DotWrangler: A Method for Assessing Fluency, Originality, and Flexibility of Concept Maps and Diagrams at Scale

Mohamed Ez-zaouia[1,2] [a] and Rubiela Carrillo[3] [b]

[1] *Le Mans Université, LIUM, Le Mans, France*

[2] *Faculty ITC, University of Twente, Netherlands*

[3] *CP Lyon, Villeurbanne, France*

Keywords: Concept Maps, Assessment, Concept Mapping, Diagrams, Creativity, Visualization.

Abstract: Visualization of interrelated ideas and concepts, widely known as *concept maps*, is a ubiquitous technique for knowledge inquiry in many areas, such as learning, design, problem-solving, and creativity. While the effectiveness of concept maps is well documented in research and practice, comprehensive methods and tools for concept map assessments are scarce. Assessing concept maps is challenging, time-consuming, and prone to errors. DotWrangler builds upon previous research and proposes a method to assess concept maps qualitatively and quantitatively at scale. We present a visual assessment authoring tool demonstrating the DotWrangler approach and show its utility through a case study. The utility of DotWrangler is to enable the design of a reliable method of concept maps' assessments, facilitate the execution of assessments at scale, and reduce the burden on instructors during the assessment process.

1 INTRODUCTION

Concept maps are node-link representations, where nodes represent concepts, such as ideas, people, places, events, etc., and links, implicit or explicit, represent relationships among concepts. They refer to a wide range of representations, such as "webs, spider maps, clusters, mind maps, semantic maps, cognitive maps, story maps, diagrams, templates, and graphic organizers" (Hyerle, 2009, p. 37). Substantial research has shown that studying or constructing a concept map can enhance significantly learners' performance (Schroeder et al., 2018). Given their well-documented benefits, concept maps are increasingly used for knowledge inquiry in a variety of educational activities, such as note-taking (D'Antoni et al., 2010), problem-solving (Wang et al., 2018), creativity (Sun et al., 2019), graphic elicitation (Crilly et al., 2006), scaffolding (Chen et al., 2012), to name a few.

While the effectiveness of concept maps is well-documented by researchers, practitioners, and commissions, methods for concept map assessments are not well examined, especially not for large-scale assessment (Ruiz-Primo and Shavelson, 1996). There exists a wide variety of heuristics to score concept maps (Strautmane, 2012). However, existing heuristics are somewhat fragmented and do not provide a comprehensive framework for a reliable assessment (Mcclure and Sonak, 1999).

We identified three challenges in assessing concept maps. Additionally, we identified three barriers to designing reliable assessment methods in existing research. DotWrangler builds upon previous research and proposes a novel method to assess concept maps qualitatively and quantitatively at scale. We designed a visual authoring tool demonstrating the DotWrangler approach and showed its utility through a case study. The utility of DotWrangler is to enable the design of a reliable method of concept map assessment at scale, and reduce the burden on teachers during the assessment process.

2 RELATED WORK

Meaningful and Productive Learning. Concept maps can support meaningful learning of the structures and transformations within a problem/knowledge area versus rote learning (Council, 2012, p. 72). Meaningful learning can foster *produc-*

[a] https://orcid.org/0000-0002-3853-0061

[b] https://orcid.org/0000-0002-2949-4257

Ez-zaouia, M. and Carrillo, R.
DotWrangler: A Method for Assessing Fluency, Originality, and Flexibility of Concept Maps and Diagrams at Scale.
DOI: 10.5220/0012037400003470
In *Proceedings of the 15th International Conference on Computer Supported Education (CSEDU 2023) - Volume 2*, pages 442-450
ISBN: 978-989-758-641-5; ISSN: 2184-5026

tive thinking, which can help learners generate novel solutions to new problems. Rote learning might only enable *reproductive* thinking. Novak et al., (Novak and Gowin, 1984) and others developed this line of thinking by using concept maps as an (external) representation technique to tape into and build upon, learners' (internal) knowledge. Concept mapping cultivates meaningful learning because learners can bridge the gap between new knowledge and relevant knowledge they already have. Concept mapping can be an effective technique to direct users' attention, e.g., signaling, focal points (Crilly et al., 2006); support productive thinking, e.g., creative thinking, reasoning, meta-cognition (Sun et al., 2019; D'Antoni et al., 2010); organize actions, e.g., sequences, processes, scaffolds, (Chen et al., 2012); engage users and increase dwell time (Sun et al., 2019).

