Evaluation and Program Planning 60 (2017) 186–193

Contents lists available at ScienceDirect

Evaluation and Program Planning

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

A systematic review of concept mapping dissertations

James P. Donnelly, Ph.D

Institute for Autism Research, Department of Counseling and Human Services, Canisius College, 2001 Main Street, Buffalo, NY, 14208, USA

ARTICLE INFO

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

Keywords:
Concept mapping
Dissertations
Systematic Review

ABSTRACT

The purpose of this study was to identify, summarize, and synthesize all doctoral dissertations completed using Trochim's concept mapping methodology between 1985 and 2014. A comprehensive search produced a set of 108 eligible dissertations; of which 104 were available as full-text or hard copy. The studies were coded on 77 variables, which were summarized in descriptive analyses. The dissertations were conducted in a wide variety of topic areas and completed at 35 different universities in the US and Canada. On comparable variables, the results were similar to two prior syntheses (Trochim, 1993; Rosas & Kane, 2012). The mean multidimensional scaling analysis stress value for 96 concept maps was 0.26 with a standard deviation of 0.05. Cumulative rates of dissertation completion and resulting citations of the dissertations and any resulting articles were plotted over the study period. Reliability and validity were considered in most studies but formally assessed in a minority of cases. The review concludes with a summary of findings and thoughts about future studies.

1. Introduction

Every day doctoral students and their advisors discuss topics and methods. Such a conversation happened one day about 30 years ago on the campus of Cornell University, when a doctoral student named Dorothy Torre discussed her topic idea with her advisor, Bill Trochim (Kane & Trochim, 2007). That first conversation turned into a quest, initially through trial and error, to find a way to define a complex construct. Ultimately, the process resulted in Professor Trochim's invention of Concept Mapping (CM). Coincidentally, the construct Torre wanted to study was empowerment. Since then, CM has empowered scores of students to complete dissertations and theses. On the occasion of this second special issue on the method, this paper will present a systematic review of dissertations that have used Trochim's concept mapping methodology since the original study.

The dissertation represents a unique facet of public knowledge, often referred to as part of the "grey literature" because it lies separate from commercially controlled, peer reviewed publications, but in an important domain with other kinds of high quality reports necessary for comprehensive synthesis (Andrés, 2009; Augur, 1998). It is now generally accepted that a comprehensive systematic review is not complete without a search of the grey literature (Hopewell, Clarke, & Mallett, 2005). The doctoral thesis is also unique because of its key role in the training of researchers, the possibility of methodological evolution through variation and innovation, and the quality control of doctoral committees. Dissertations generate all sorts of important and highly variable outcomes in terms of publications, careers, programs, grants, citations and other products, such as new measures. Theoretically, the file drawer problem in dissertations is much smaller than other literatures because all dissertations are part of the public record of degree granting institutions, which generally require students to submit their final dissertation to the UMI or another publically available database.

1.1. The first two concept mapping dissertations

The first concept mapping dissertation to be completed was written by Rhoda Linton at Cornell with Professor Trochim's supervision (Linton, 1985). This study was built on Trochim's (1985) extension of Campbell's Pattern Matching theory (1966). It was also a timely study in the context of the emerging area of Program Theory (Bickman, 1987; Chen & Rossi, 1984). The substantive focus of the study was construction and evaluation of a theory of an employer-sponsored child-care program. The study included a ladder graph to examine the match of predicted and observed effects, with a correlation to represent the overall degree of pattern match. This type of analysis became generally available to researchers using the CM software that Trochim went on to develop: The Concept System.

Simultaneously, Torre was conducting her study, entitled "Empowerment: Structured Conceptualization and Instrument

E-mail address: donnell6@canisius.edu (J.P. Donnelly).

http://dx.doi.org/10.1016/j.evalprogplan.2016.08.010
0149-7189/

Development" (1986). It was an ambitious undertaking beyond the implementation of the new methods because the study goals included development of an instrument, including initial psychometrics. Torre's study made use of early Concept System software developed by Trochim, along with multidimensional scaling (MDS) via ALSCAL in SAS and the SAS VARCLUS procedure to generate the hierarchical cluster analysis. The procedures utilized in the first two dissertations largely set the template for future dissertations and reflect the methods that became standard following the first *EPP* special issue. While the procedures in use today are very similar to those evident in the first studies, there were some that seem not to have been repeated in future studies. For example, in addition to study of item and cluster meaning, Torre examined the correlations between clusters, sorting them into a table of positive and negative coefficients to aid understanding of cluster relationships.

