Evaluation and Program Planning 62 (2017) 56–63

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Concept mapping internal validity: A case of misconceived mapping?

Normand Péladeau[a,*], Christian Dagenais[b], Valéry Ridde[c,d]

[a] Provalis Research, 1255 Robert Bourassa Blvd Suite 1604, Montréal, QC H3B 3X3, Canada
[b] Department of Psychology, University of Montreal, Pavillon Marie-Victorin, Room C355, P.O. Box 6128, Centre-ville Station, Montreal, Quebec, H3C 3J7, Canada
[c] Department of Social and Preventive Medicine, University of Montreal School of Public Health (ESPUM), Montreal, Canada
[d] University of Montreal Public Health Research Institute (IRSPUM), Montreal, Canada

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form 15 January 2017
Accepted 5 February 2017
Available online 11 February 2017

Keywords:
Concept mapping
Multidimensional scaling
Hierarchical cluster analysis
Evaluation methods
Internal validity

ABSTRACT

Since the early 1990s, the concept mapping technique developed by William M. K. Trochim has been widely used by evaluators for program development and evaluation and proven to be an invaluable tool for evaluators and program planners. The technique combines qualitative and statistical analysis and is designed to help identify and prioritize the components, dimensions, and particularities of a given reality. The aim of this paper is to propose an alternative way of conducting the statistical analysis to make the technique even more useful and the results easier to interpret. We posit that some methodological choices made at the inception stage of the technique were ill informed, producing maps of participants' points-of-view that were not optimal representations of their reality. Such a depiction resulted from the statistical analysis process by which multidimensional scaling (MDS) is being applied on the similarity matrix, followed by a hierarchical cluster analysis (HCA) on the Euclidian distances between statements as plotted on the resulting two-dimensional MDS map. As an alternative, we suggest that HCA should be performed first and MDS second, rather than the reverse. To support this proposal, we present three levels of argument: 1) a logical argument backed up by expert opinions on this issue; 2) statistical evidence of the superiority of our proposed approach and 3) the results of a social validation experiment.

1. Introduction

Mapping means knowing. At the beginning of the 1970s, Joseph D. Novak of Cornell University developed a technique of Concept Mapping (CM) which made it possible to visualize the relationships among various concepts (Novak, 1990). Results obtained without the help of statistical analysis, were presented in the form of a diagram in which concepts were linked by arrows and the relationships explained in short sentences. Concept maps have been used in several disciplines, particularly in Education and Philosophy, to give a visual representation of knowledge (Kremer & Gaines, 1994).

Towards the end of the 1980s, William M. K. Trochim, from Cornell University, perfected a Concept Mapping technique that combined strategies for qualitative and quantitative analysis and was based on the active participation of interested parties. The technique is designed to help identify the components, dimensions, and particularities of a given reality, to prioritize them, and relate them to one another (Caracelli, 1989; Daughtry & Kunkel, 1993). The concept maps are based on information produced to answer a single question. The method typically involves five steps: (1) The first step is to *formulate the question*. (2) A group of participants is then invited to collectively answer this question by generating statements during a *brainstorming session*. (3) Participants are then asked to sort the topics in piles, creating distinct categories representing an idea or a concept. They also rate each statement in order of importance on a scale of 1–5. (4) The *data analysis step* starts with the creation of a distance matrix between all statements transforming the number of times statements are grouped together in a pile into a distance measure (the more often they appear together, the smaller the distance). A Multidimensional scaling analysis (Kruskal & Wish, 1978) is then applied to this matrix to create a two-dimensional map where the position of all statements tends to reflect, as much as possible, the computed distances between them. Next, a hierarchical cluster analysis (Aldenderfer & Blashfield, 1984) is conducted on the map coordinates to group statements that are close to each other,

* Corresponding author.
E-mail address: peladeau@provalisresearch.com (N. Péladeau).

http://dx.doi.org/10.1016/j.evalprogplan.2017.02.005

Table 1
Distances Between 9 US Cities.

	Boston	NYC	Washington	Miami	Chicago	Seattle	San Francisco	Los Angeles	Denver
Boston	0	206	429	1504	963	2976	3095	2979	1949
NYC	206	0	233	1308	802	2815	2934	2786	1771
Washington	429	233	0	1075	671	2684	2799	2631	1616
Miami	1504	1308	1075	0	1329	3273	3053	2687	2037
Chicago	963	802	671	1329	0	2013	2142	2054	996
Seattle	2976	815	2684	3273	2013	0	808	1131	1307
San Francisco	3095	2934	2799	3053	2142	808	0	379	1235
Los Angeles	2979	2786	2631	2687	2054	1131	379	0	1059
Denver	1949	1771	1616	2037	996	1307	1235	1059	0

forming clusters that represent concepts.[1] (5) The last step involves a second meeting with the participants where they are asked to assess, name and interpret the concept map obtained in the previous step.