Cognitive Scaffolding and Offloading. The effectiveness of concept maps resides in three main factors. First, the act of externalization of concepts and relationships in a visual artifact support thinking and reasoning (Crilly et al., 2006). Externalized ideas can prime one another and generate new ideas and associations (Sun et al., 2019). Second, the simplicity of the technique accommodates users with varying backgrounds (e.g., computer science, social science, liberal art) as well as various needs and uses (learning, problem-solving, creativity, design). And finally, the expressiveness of the concept maps elicits and communicates literal and metaphoric content, structure, and relationships directly.

Existing Assessments. In education, assessment or, as defined: "a systematic method with which students' concept maps can be evaluated accurately and consistently" is important, yet not thoroughly examined (Ruiz-Primo and Shavelson, 1996, p. 581). Assessment can be holistic, relational, and structural. It can combine qualitative and qualitative analyses (Besterfield-Sacre et al., 2004; Carrillo et al., 2017). A main challenge of assessment is that it should be *objective*, *reliable*, and *capture unique insights* into the subjects' knowledge (Mcclure and Sonak, 1999). Traditionally, concept mapping assessments are achieved using posthoc tests, often conducted through questionnaires and essays (e.g, Wang et al., 2018; Chen et al., 2012). Although posthoc tests might facilitate the assessment, they might not be authentic and reliable because the test structure might impose cognitive biases on learners. Furthermore, such tests do not capture learners' differences in structuring and communicating their conceptual knowledge in a subject area.

Other ways of concept map assessments are achieved through scoring heuristics. A review identified more than 42 heuristics and measures (Strautmane, 2012). However, existing heuristics are fragmented; some are context-depend (e.g., "amount of help used" when concept mapping); others are applicable to specific types of concept maps (e.g., "frequency of branching" in tree-based concept maps) (Strautmane, 2012). The result is that only a few heuristics are adopted in practice, namely Novak's heuristics concerning the validity of propositions, hierarchies, and cross-links (Novak and Gowin, 1984). However, Novak's heuristics capture a narrow set of concept maps' features. In addition, concept maps might differ substantially from one learner to another (Hyerle, 2009). Existing heuristics, do not capture fine-grain measures of individual differences among learners. Furthermore, such heuristics are mainly designed to be conducted by users manually, which can be tedious and time-consuming. There is still a lack of comprehensive approaches to assess concept maps reliably and at scale (Mcclure and Sonak, 1999).

We build upon previous research and design a method to assess concept maps' qualities in a flexible manner and at scale. In particular, Fardhila and Istiyono (2019) work was inspiring to us. The authors developed a 10 items instrument to assess creative thinking skills using mind maps for biology subjects. The instrument spans fluency, originality, flexibility, and elaboration. However, the 10 items were designed for manual assessment and are context-dependent. In contrast, we aim to design a generalizable approach that maximizes the assessment's reliability by (1) combining quantitative and qualitative measures and (2) optimizing the assessment consistency using a tagging system. Our approach can support at-scale assessments and assess items from various instruments.

3 ASSESSMENT DESIGN

Following a Design-Based Research (Barab and Squire, 2004), we first report three assessment challenges from our fieldwork. We then report three barriers to designing reliable assessment methods in existing research. And finally, we present our method.

3.1 Field Challenges of Assessments

Given the widespread use of concept maps in education, during the academic year 2021-2022, we designed four activities to engage the students in learning by constructing concept maps. In total, 88 students (N=88) participated in the four activities in individual and group work (see examples in the appendix [1]). Yet, after each activity, we faced three main challenges.

C1: There Is a Lack of Well-Defined Methods for Assessing Diverse Students' Concept Maps. A concept map leverages three main facets: conceptual contents, relationships, and structures. Students' productions differ substantially from one to another in these three facets. There is a wide range of aspects that we can assess about concept maps, whether externalization processes, cognitive processes, and outcomes. Comprehensive measures of the contents, relationships, and structures of a concept map are not well-defined.

C2: Assessing Students' Productions Is Challenging, Time-Consuming, and Prone to Errors. Exploring and making sense of all the concept maps made by the students can be challenging due to the difficulty of maintaining awareness of the overall outcome at the group level (e.g., classroom) and the individual level (e.g., a student). Additionally, a concept map can be challenging to grasp at a glance by someone other than the creator. In addition, we can only allocate a few information items at a time in our working memory for active cognitive processing (e.g., when comparing and contrasting different concept maps). Thus, assessing numerous concept maps at a time is prone to errors. Further, *subjectivity* can easily build up and might lead to overlooking or overseeing aspects of concept maps.

