**(3), pp. 227–239, 2005.
40. R. B. Clariana and E. M. Taricani, The Consequences of Increasing the Number of Terms Used to Score Open-ended Concept Maps, *International Journal of Instructional Media*, **37**(2), pp. 163–172, 2010.
41. V. P. Gurupur, G. Pankaj Jain and R. Rudraraju, Evaluating student learning using concept maps and Markov chains, *Expert Systems with Applications*, **42**(7), pp. 3306–3314, 2015.
42. D. Luckie, S. H. Harrison and D. Ebert-May, Model-based reasoning: Using visual tools to reveal student learning, *American Journal of Physiology – Advances in Physiology Education*, **35**(1), pp. 59–67, 2011.
43. T. Conlon, But is Our Concept Map any Good?: Classroom Experience with the Reasonable Fallible Analyser, *Proceedings of the First International Conference on Concept Mapping*, Pamplona, Spain, September, pp. 159–166, 2004b.
44. J. L. Pruett and E. G. Weigel, Concept Map Assessment Reveals Short-Term Community-Engaged Fieldwork Enhances Sustainability Knowledge, *CBE – Life Sciences Education*, **19**(3), pp. 1–10, 2020.
45. A. J. Cañas, G. Hill, R. Carff, N. Suri, J. Lott, G. Gómez, T. Eskridge, M. Arroyo and R. Carvajal CmapTools: A Knowledge Modeling and Sharing Environment, *First International Conference on Concept Mapping*, *Concept Maps: Theory, Methodology, Technology, Proceedings*, Pamplona, Spain, September 14–17, Editorial Universidad Pública de Navarra, 2004.
46. M. L. Cano Morales and E. Rodríguez Mejía, Github, 2023, https://github.com/EM-Cmap-Scoring-Tool/EM_Cmap_Scoring_Tool
47. The Kern Family Foundation, *About Us*, Engineering Unleashed. https://engineeringunleashed.com/about, Accessed June 2024.
48. E. Barrella, C. Bodnar, M. L. Cano Morales, M. I. Carnasciali, J. Cruz, H. E. Dillon, K. Kecskemety, A. M. Jackson, E. Miskioglu, E. Rodríguez Mejía, EM Concept Map Toolkit. Licensed under CC BY-NC-SA, 2023.
49. V. Braun and V. Clarke, Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology*, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. pp. 57–71. American Psychological Association, 2012.
50. M. B. Miles, A. M. Huberman and J. Saldana, *Qualitative data analysis: a sourcebook of new methods*. Sage Publications, 2014.
51. J. Walther, N. W. Sochacka and N. N. Kellam, Quality in interpretive engineering education research: Reflections on an example study, *Journal of Engineering Education*, **102**(4), pp. 626–659, 2013.
52. D. Stevens, Optimizing Visual Cues in Educational Software, *Augmented Cognition*, pp. 287–303, 2022.

Eduardo Rodriguez-Mejia: Received his BS and MSc degrees in Electronics Engineering from Pontificia Universidad Javeriana, Bogotá-Colombia. Currently acts as advisor for WIE and the research seminar InGenias for the encouragement, participation, and leadership of women in engineering. He is also a scout and is currently pursuing his PhD degree in Engineering Education at Rowan University in the USA.

Martha L. Cano-Morales: MSc. She is currently serving as an adjunct professor in the Department of Electronics at Pontificia Universidad Javeriana, Colombia, while pursuing her doctoral studies in Engineering Education at Rowan University in the United States. Additionally, she acts as an advisor for the Women in Engineering student group at Pontificia Universidad Javeriana.

Alexandra Jackson: PhD student at Rowan University in Engineering Education. She began her research in Rowan's Experiential Engineering Education Department in the Fall of 2019, and has developed interests in entrepreneurial mindset and student development. In particular, she is interested in assessment of entrepreneurial mindset through both quantitative and qualitative methods, and is currently working in both survey and concept map assessment. She was awarded an NSF Graduate Research Fellowship in April, 2022, and hopes to continue her research in entrepreneurial mindset assessment using narrative inquiry.

Elise Barrella: PhD, PE, is the founder and CEO of DfX Consulting LLC and has 20 years of transportation and engineering education research, consulting, and teaching experience. Dr. Barrella's education research investigates how students learn and apply systems thinking in entrepreneurial and sustainable design contexts. She has served as an investigator or evaluator on projects funded by National Science Foundation, Kern Family Foundation, Federal Highway Administration, North Carolina Humanities, and other agencies and foundations. In addition to academic pursuits, Dr. Barrella lends her expertise to local non-profit work focused on infrastructure systems, spatial justice, and design education.

Cheryl A. Bodnar: PhD is a Professor in the Experiential Engineering Education Department at Rowan University. Currently, her research is funded by the National Science Foundation (NSF) and the Kern Family Foundation. Her research interests relate to the incorporation of active learning techniques such as game-based learning in undergraduate classes as well as integration of innovation and entrepreneurship into the engineering curriculum. In particular, she is interested in the impact that these tools can have on student perception of the classroom environment, motivation, and learning outcomes.