1.2. Evolution of methodology and the "Constructs in students' heads"

It was only four years between Linton's dissertation defense and the appearance of the 1989 *EPP* special issue on concept mapping. Clearly the work with his students was ongoing and in the forefront of Professor Trochim's mind as he described Concept Mapping in the first paper of the special issue:

> "Consider the plight of the graduate student who needs to define the major constructs for a dissertation project. While all of the texts on research say that it is important to define constructs, there is no concrete advice given on *how* to articulate a conceptual framework. The concept mapping approach views concept definition as a measurement task – much like that of developing a scale" (1989a, p. 109).

In fact, the problem of conceptualization has always been, and will always be, the first issue the doctoral student faces in constructing a dissertation (Heppner & Heppner, 2004). Concept mapping may be a unique method, but is not the first approach to defining constructs evident in the dissertation literature. Prior dissertations have included some similar methods, including multidimensional scaling of similarities (e.g., Messick, 1954). Writing about Samuel Messick's dissertation, Jackson reported the following about what he learned about the study of "constructs in peoples' heads" Jackson (2002, p. 3):

> "I think about how Sam Messick strongly influenced me in the development of my thinking about constructs and construct validity by introducing me to the quantitative analysis of judgments of psychological similarity. Messick convinced me that multidimensional scaling, regarded by some as an arcane method for evaluating psychological judgments, had the potential to permit the representation of the constructs of ordinary people as projections on dimensions in a geometric space. This in turn provided an orderly, rigorous way of measuring and thinking about people's constructions of important entities in their psychological world"

Indeed, Messick showed that MDS could be used with confidence beyond the laboratory in examination of complex aspects of thinking and behaving, such as attitudes. Trochim's concept mapping procedures subsequently connected MDS to cluster analysis, pattern matching, and other procedures that have

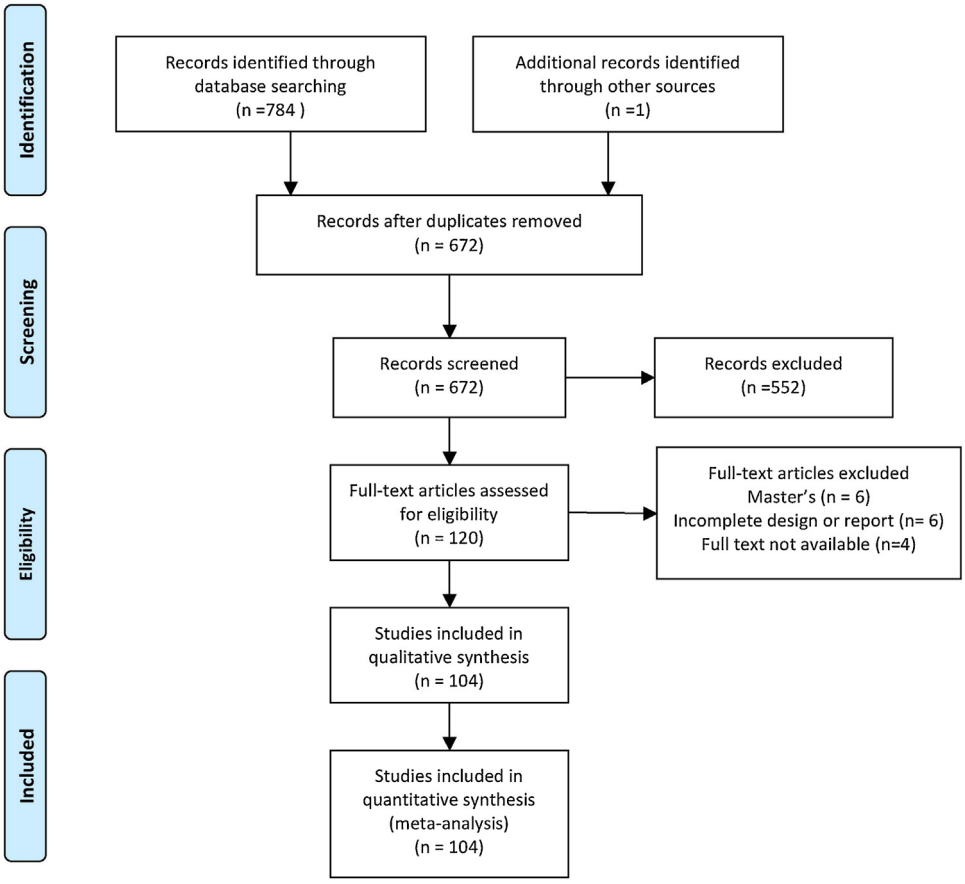

Fig. 1. PRISMA Flow Diagram.

extended the researcher's ability of identify detailed aspects of the shared conceptualization of people of all kinds. Similar to Jackson's comment on the usefulness of MDS, concept mapping can certainly be described as an orderly and rigorous way to obtain data on the way people think about things.