C3: Tools That Relieve Some of the Burdens of Assessing Students' Concept Maps Are Scarce. Teachers might lack the time and resources to objectively design methods and tools that assess students' concept maps. Authoring tools for concept map assessments can support users in the assessment process. However, apart from spreadsheets and ad-hoc analyses, we found no commercial or academic tools to effectively ease the assessment for teachers who might have a lower digital and analytical literacy.

3.2 Barriers to Reliable Assessments

We identified three barriers to designing reliable assessment methods in existing research.

B1: There Are Varying Conceptualizations of "what is a concept map". Researchers draw upon varying conceptualizations for concept maps. This spans: (1) the visual artifacts, (2) the cognitive processes, and (3) the semantics and knowledge organizations. For example, in relation to the visual artifacts, a concept map was defined as "drawing pictures", "visual form" (Sun et al., 2019), and "arrangement of the graphical objects (e.g. proximity, inclusion, and adjacency)" to represent and communicate knowledge (Crilly et al., 2006, p. 7). In relation to cognition, concept mapping is often referred to as a technique to support a wide range of cognitive processes, namely higher-order thinking (D'Antoni et al., 2010), visual thinking (Crilly et al., 2006), spatial thinking (Hou et al., 2016), and creative thinking (Sun et al., 2019). And finally, concept maps are often referred to as techniques to communicate semantics and knowledge organizations (Wang et al., 2018). A concept map carries semantics and organizations, such as hierarchy (e.g., part of, kind of), centrality, similarity, connectedness (e.g., tightly connected information), and ordering (e.g., sequence, process, procedure).

One main challenge is that, often, the assessment differs depending on the researchers' considered conceptualization. For example, when a concept map was designed to support higher-order thinking, such as critical thinking (D'Antoni et al., 2010), the focus was more on assessing the knowledge that students gained and less on the visual artifact itself. However, when a concept map was designed to support visual thinking, such as graphical elicitation (Crilly et al., 2006), the focus was more geared towards the visual artifact.

B2: Concept Maps Have Various Context-Dependent Uses, and So Does the Assessments. Researchers leverage concept maps for various contexts. One context of use is to support comprehension of learning materials (e.g., text passages) through both studying and constructing concept maps (Schroeder et al., 2018). A second use relates to guided and adaptive learning, where concept maps are designed to guide learners in acquiring knowledge incrementally through scaffolding and fading strategies (Chen et al., 2012). Another use relates to capturing learners' knowledge and understanding of a subject for feedback and assessment (e.g, Amadieu et al., 2009). And finally, concept maps are used for graphic elicitation (Crilly et al., 2006), such as brainstorming (e.g, Sun et al., 2019), problem-solving (e.g, Wang et al., 2018), and note-taking (e.g, D'Antoni et al., 2010).

While the usefulness of concept maps for the activities mentioned above is interesting, existing assessment methods follow the context of use, making the assessment's design harder to generalize. For example, when a concept map was designed to support comprehension, the evaluation focuses on learners' understanding, which is usually achieved through posthoc tests (e.g, Wang et al., 2018). However, tests are limited because they do not capture individual differences in structuring knowledge (Mcclure and

[1] Online supplementary examples: http://bit.ly/3mCqiwc

Sonak, 1999). Similarly, scaffolding strategies, are often used to help learners in concept mapping activities (e.g, Chen et al., 2012) and they might ease the assessment because the learners follow a well-defined structure. However, they might hinder learners' creativity and advancement (e.g, Amadieu et al., 2009).

B3: There Is Less Consensus on Measures and Heuristics of "what is a good concept map". Well-known heuristics were proposed by Novak, which relate to the validity and significance of (1) propositions, (2) hierarchies, (3) cross-links, (4) examples, and (5) comparison (to experts' maps) (Novak and Gowin, 1984). Although Novak's heuristics are widely used, they have subtle limitations. They do not capture attributes of the artifacts that communicate meaning (e.g., similarities, relatedness, ordering, prominence, adjacency, proximity, etc.). Additionally, Novak heuristics focus mainly a hierarchical (top-down) concept maps. As the supplementary demonstrates[1], we found that students use varying ways to structure concept maps (e.g., networks, mind maps, grids, etc.). And finally, existing heuristics do not capture critical relational qualities, such as (i) fluency, i.e., ease in generating concepts, relations, and relations' types; (ii) originality, i.e., uniqueness, rarity, the relevance of concepts and associations; and (iii) flexibility, i.e., conceptual categories, themes, depth/breadth of thinking underlying a concept map.

