Evaluation and Program Planning 60 (2017) 277–283

Contents lists available at ScienceDirect

Evaluation and Program Planning

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

Use of multiple cluster analysis methods to explore the validity of a community outcomes concept map

Rebecca Orsi[a,b,*]

[a] School of Education, Colorado State University, United States
[b] Social Work Research Center, School of Social Work, Colorado State University, United States

ARTICLE INFO

Article history:
Received 26 July 2016
Accepted 22 August 2016
Available online 28 August 2016

Keywords:
Concept mapping
Multidimensional scaling
Cluster analysis
R statistical software
Validity

ABSTRACT

Concept mapping is now a commonly-used technique for articulating and evaluating programmatic outcomes. However, research regarding validity of knowledge and outcomes produced with concept mapping is sparse. The current study describes quantitative validity analyses using a concept mapping dataset. We sought to increase the validity of concept mapping evaluation results by running multiple cluster analysis methods and then using several metrics to choose from among solutions. We present four different clustering methods based on analyses using the R statistical software package: partitioning around medoids (PAM), fuzzy analysis (FANNY), agglomerative nesting (AGNES) and divisive analysis (DIANA). We then used the Dunn and Davies-Bouldin indices to assist in choosing a valid cluster solution for a concept mapping outcomes evaluation. We conclude that the validity of the outcomes map is high, based on the analyses described. Finally, we discuss areas for further concept mapping methods research.

1. Introduction

Literature in the field of theory-based evaluation (Chen, 1990) has suggested that statistical tools may be used in the development of program theory, particularly in the area of outcomes evaluation. Also, such literature suggests (Leeuw, 2003) that three primary methods for reconstructing program theories can be used: a policy-scientific approach, a strategic assessment approach and an elicitation approach. The elicitation approach draws on ideas from cognitive and organizational psychology, and Leeuw notes that Trochim's concept mapping method (1989) is an example of the elicitation approach. Programmatic outcomes can be understood as one domain in the context of Chen's (1990) six-domain framework for program theory. Alternatively, outcomes can be regarded as part of a simpler program theory framework based on a linear logic model to represent inputs, outputs and outcomes (University of Wisconsin-Extension, 2002; W.K. Kellog Foundation, 2001). The purpose of this study is to explore the utility of concept mapping (Trochim, 1989) as a tool for articulating outcomes from a complex social intervention. Specifically, we examine variations of the basic concept mapping process and how such variations can assist evaluators in validly articulating programmatic outcomes.

Concept mapping was first presented as a cohesive evaluation tool more than 25 years ago (Trochim, 1989). Briefly, the six steps for the process are: (1) *preparation*, which includes selection of participants and a brainstorming focus statement, (b) *generation* of focus response statements via brainstorming, (c) *structuring* of statements via sorting and rating, (d) *concept mapping analysis* (also known as *representation* of statements on a map), (e) *interpretation* of maps and (f) *utilization* of maps (Kane & Trochim, 2007). The current study constitutes an in-depth look at step four: concept mapping analysis. It will explore several alternative cluster analyses in an effort to produce the most valid representation of participant responses as possible. The discussion is primarily methodological with a focus on how valid results were obtained. Readers interested in substantive concept mapping outcomes as they relate to the social program in question are referred to Orsi (2014).

Shadish, Cook, and Campbell (2002) note that validity refers to whether or not an inference or knowledge claim or proposition is approximately true. They further note that validity is properly understood as a property of inferences, not a property of methods. Thus, when considering validity and concept mapping, the question is not whether concept mapping is a valid method, but whether concept mapping produces valid knowledge propositions for a specified situation or context. As noted briefly above, the fourth step of concept mapping is the *representation of statements* on a map. This involves conducting one or more cluster analyses.

* Correspondence to: School of Social Work, Campus Delivery 1586, Colorado State University, Fort Collins, CO 80523-1586, United States.
E-mail address: becky.orsi@colostate.edu (R. Orsi).

http://dx.doi.org/10.1016/j.evalprogplan.2016.08.017
0149-7189/

Ward's hierarchical method of clustering is commonly used, although other types of cluster analysis are available. Furthermore, cluster analysis typically involves judgement, as more than one set of clusters may represent concept mapping data in a meaningful way. Decisions about what type of cluster analysis to use and how to select a final set of clusters have implications for the validity of the evaluation results produced. Therefore, in the current study, we ask whether or not using several clustering methods (e.g. agglomerative, divisive or non-hierarchical) and calculating clustering fit metrics serves the purpose of better enabling valid articulation of programmatic outcomes via concept mapping.