1.3. Purpose of the study

The purpose of this study is to identify, summarize, and synthesize all doctoral dissertations completed using Trochim's concept mapping methodology. The review is systematic and largely descriptive. It may add to general knowledge about the method, as well as provide a point of comparison with the recent Rosas and Kane (2012) *meta*-analysis of 69 studies and the more remote but widely cited paper by Trochim (1993) that integrated 33 concept maps and provided guidance on assessment of CM reliability.

2. Method

2.1. Search strategy

The ProQuest Dissertation Database was utilized as the primary source of dissertations. A secondary search was conducted with Google Scholar. In addition, the list of 66 dissertations listed in Kane and Trochim (2007) was used as a check for inclusiveness. Multiple searches of dissertation titles, abstracts and full-text using key terms "concept mapping" and "structured conceptualization" in combination with author designations (all including Trochim) were conducted. The search results are presented in Fig. 1, which employs the PRISMA template (Moher, Liberati, Tetzlaff, & Altman, 2009). The primary inclusion criterion was that the study used and cited Trochim's Concept Mapping/Pattern Matching (CM/PM) methodology, including MDS and hierarchical cluster analysis (HCA) at a minimum. Master's theses found in the ProQuest database ($n = 6$) were excluded to maximize homogeneity in the document set. Six dissertations did not complete a full CM study or did not provide enough information to code fully. Missing data included such things as details on conduct of standard CM/PM steps, number of participants by step, minimal or inconsistent reporting of results, and other basic study information. Full-text was available for 104 eligible studies utilized in the analysis (11 hard copy; 93 pdf).

2.2. Coding scheme

Studies were coded on 77 characteristics of the study context, methods, and results. Study context characteristics included year, area of study, institution (public vs. private, geographic location, academic department) and so on. Method characteristics included participants in each phase (sorters, raters, interpreters), data collection modality (in-person, web-based, mail, and various combinations), statistical procedures (including cited software), and statement statistics (number of initial and final items). Citation counts for dissertations were obtained from both the ProQuest database and Google Scholar. In most cases the counts matched, but in some cases Google Scholar listed a higher number. The higher numbers were used based on the continuous updating in Google Scholar's automated system. Articles published from dissertations were also obtained, with care to exclude publications that may have been authored by the student on a similar topic, but not based directly on the dissertation. Statistical variables summarized include stress values and any reliability or validity indicators. Finally, the study database included notes on methodological problems and innovations, typically described in greater detail in the dissertation than in more concise kinds of reports such as journal articles or published abstracts.

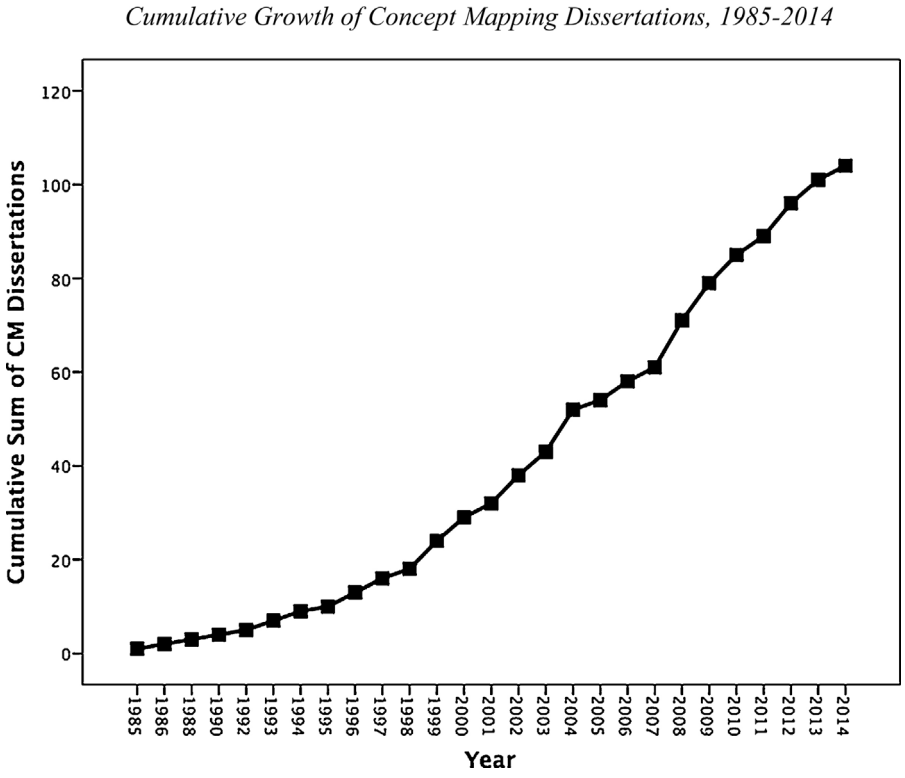

Fig. 2. Cumulative Growth of Concept Mapping Dissertations, 1985–2014.