3.3 A New Method for Assessments

To overcome the challenges mentioned earlier, we decided to instantiate the practice of concept mapping as a cognitive and creative activity of externalization of concepts and associations (Crilly et al., 2006). Here, a concept map can be seen as technique of brainstorming (Al-Samarraie and Hurmuzan, 2018). Brainstorming is an act of externalization of ideas and associations that leads to the production of spatial, visual, and conceptual artifacts (e.g., ideas, concepts, designs, diagrams, writings, etc. (Crilly et al., 2006)).

Research into brainstorming as a tool for problem-solving, creativity, and concept generation, has yielded measures to evaluate the results of a brainstorming activity. Primary measures involve the quantity of ideas, quality of ideas, novelty of ideas, resource utilization (e.g., initial ideas), redundancy of ideas, and categorization of ideas, among others (Al-Samarraie and Hurmuzan, 2018). Quantity of ideas, also known as *fluency*, represents the degree of ease in processing inputs, such as understanding a problem, or the degree of ease in producing outputs, such as generating ideas, concepts, or solutions (Thompson et al., 2013). Fluency is widely quantified as the number of ideas generated for a given situation.

Additionally, ideas can have several qualitative attributes. One quality is *originality*, which refers to the pertinence, novelty, and rarity of ideas (Puccio and Cabra, 2012). Originality can be essential to quantify unique, clever, and less frequent ideas but still valuable and appropriate for the subject. Another quality is *flexibility*, which refers to the conceptual categories and shifts in thinking underlying ideas, and indicates heuristics and strategies adopted when resolving a problem or a challenge (Puccio and Cabra, 2012). Qualities of flexibility are primarily the results of thematic analysis of the content. Thus, flexibility can be an umbrella for concept maps' qualities. Qualities can be conceptual, relational, structural, or visuospatial, which can be framed depending on the context. In this view, we can define qualitative and quantitative measures for fluency, originality, and flexibility of concept maps.

• **Fluency Measures.** We quantify three fluency measures for concept maps. *Concept fluency* (CFlue): the number of generated concepts. *Relation fluency* (RFlue) the number of generated relations. *Relation-type fluency* (RTFlue): number of generate relations' types (i.e., unique relations' labels). Fluency measures are quantitatively quantified.

• **Originality Measures.** We quantify five originality measures for concept maps. We do so in two ways. First, we qualitatively quantify originality through novelty, uniqueness, or rarity of ideas. Thus we quantify *concept originality* (COrig): the number of original concepts, *relation originality* (ROrig): the number of original relations, and *relation-type originality* (RTOrig): number of original relations' types (i.e., unique relations' labels). Second, we use Natural Language Processing (NLP) approaches to quantitatively quantify the rarity scores of ideas. We quantify the rarity score of ideas as the sum of the frequency of each idea's stem words. After cleaning up misspellings and abbreviations, we tokenize each idea using 1-gram (one word). We remove stop-words. In NLP, stop words are common words of a language, such as articles and prepositions. We generate the stem of each word using a dictionary of stems. Stemming unifies the wording used for all ideas, which is appropriate for computing the frequency of words in a corpus of ideas. And finally, we compute the rarity score of each idea as the sum of the frequency scores of its stem words. A lower rarity score means that the words used for ideas are unique or less frequent. Using this approach, we quantify *concept-stem originality* (CSOrig): the rarity score of concepts and *relation-stem originality* (RSOrig): the rarity score of relations.

• **Flexibility Measures.** We refer to flexibility as a placeholder for any qualitative measure of a concept map, whether contents, structures, or relations. We use *tagging*, a widely used technique for knowledge organization, categorization, and thematic analysis, to quantify flexibility measures. A tag is a code that we associate with a piece of data. Tags capture insights into the information at hand. The tagging system is often context-depend. In addition, tags can have different weights. We do so by associating a weight multiplier with a tag, which is set to one initially. Therefore, tagging is a flexible way to map out the qualities of a concept map with varying granularity and weights. We can tag a concept map as a whole. We can tag components of a concept map. We can tag fine-grain elements of a concept map, such as concepts and relations. Once a concept map is tagged, we compute quantitative measures of the frequency of tags, whether they are related to a concept map as a whole or its elements. A tag_x *flexibility* is the number of occurrences (frequency) of a tag_x times the multiplier.