A few past studies have examined the validity of concept mapping results. Dumont (1989) considered whether maps formed by multidimensional scaling (MDS) are a valid representation of a participants' individual conceptualizations for a construct of interest. This study concluded that the maps were not entirely valid representations of participants' experiences. Dumont did not, however, entirely use Trochim's methodology. Cacy (1996) produced three concept maps relating to the nature of doctors' practice-based research networks. Participants were asked to choose a concept map that "makes the most sense" to them, based on professional experience (1996, p. 95). Results showed that faculty doctors consistently chose a community practitioners' concept map, rather their own map. Community practitioners did not consistently pick any of three possible maps. Cacy concluded that the study did not provide any certain evidence for the validity of concept maps (1996). More recently, Rosas and Kane (2012) examined possible measures of representational validity from a wide variety of concept mapping studies, focusing primarily on internal representational validity. Measures discussed include acceptable values for the stress statistic and also configural similarity. These measures focus on validity understood as "determining the overall match between the participant-structured input and the mathematically generated output" (Rosas & Kane, 2012, p. 237). Internal representational validity was found to be good across the studies examined.

Although the focus of the current study is on validity from a quantitative and statistical point of view, concept mapping also includes qualitative data. Therefore, alternative considerations of validity are also appropriate. For example, Jackson and Trochim (2002) take a perspective from content analysis and suggest that collective conceptualizations from concept mapping are potentially more valid than are results from other methods which rely substantially on the researcher's role or interpretations. However, with Krippendorf, they note that because concept mapping deals with social constructions, " . . . there is really no way to establish a standard by which to judge the degree of error" in the expression of participants' perceptions (2002, p. 330; Krippendorf, 1980). Also, Orsi (2014) discusses the *credibility* of outcomes using data from the current study and finds that outcomes are indeed credible to program participants who review concept maps. It remains the case, however, that no single and universally accepted measure of validity for concept mapping yet exists (Bedi & Alexander, 2009; Trochim, Cook, & Setze, 1994).

2. Methods

The context for applying concept mapping in the current study was a grassroots community organization in a western United States city. At the time, the organization was working to address several community issues including access to healthcare, local and state education reforms, college access, citizenship, and neighborhood safety. To create an outcomes map for the community organizing program, twenty-one grassroots community leaders from schools and local religious congregations participated. The twenty-one participants generated responses to the following focus statement: "Think about yourself, your family, your child's school, your church and your neighborhood. When [our organization] does community organizing, this is what happens: _____." This resulted in 125 statements for sorting and rating. However, Kane and Trochim (2007) suggested limiting the number of statements for sorting to about 100. Experience from the author's pilot study suggested there should be even fewer statements to reduce the time for sorting, a task which pilot participants stated was burdensome. Therefore, statements reflecting a similar theme or topic were combined to remove redundancy and to reduce the number of sort statements. The final list of brainstormed statements numbered 89. This is consistent with the average number of statements per study (about 96) found by a recent overview of concept mapping studies (Rosas & Kane, 2012).

The next step of the concept mapping process was the sorting of statements. In total, twenty-one sorted solutions for the 89 statements were provided by participants. These formed the data set for multidimensional scaling and cluster analyses. Table 1 displays information which characterizes the entire group of participants in terms of experience with community organizing, personal education level, age, and childcare responsibilities. Detailed results concerning community organizing outcomes are reported elsewhere (Orsi, 2014). In the current study, we focus on methods to ensure validity and on providing details from statistical analysis using the R statistical software environment (R Development Core Team, 2011).

2.1. Data preparation

As noted by McLinden (2013), in order to perform multidimensional scaling (Bartholomew, Steele, Moustaki, & Galbraith, 2002; Kruskal & Wish, 1978) and cluster analysis (Johnson & Wichern, 2007; Kaufman & Rousseeuw, 1990), it is necessary to reformat data from the sort solutions into a matrix for analysis in the R statistical package. Each response statement was numbered from 1 to 89. Each participant's sort results were transcribed into a Word document, then cut, pasted and edited in the R package as a vector and finally, transformed from a vector into a symmetric binary

Table 1
Participant Characteristics (n = 21).

Characteristic	
Experience with organizing	
1 year or less	5%
2–3 years	33%
4 or more years	62%
Committee affiliation	
Congregation	72%
School	33%
Education level	
No HS diploma	14%
HS diploma	33%
2-year degree	5%
4-year degree +	43%
Missing	5%
Age	
Under 35 years	19%
35–50 years	33%
Over 50 years	48%
Caring for children?	
Yes	38%
No	62%

Note: Percentages of committee affiliation do not add to 100% because some participants are affiliated with both a congregation and school organizing committee.