2.3. Data management

Several data management tools were utilized, including ProQuest's My Research and Flow programs, Adobe Acrobat Professional and SPSS (v. 22). Despite review by multiple readers, dissertations inevitably include some errors. In this study, errors that could be corrected with confidence (e.g., year of a key Trochim citation, decimal points in a stress value) were changed. In cases where a correct replacement could not be made, variables were coded as missing.

3. Results

3.1. Descriptive summary of the dissertations

The descriptive summary of coded dissertation characteristics includes items about the source documents as well as their academic origins. The set of 104 dissertations averaged 208.5 pages (*SD* = 73.92), with a range of 64–405 pages. The total number of pages was 21,689. Concept mapping dissertations were conducted at 35 institutions in the USA (*n of studies* = 91, 87.5%) and Canada (*n of studies* = 13, 12.5%). Of the 35 institutions, three accounted for 44 (42.3%) of the studies. These included Professor Trochim's university, Cornell (*n* = 19), the University at Buffalo (*n* = 14, disclosure: my former institution) and the University of Alberta (*n* = 11). The doctoral degrees obtained included 99 PhD, 2 EdD, 2 PsyD, and 1 DrPH. Fig. 2 shows the cumulative growth in CM dissertations between 1985 and 2014. The largest number in any single year was 10 in 2008. By decade, the dissertations were completed as follows: 1980–89: 3, 1990–99: 21, 2000–09: 55, and from 2010 until 2014: 25.

Areas of study are shown in Table 1. In most cases, the code for area of study was apparent from the title page or other front matter of the document. In some cases, similar areas were combined. This included one study listed as Human Services that was combined with the Social Work studies. The single Architecture study was merged with the single Civil Engineering study. Business and management studies were also combined. The most frequently occurring areas of study included Psychology (*n* = 38, 36.5%), Education (*n* = 23, 22.1%), and Evaluation (11, *n* = 10.6%).

3.2. Dissertation purpose

Study purpose was extracted verbatim from each document. The content areas described in the purpose statements are extraordinarily diverse, and probably suitable for a separate content analysis. But all are consistent with the method's purpose of structured conceptualization. Purpose statements typically present a primary focus, which served as the basis of the student's CM analysis. In some cases, a more complex design was described, in which concept mapping was a sub-study in a sequence leading to a follow-up sub-study. The vast majority (*n* = 96, 92.3%) of studies focused on conceptualization of a domain described as previously unstudied or understudied. The balance included five studies whose primary purpose was measure development and three that focused on some aspect of research design methodology. In 18 cases (17.3%), the student conducted multiple map studies with separate participants. All of these studies examined similarities and differences in perspectives on the same concept by different groups.

Table 1
Area of Study.

Area	Number	Percent
Psychology	37	35.6
Education	23	22.1
Evaluation	11	10.6
Health	10	9.6
Social Wk/Human Svcs	9	8.7
Business/Management	5	4.8
Gender Studies	3	2.9
Environmental	3	2.9
Architecture/Civil Engineering	2	1.9
Total	**104**	**100**

3.3. Participation

Participation was coded in a number of ways, including reports of sampling strategies, IRB review, data collection modality for each phase, and number of participants in CM phases. More than half of the dissertations did not provide a total sample size across all phases. In these cases, the rate of overlapping participation between phases was not reported. A few studies were very clear and precise with reports of participation, but it was not possible to calculate participation rates for each phase for most of the studies (i.e., number invited or consented/number completed).

Sampling strategies were initially extracted verbatim and then recoded to a smaller number of categories. Given the frequency of conceptualization as the primary study goal, it is not surprising that purposive sampling strategies were by far the most frequently employed (83/104 cases, 79.8%). Most of these described specific inclusion criteria. Nine studies reported using random selection methods (8.7%), seven used multiple methods (e.g., purposive followed by snowball), four (3.8%) used samples of convenience and one (1%) used a pure snowball sampling strategy. Of the 104 studies, 60 (57.7%) specifically discussed IRB review and approval. Of these, 40 reported on review and approval, but not a specific level of review. Twelve studies reported that they received an exemption from IRB oversight and three specifically discussed full board review. Not surprisingly, attention to IRB review appeared to increase as time goes on, with 79.1% (19/24) of the studies completed in the last five years (2010–2014) reporting IRB review.