4 A VISUAL ASSESSMENT TOOL

We designed an assessment tool using our approach. We derived four design principles to guide our design.

4.1 Design Principles

DP1: Reduce the Burden of Assessing Multiple Concept Maps. The assessment process is a time-consuming task and might require the user to perform several iterations on different graphs. Such a process requires maintaining a considerable amount of information in the working memory, such as information about the graph under assessment and contextual information about other graphs, which is prone to errors. Depending on the need, users need to seamlessly navigate between views that aggregate all concepts and relations in a workspace and views of a selected graph or selected concepts and relations.

DP2: Promote Data Entry and Ability to Modify Information. Data entry and modifying relational information are essential to support powerful analysis, such as natural language processing and tagging. A common use case is to unify the wording, such as abbreviations, the naming of concepts, the naming of relations, etc. Users need to be able to add notes, descriptions, and conceptual tags to concepts and relations while performing assessments. Some data entries need to alter the underlying data. Other data entries (e.g., descriptions) can serve as annotations for later examination, collaboration, or feedback.

DP3: Promote Interactivity and Highlighting. Assessing a vast number of concepts and relations can be overwhelming. Thus, users usually scaffold the assessments of concept maps over several incremental iterations. Users make decisions to navigate and explore further information based on what they are focusing on or interacting with at a given time. Thus, users need to be able to explore information in context while navigating between different views. Users' interactions, such as hovering over, selecting, or searching concepts and relations, should be highlighted in different views. Using coordinated views, users can explore relational content, in a linear (i.e., interactive lists) and non-linear way (i.e., graph view).

DP4: Promote Qualitative and Quantitative Analysis. Qualitative assessments are important because a concept map can differ in several ways. DotWrangler aims to strike a balance between qualitative and quantitative analysis. Quantitative measures, namely *fluency* and *originality* of concepts, relations, and relations types need to be computed automatically. Additionally, all the qualitative analyses need to be achieved using a unified tagging system. We informed tagging based on thematic analysis of content. A tag is a theme that captures some insight into a concept map. Users can use tags to perform various analyses (e.g., conceptual, visual, structural, relational, etc.). Quantitative measures about tags should be computed automatically.

4.2 User Interface and Interaction

Using our prototype learners can create concept maps. Teachers can assess learners' productions. The concept map is rendered on an interactive canvas view that supports data entry and manipulation (Figure 1-(a) —**DP2-3**). We can zoom in/out on the graph using the mouse wheel. We can select multiple concepts or links by holding the shift key down while selecting using the mouse, or brushing on the canvas. We can explore neighboring concepts and relations by hovering over a concept (Figure 1-(e)). We can use a contextual menu to edit, delete, and tag concepts and relations. We can select multiple concepts and tag them simultaneously.

The *Contextual Sider* (Figure 1-(b)), has four main coordinated views: graphs view, tags view, concepts view, and dashboard view. The *Graphs View* (Figure 1-(a)) lists all the graphs in the workspace. We can create, delete graphs, and edits their details (**DP2**). We can open one graph or all the graphs in a