Since the early 1990s, the technique has been widely used by evaluators for program development and evaluation. Most program evaluation journals have published articles on projects that have employed this method. Today, over 200 references to this specific technique have been published in peer review journals. This data collection and analysis technique has proven to be very useful for logic model development, outcome evaluation, needs assessment, concept definition, theory creation, instrument development, etc. The method has been proven to be an invaluable tool for evaluators and program planners.

Although the technique of CM has been widely employed, it should be noted that articles published on the subject rarely critically appraise the statistics used in the method. In fact, in our review of 190 articles published in peer review journals from 1989 to 2012, only 12 of them mention the statistical procedures underlying CM, without raising any concerns. This may be due to the fact that most of the authors of these articles are using the technique within the framework of research projects that deal with the advancement of knowledge in a specific field of research and they are not analyzing the method itself. This could also be explained by the fact that the concept mapping (CM) technique developed by Trochim (1989a, 1989b) has been integrated into a software by Concept Systems Incorporated® that runs all the different statistical analyses automatically. Still, the reflexive analysis must be a quality of researchers in general, particularly in the field of evaluation.

We believe that some methodological choices made at the inception stage of the technique and still in use today were ill informed, producing maps of participants' points-of-view that were not optimal representations of their reality, making them unnecessarily harder to interpret. For example, several authors reported the difficulty participants had in understanding and naming clusters and the necessity of removing from the clustered solution, statements with no obvious connection to the others (e.g. Campbell & Salem, 1999; Dagenais, Ridde, Laurendeau, & Souffez, 2009; Gol & Cook, 2004; Mercier, Piat, Péladeau, & Dagenais, 2000; Rosas & Camphausen, 2007; Sutherland & Katz, 2005).

The aim of this paper is to propose an alternative way of conducting the statistical analysis to make the results easier to interpret and the technique even more useful for evaluators and program planners. We do not challenge the whole CM method, but put into question the statistical analysis performed to create maps (step #4). We propose a change that should result in the creation of more coherent clusters of statements and thus facilitate their interpretation and naming. We will establish that the main problem lies in the statistical analysis sequence by which multidimensional scaling (MDS) is first being applied on the similarity matrix, followed by a hierarchical cluster analysis (HCA) on the Euclidian distances between statements as plotted on the resulting two-dimensional MDS map. We will also demonstrate that this problem is exacerbated by two interrelated factors, namely the high stress values typical of CM studies and the initial choice of restricting the number of dimensions extracted with MDS to two dimensions only.

As an alternative, we will advance the idea that the clustering process must be performed on the original similarity (or distance) matrix rather than on the one obtained through a MDS transformation. In other words, HCA should be performed first and MDS second, rather than the reverse as suggested by Trochim (1989a, 1989b). To support this proposal, we will present three levels of arguments. First, we will present a logical argument backed up by expert opinions on this issue. Second, we will attempt to present statistical evidence of the superiority of our proposed approach (HCA → MDS) over the approach implemented by the Concept System software (MDS → HCA). Third, we will present the results of a social validation experiment that demonstrates the superiority of our proposed approach in representing the participants' points-of-view. Recommendations and suggestions for further research and for alternative ways of producing concept maps will then be presented.

2. The logical arguments

Multidimensional scaling is a technique that attempts to represent a matrix of distances (or dissimilarity) between multiple data points on a multidimensional Euclidian space as accurately as possible. It has its origin in psychometrics, where it was developed to identify the underlying dimensions used by people to judge the similarity of a set of objects (Torgerson, 1952). It has been used since then in a variety of fields, especially marketing, but also sociology, political sciences, as well as physics and biology (e.g., Young and Hamer, 1994). In CM, the objects are the statements generated in the brainstorming session, while the similarity is obtained through the grouping of statements by participants into piles during the classification and rating step. A common example given to illustrate MDS is to start with a matrix of distances between cities like the one used in Borgatti, Everett, and Johnson (2013) and reproduced in Table 1.

Applying MDS on such a matrix and plotting all nine cities on a two dimensions Euclidian plane results in a map that will represent the positions of those cities relative to each other with such a precision that it is possible to overlap a geographic map of the United States and find the cities located very close to their actual location. Since only relative distances between these cities are being used, such a map may have to be rotated and sometimes flipped to reproduce a US map as we are used to seeing it (north on top and the east on the right).

In CM, the co-occurrence of statements in piles created by participants resulting from the sorting process is transformed into

[1] For more details, see Kane and Trochim (2007) and Trochim (1989a, 1989b).