Table 2 summarizes participation in Brainstorming, Sorting, Rating and Interpretation. The rates of participation in each phase are as expected, with the greatest participation in brainstorming and least in interpretation. Variability in brainstorming, sorting and rating is striking. For example, the minimum coded in the brainstorm phase was one, with a high of 555. This is largely a reflection of the data collection modality. For example, the case of a single brainstormer was a study in which the author obtained items from prior literature, and the case of 555 brainstorming participants utilized an Internet-based national sampling strategy.

Only the first rating phase was coded, but it should be noted that 23 studies included a second rating (mean *n of participants* = 39.17, *SD* = 30.12) and five included a third rating scale (mean *n of participants* = 26.8, *SD* = 15.39). Preparation of the focus prompt was typically accomplished by one person (the student), or with a small group of experts (e.g., dissertation chair and/or community-based experts). Less than 10 studies reported a formal pilot test of the focus prompt as part of the design. In a small number of studies, utilization was readily apparent because the CM results were a sub-study in a sequence. The sequence of studies in these cases was typically development of a measure or a survey employed in a descriptive as well as outcome studies. In a little over half of the cases (*n* = 55, 52.9%) utilization was described in theoretical terms as part of the study justification and in the discussion section implications, but the specific number of participants in interpretation was not reported.

Participation in CM phases by modality is presented in Table 3. These categories were reduced from the original codes, which described the modality more specifically. In total, 13 different methods or combinations of methods were utilized in

Table 2
Participation in CM Phases.

Summary	Brainstorming *n*	Sorting *n*	Rating *n*	Interpreting *n*
Mean	48.84	27.39	34.90	8.63
Median	31.50	20	26	4
SD	78.16	12.81	30.84	16.46
Min/Max	1/555	5/152	0/152	1/112
Number (%) not reported	10 (9.6%)	8 (7.7%)	17 (16.3%)	55 (52.9)

Table 3
Distribution of Data Collection Modalities (n = number of studies).

Modality	Brainstorming n(%)	Sorting n(%)	Rating n(%)
Internet	14 (13.5)	17 (16.3)	17 (16.3)
In person	49 (47.1)	49 (47.1)	43 (41.3)
Mail or Email	8 (7.7)	14 (13.5)	16 (15.4)
Mixed	29 (27.9)	23 (22.1)	2 (1.9)
Document-based	4 (3.8)	N/A	N/A
Not Reported	0	1 (1)	22 (21.0)
Phase Not Included in Design	0	0	4 (3.8)

brainstorming. Seven methods were used in sorting and six were reported in rating procedures. By far the most frequently used method across all three phases was an in-person procedure, with nearly half of the brainstorming and sorting done face to face (47.1%) along with 41.3% of the rating procedures. The combinations of modalities were recoded to the mixed methods category. In some cases, the mixing of methods was planned in advance (e.g., document-based followed by a person-based method such as a group setting), but in other cases students described adaptation of methods as a result of disappointing response rates or requests from participants (e.g., some requested a paper-based method instead of Internet-based). The first Internet-based study was completed in 1999, and this mode has gradually grown in usage since then.

3.4. Analytic software

The software used for the primary CM analyses of multidimensional scaling (MDS) and hierarchical cluster analysis (HCA) was reported in most dissertations (MDS missing = 7, HCA missing = 9). A version of The Concept System (from early incarnations to the most recent web-based version) was used in the MDS analysis of 80 dissertations (76.9%) and the HCA analysis in 81 studies (77.9%). Other MDS programs included SPSS (5, 4.8%), SAS (5, 4.8%), R (5, 4.8%), and Anthropac (2, 1.9%). HCA was accomplished with SPSS in six studies (5.8%), SAS in four studies (3.8), and R in four studies (3.8%). All of the studies that reported Bridging Values used some version of the Concept System. Similarly, the Concept System was frequently used to produce various kinds of maps (Point, Cluster, Cluster Rating, Item Bridging, Spanning, etc.). Pattern matching, built into the Concept System, was employed in exactly half of the studies ($n = 52$, 50%). More recent additions to the Concept System including Go Zones and Spanning analysis were less often reported (Go Zone $n = 17$, 16%; Spanning $n = 5$, 4.8%). Table 4 summarizes the kinds of CM analyses reported. All studies that conducted reliability or validity analyses beyond pattern matching used SPSS, SAS or R.

Table 4
Frequency of Graphic and Supplementary Analysis.