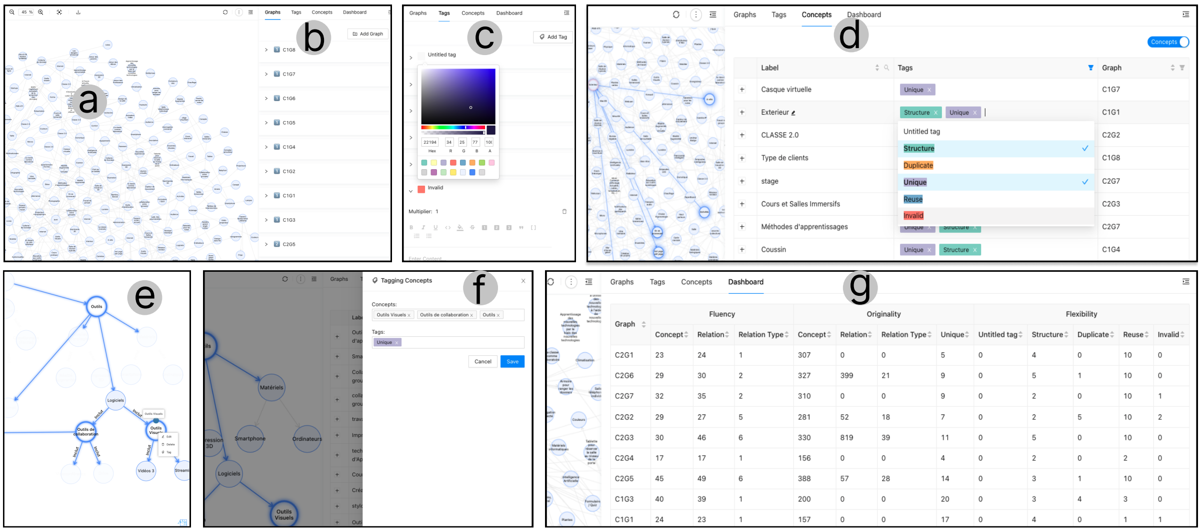

Figure 1: (a) a Graph View, (b) Contextual Sider, (c) Tags View, (d) Concepts View, (e, f) Tagging view, and (g) Dashboard View. We can open one graph at a time, or all the graphs combined. A higher resolution Figure : http://bit.ly/3Td1ze1.

workspace. Similarly, the *Tags View* (Figure 1-(c)) lists all the tags in the workspace. We can create and delete tags. We can edit tags' details, such as label, description, multiplier, and color (**DP2**). The multiplier is the weight coefficient parameter of a tag (equals 1 by default).

The *Concepts View* (Figure 1-(d)) lists all the concepts in the workspace using an interactive table (**DP2-3**). The table enables (1) navigating between different graphs in the workspace, (2) exploring concepts and relations, and (3) tagging concepts and relations (**DP1**). The table lists concepts or relations in rows with three main columns: Label, Tags, and Graph label. We can switch between rendering concepts or relations in the table. We can hover over the labels to edit them. We can expand the rows to explore or edit the descriptions of the concepts and relations using a rich text editor. We can click on a graph label to open it in the graph view. We can sort, filter, and search the table. The table is coordinated with the graph view. When we hover over a row in the table, we highlight the item and its relations with other items in the graph view. Similarly, when we text search in the Label column, we highlight the search results in the graph view. The table enables adding/removing tags to concepts and relations in two ways (**DP2**). First, while exploring the graph using the table, we can tag concepts and relations by selecting tags from the Tags column. The list of the tags is automatically populated from the Tags view. Also, when selecting multiple concepts or relations in the graph view, we can tag them simultaneously using the contextual menu (Figure 1-(e, f)).

The *Dashboard View* (Figure 1-(g)) presents qualitative and quantitative indicators about the concepts and relations for each graph in the workspace, namely fluency, originality, and flexibility (**DP4**). We use flexibility as an umbrella for qualitative and custom indicators needed to evaluate a concept map, which we can achieve using tagging. The Dashboard view is coordinated with other views of the Contextual Sider. Added tags are automatically added to the dashboard. The dashboard is populated and updated automatically.

We implemented our tool using Typescript, Reactjs (UI), G6 (graph), NLPJs (NLP), and supabase (server).

5 CASE STUDY AND FINDINGS

We conducted a case study using our approach.

Participants. The participants were 40 third-year graduate students, of two classes of 21 and 19 students (N = 40), enrolled in the course "Information Systems Modeling, 2021-2022", (gender: [M = 38, F = 2], age: [>25 = 1, 20-25 = 33, <20 = 6]). The participants were French native speakers. They voluntarily participated in the study as part of the course. They signed an informed consent for analyzing their data for research.

Procedure. The activity focused on the conception of a design using collaborative concept maps. We used Miro for the collaborative concept mapping[2]. We

[2] We could not use our tool for concept mapping because it was not designed at the time of conducting the learning activities. We designed our tool thereafter because of the challenges (section 3) that we faced with assessing a large number of concept maps using

provided nine initial concepts (ideas) to the students in each class to stimulate their thinking, referred to as *initial ideas*. Each concept mapping session took about 2 hours. For this case study, we collected the participants' results: 15 concept maps. We manually typed the collected concept maps into our tool for assessment. The two authors assessed collaboratively the 15 concept maps in videoconferencing meetings and wrote down notes about the assessment process.

5.1 Findings: Assessment Process

F1: Explore and Develop Initial Insights. We used the Graph view in the Contextual sider panel to explore graph by graph. We looked for visual and conceptual features, themes, and meanings. We wrote down a few notes about each graph's salient features (in the description field). Initial notes were related to differences between graphs, overall structures, duplicated concepts, the naming of concepts (e.g., abbreviations), and the use of initial ideas —provided to the students for stimulation. As we noticed that students rarely labeled the relationships between concepts, we focused our assessment on the concepts.

F2: Perform a Fine-Grained Review of Ideas. We performed a fine-grained review of concepts for validity using the Concepts view in the Contextual sider panel. We sorted the Label column in the Concepts view alphabetically. We filtered the Graph column to focus on the graphs of each class because the two classes had two slightly different subject statements. We corrected some spelling issues to unify the wording. This step is important for quantitative *originality measures* because we use Natural Language Processing (NLP). We tagged concepts that were not meaningful using DotWrangler tag *Invalid*. Invalid concepts are not included in the *fluency* and *originality* measures, but shown under the *flexibility* measures (see Figure 1-(g)).

F3: Evaluate Ideas. We checked concepts to remove duplicated concepts. Sorting the Label column and filtering by a graph in the Concept view helped spot duplicated concepts. The graph view on the left-hand side was useful for exploring concepts in context as we hover over them in the Concepts View. For each duplicate, we selected one concept to keep and tagged the remaining occurrences using DotWrangler tag *Duplicate*. Similar to *Invalid* tag, duplicated concepts are not included in the *fluency* and *originality* measures, but the number of duplicated concepts is shown under the *flexibility* measures.

F4: Quantify Resource Utilization. We examined whether the students used the initial ideas (concepts) that we provided them for stimulation. We tagged the initial ideas using a new tag *Reuse*. In the Concepts view, we filtered the graphs of each class and searched by labels for the initial ideas. Because the concepts resulting from the search are selected in graph view, we used the textual menu to tag multiple concepts simultaneously.

F5: Quantify Originality and Flexibility of Ideas. We iterated on the concepts in the Concepts view, and we tagged unique, original, and relevant concepts using a new tag, *Unique*. Similarly, we quantified the level of structure, flow, and clarity of each concept map. We created a new tag *Structure* on a scale of 1 (less structured) to 5 (highly structured). For each graph, we tagged up to five selected concepts using the *Structure* tag, depending on the level of structure that we assessed.

F6: Review the Assessment Results. Along the way, we kept reviewing the dashboard which displays automated measures out of the box, namely fluency and originality of concepts, relation, and relation type. The dashboard is automatically updated as we added tags or updated wordings of concepts. The tagging approach and the dashboard make it easier to conduct and capture important aspects of students' concept maps in a flexible manner.

6 DISCUSSION

Reflections and Limitations. Future studies with learners and teachers are needed to further examine our approach. It might prove useful to design a collaborative process (and tool) for our approach so that multiple users can collaborate on the assessment, perhaps with a built-in inter-rater agreement. Future studies can engage with the design and use of a dashboard systematically (e.g., teacher-centered design (Ez-Zaouia, 2020)). Similarly, future studies can combine measures (e.g., using formulas) to build holistic ratings/rankings of students (e.g., Ez-zaouia. et al., 2020). And finally, other measures can be examined, such as comparisons with expert maps, topic mining, and sentiment and emotion analysis (e.g., Ez-Zaouia et al., 2020).

Case Study Applications. We envision that our approach can support the assessment of concept maps (and diagrams in general) for activities spanning various domains of (1) art, design, and creativity; and (2) STEM and non-STEM.

Art, Design, and Creativity. During art, design, or creative problem-solving, students are usually tasked

adhoc methods.

to produce design concepts by analyzing, summarizing, and representing design thinking processes. Node-link diagrams are common productions for this work. Measures of fluency, originality, and flexibility of DotWrangler can make it easier for teachers to evaluate students' work and devise informed interventions.

STEM and non-STEM. In France, for example, the reform of the UBT level (University Bachelor of Technology) that took place in 2022, has put forth a new learning form, referred to as "situation of learning and assessment (SLA)." Writing documents and diagrams are common productions of SLA. We engaged with one UBT teacher to understand how DotWrangler can help them in the assessment process. The teacher shared with us the assessment grid they used in 2022. Four out of eight (4/8) criteria of assessment involved diagramming. This includes "completion of functionalities and diagrams," "readability of UML diagrams", "respect of UML rules" and "overall design concept." The teacher shared with us a total of 101 (N=101) anonymous diagrams. We hypothesize that assessing the 101 diagrams *manually* with respect to the four assessment criteria would be difficult to perform *objectively* and *reliably*. Following a DBR (Barab and Squire, 2004), we plan to conduct studies with teachers to examine how this approach can support assessing concept maps in different contexts.