Analysis	n (%) Reporting
Point Map	78 (75%)
Cluster Map	84 (80.8%)
Cluster Rating Map	47 (45.2%)
Bridging	53 (51%)

3.5. MDS and HCA summary

As all concept mappers know, MDS and HCA are the multivariate engines in CM methodology (Trochim & McLinden, 2015). The analytic process and statistical algorithms to accomplish these analyses described by Trochim in the 1989 special issue were consistently employed in the dissertations. As noted, most of the dissertations used the Concept System. A small number of students examined more than two MDS dimensions, but all eventually accepted and reported a two-dimensional model. Great interest in MDS stress values has been shown in the CM literature, and the summary statistics in Trochim's (1993) paper were cited in 37 dissertations. Trochim's synthesis of 33 CM projects reported a mean stress value of 0.28,527 with a standard deviation of 0.04. Similarly, Rosas and Kane (2012) reported a mean stress value of 0.28 with an *SD* of 0.04 in their synthesis of 69 CM studies. The dissertation set of 104 studies included 86 (82.7%) that included a single concept map and 18 (17.3%) with multiple maps, with a combined report of 96 stress values. The 96 stress values were normally distributed with a mean of 0.26 and a standard deviation of 0.05. The range was from 0.13 to 0.36.

In order to explore potential correlates of stress values, the following variables were studied: number of maps in study, number of sorters, sort modality, number of items, and number of clusters. In Table 5, a breakdown of stress values by number of maps and sort modality is presented. The estimates are very consistent around the overall mean of 0.26, with the exception of the two studies with three maps, with a mean of 0.31. However, the range in this very small subset was 0.29–0.33, toward the higher end of other reported distributions (Rosas & Kane, 2012; Trochim, 1993), but within the typically accepted range.

Stress and other key variables were further examined with correlations shown in Table 6. Stress was negatively correlated with the number of sorters ($r = -0.30$), and a modest positive correlation with the number of items ($r = 0.27$) was obtained. The number of items was also correlated with the number of brainstormers ($r = .23$) and the number of clusters ($r = 0.25$). The number of brainstormers and clusters were also associated in a positive direction with a correlation of 0.28. Means and standard deviations are also given in the table. The mean number of statements was 87.76 ($SD = 34.48$), only slightly less than the number reported by Rosas and Kane (2012), which was 96.32 ($SD = 17.23$). The average number of clusters was 7.86 ($SD = 3.0$), also similar to Rosas and Kane's report of 8.93 ($SD = 1.55$) in their analysis of 69 studies.

3.6. Reliability

Nearly all of the dissertations included some attention to reliability. As noted above, Trochim's (1993) analysis of CM reliability was cited by 67 (64.4%) of the studies. Eighteen ($n = 17.3\%$) included a specific reliability analysis. Trochim's paper described six methods of reliability estimation. Seven dissertations

Table 5
MDS Stress Values by n of Maps in Study & Sort Modality.

Variable				
Number of Maps in Study	Value (*n* of studies)	Stress *M* (*SD*)	Min/Max	95% CI
	1 (*n* = 62)	.26 (.05)	.13/.35	.25, 0.27
	2 (*n* = 13)	.25 (.03)	.20/.29	.23, 0.27
	3 (*n* = 2)	.31 (.03)	.29/.33	.04, 0.57
Sort Modality				
	Internet (*n* = 13)	.26 (.04)	.18/.32	.23, 0.28
	In person (*n* = 40)	.26 (.05)	.13/.35	.24, 0.28
	Mail (*n* = 14)	.27 (.05)	.18/.33	.23, 0.29
	Mixed (*n* = 10)	.26 (.05)	.17/.33	.23, 0.29
All Maps	*N* = 96	.26 (.05)	.13/.36	.25, 0.27

Table 6
Summary of Correlations, Means, and Standard Deviations for Primary CM Variables.

Variable	1	2	3	4	5	6
1. Items	–					
2. Clusters	.25*	–				
3. Stress	.27*	0.17	–			
4. Brainstormers	.23*	.28*	0.03	–		
5. Sorters	−0.04	0.11	−0.30*	0.05	–	
6. Raters	−0.04	0.07	−0.04	0.09	.63**	–
M	87.76	7.86	0.26	48.84	27.40	34.91
SD	34.48	3.00	0.05	78.16	23.81	30.84

Note: $^{*}p < 0.05$ $^{**}p < 0.01$.

employed more than one of the six methods. Two studies (Cabrera, 2006; Lindstrom-Johnson, 2009) reported split-half matrix estimates ranging from 0.63 to 0.84. Johnson (2012) examined the individual-total matrix correlations and reported an overall estimate of 0.88. Four authors supplemented one or more of Trochim's procedures with examination of Cronbach's alpha on scales comprised of cluster item ratings (Bedi, 2004; Cacy, 1995; Florio, 1997; Heisler, 2014). Six additional studies reported alpha, but not one of the six Trochim methods. The alpha estimates for the 10 studies ranged from 0.49 to 0.96. Pope (2010) looked at consistency in cluster naming by two subgroups in her study by computing Kappa on agreement in her nine-cluster solution. She obtained an estimate of 0.62. One study also provided a reference to Weller and Romney's (1988) volume that included a detailed review of the study of similarities via sorting techniques in ethnography. Their analysis included a recommendation of a minimum sample size of 30 to obtain reliability of 0.90.