Conclusion. In this work, we formulated a method to assess concept maps, designed an assessment tool demonstrating our approach, and showed its utility through a case study. We discussed our findings and envisioned future case study applications. We hope our work help spark new ideas for generalizable and reliable methods that reduce the burden and facilitate large-scale assessments of concept maps and diagrams.

REFERENCES

Al-Samarraie, H. and Hurmuzan, S. (2018). A review of brainstorming techniques in higher education. *Thinking Skills and creativity*, 27:78–91.

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., and Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and instruction*, 19(5):376–386.

Barab, S. and Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, 13(1):1–14.

Besterfield-Sacre, M., Gerchak, J., Lyons, M. R., Shuman, L. J., and Wolfe, H. (2004). Scoring Concept Maps: An Integrated Rubric for Assessing Engineering Education. *Journal of Engineering Education*, 93(2):105–115.

Carrillo, R., Renaud, C., Prié, Y., and Lavoué, É. (2017). Dashboard for Monitoring Student Engagement in Mind Mapping Activities. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pages 433–437.

Chen, H.-H., Chen, Y.-J., and Chen, K.-J. (2012). The design and effect of a scaffolded concept mapping strategy on learning performance in an undergraduate database course. *IEEE Transactions on Education*, 56(3):300–307.

Council, N. R. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. National Academies Press.

Crilly, N., Blackwell, A. F., and Clarkson, P. J. (2006). Graphic elicitation: using research diagrams as interview stimuli. *Qualitative research*, 6(3):341–366.

D'Antoni, A. V., Zipp, G. P., Olson, V. G., and Cahill, T. F. (2010). Does the mind map learning strategy facilitate information retrieval and critical thinking in medical students? *BMC Medical Education*, 10(1):61.

Ez-Zaouia, M. (2020). Teacher-centered dashboards design process. In *Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge LAK20*, pages 511–528.

Ez-Zaouia, M., Tabard, A., and Lavoué, E. (2020). Emodash: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies*, 139:102411.

Ez-zaouia., M., Tabard., A., and Lavoué., E. (2020). Progdash: Lessons learned from a learning dashboard in-the-wild. In *Proceedings of the 12th International Conference on Computer Supported Education - Volume 2: CSEDU,*, pages 105–117. INSTICC, SciTePress.

Fardhila, R. R. and Istiyono, E. (2019). An assessment instrument of mind map product to assess students' creative thinking skill. *REID (Research and Evaluation in Education)*, 5(1):41–53.

Hou, H.-T., Yu, T.-F., Wu, Y.-X., Sung, Y.-T., and Chang, K.-E. (2016). Development and evaluation of a web map mind tool environment with the theory of spatial thinking and project-based learning strategy. *British Journal of Educational Technology*, 47(2):390–402.

Hyerle, D. (2009). *Visual Tools for Transforming Information into Knowledge*. Corwin Press, 2nd ed edition.

Mcclure, J. R. and Sonak, B. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, pages 475–492.

Novak, J. D. and Gowin, D. B. (1984). *Learning How to Learn*. cambridge University press.

Puccio, G. J. and Cabra, J. F. (2012). Idea generation and idea evaluation: Cognitive skills and deliberate practices. In *Handbook of organizational creativity*, pages 189–215. Elsevier.

Ruiz-Primo, M. A. and Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6):569–600.

Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., and Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review*, 30(2):431–455.

Strautmane, M. (2012). Concept map-based knowledge assessment tasks and their scoring criteria: An overview. In *Concept maps: Theory, methodology, technology. Proceedings of the fifth international conference on concept mapping*, volume 2, pages 80–88.

Sun, M., Wang, M., and Wegerif, R. (2019). Using computer-based cognitive mapping to improve students' divergent thinking for creativity development. *British Journal of Educational Technology*, 50(5):2217–2233.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., and Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, 128(2):237–251.

Wang, M., Wu, B., Kirschner, P. A., and Michael Spector, J. (2018). Using cognitive mapping to foster deeper learning with complex problems in a computer-based environment. *Computers in Human Behavior*, 87:450–458.