3.7. Validity

As with reliability, validity was frequently discussed and sometimes assessed. Not surprisingly, the greatest emphasis in

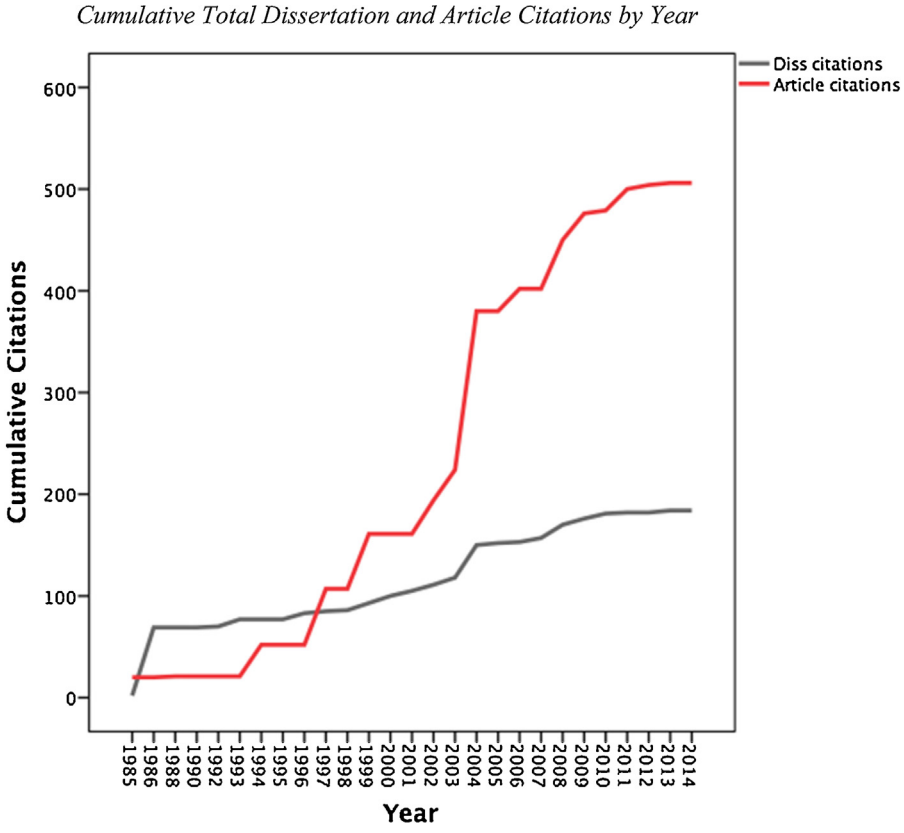

Fig. 3. Cumulative Total Dissertation and Article Citations by Year.

both narrative and data analysis was on construct validity, including statistical tests of structure, formal content validity evaluation with participants, and some attention to translational validity. Pattern matching analyses were included in half of the studies ($n = 52$). Trochim's (1985, 1989c) foundational papers on pattern matching described process and outcome pattern matching models in validity assessment. The dissertations included a small number with explicit attention to these ideas, but many with more exploratory approaches, primarily in examination of patterns apparent in categorical breakdowns by participant characteristics. The analysis is readily accomplished in the Concept System. Four studies included a formal analysis of internal structure using factor analysis or SEM. Convergent and discriminant validity were examined in nine studies that included correlations with previously published standardized measures. Attention to content validity was apparent in nearly all studies in the editing of items and interpretation phases. Three studies included what the authors termed a "testimonial validity" procedure in which participants were given the opportunity and responsibility for cluster naming and interpretation. External validity considerations included commentary of sampling strategies. As noted above, a small number ($n = 9$) used some form of random sampling. Most used a restrictive set of inclusion criteria for their purposive samples of expert stakeholders.

3.8. Citations

As noted, citation data were obtained from the ProQuest database as well as from Google Scholar. This included citations for the dissertation itself and a search for resulting articles with associated citation counts. For the original dissertations, 64 (51.9%) were cited at least one time. Most of these dissertations were cited once (18, 17.3%) or twice (14, 13.5%). The first CM dissertation by Torre (1986) was cited 67 times, with the next closest number being 16, for Bosch's (2004) architectural study of school environments and outcomes. The overall mean number of citations per CM dissertation was 1.9 ($SD = 7.17$).

Of the 104 dissertations, 28 (26.9%) resulted in peer reviewed journal articles. The citation range for these articles was 0–92, with a mean of 17.45 ($SD = 21.17$). Sixteen (15.4%) of these papers have been cited at least 10 times, and 12 studies (11.5%) have been cited by more than 20 other authors. Three have been cited more than 50 times (Florio, Donnelly, & Zevon, 1998; Bedi, 2006; Michalski & Cousins, 2000). Fig. 3 shows the cumulative total of dissertation and article citations from 1985 until 2013. The total number of dissertation citations was 184 and the total number of article citations from these studies was 506.

Given the interest in the impact of the original 1989 Trochim papers, dissertation reference lists were examined to identify the rate of citation of the original and subsequent CM methods papers. A total of 15 articles were identified, with the ten most frequently cited listed in Table 7. The introductory paper (Trochim, 1989a) was cited in exactly three fourths of the studies ($n = 78$, 75%), closely followed by the 1993 review and exposition of methods of reliability assessment ($n = 67$, 64.4%). The more recent book by Kane and Trochim (2007) is already the fourth most cited CM methods guide, with 26 citations. The Kane and Trochim text was cited in 25% of all CM dissertations and over half (56.5%, $n = 46$) of the dissertations completed since its publication in 2007.

Table 7
Top Ten Most Frequently Cited CM Method Publications.

Article	Frequency (%)
1) Trochim (1989a)	78 (75%)
2) Trochim (1993)	67 (64.4%)
3) Trochim (1989b)	36 (34.6%)
4) Kane and Trochim (2007)	26 (25%)
5) Trochim (1985)	18 (17.3%)
6) Trochim and Linton (1986)	17 (16.3%)
7) Trochim, Cook, & Setze (1994)	17 (16.3%)
8) Trochim (1989c)	13 (12.5%)
9) Trochim, Milstein, Wood, Jackson, & Pressler (2004)	8 (7.7%)
10) Trochim and Kane (2005)	6 (5.8%)

4. Discussion

The purpose of this review was to examine the breadth and depth of concept mapping in the set of identified dissertations written since 1985. Among the first conclusions to be drawn is support for the remarkable flexibility of the method. As Kane and Trochim (2007) put it: " . . . the concept mapping analysis is a standardized approach, but it allows for tremendous flexibility and adaptability to address different kinds of questions and problems" (p.108). The variety of topics and research questions observed in the dissertation review ranged from topics near its origins in structured conceptualization in evaluation research to distal topics in medicine, civil engineering, business and many other fields. In addition, the procedures employed generally followed the standard steps closely but evidenced adaptability in extensions to theoretical and applied problems of many kinds.

Some of the innovative methods used by students merit replication and refinement as the CM/PM method evolves. For example, internal validity did not receive as much attention as other forms of validity, but one study (Adu, 2011) developed a video to demonstrate sorting and rating. Further study of methods of reducing noise in sorting and rating data may prove to be valuable in enhancing internal validity. Most dissertations did not systematically collect feedback on the process from participants, and usability was not a focus of the present analysis, but there was evidence of participant confusion, frustration and fatigue in some of the studies, which may have negatively impacted internal validity.

From the quantitative perspective, perhaps the most important finding was that the 96 stress values reported followed the pattern previously seen in the Trochim (1993) and Rosas and Kane (2012) systematic reviews. The distribution of these values was remarkably similar to the prior syntheses, with an overall mean of 0.26 with a standard deviation of 0.05 in a normal distribution. It was also clear that stress values are not dependent on data collection modality. This may be particularly good news for future students and others who need to employ the most cost effective methods of data collection. Modest correlations were obtained between stress and numbers of items and sorters, a finding that may also influence study planning. One area for future *meta*-analytic study might be a detailed analysis of participant characteristics with stress and other concept mapping variables. For example, participant age and educational levels might be associated with variations in procedures and study outcomes. Clinical vs. non-clinical samples might be another variable related to participation, stress values or other outcomes.

Future students and other concept mappers may want to focus on the use of a single rating scale, or strategies for encouraging participation in multiple scale completion, because use of more than one scale was generally associated with poorer response. One limit of the present review was that it did not examine the extent to which floor and ceiling effects were evident in the rating scale data. This may be a valuable target for future study, especially in the area of importance ratings. Importance ratings may naturally suffer from ceiling effects when the participants are stakeholders in the focal area of the study.

The essential and distinct steps in conducting concept mapping research have generally been followed by students

conducting CM dissertations. Only a very small number were excluded from this review because they either did not follow the standard steps or did not report sufficient information to evaluate whether they had completed a standard CM study. Nonetheless, reporting of concept mapping procedures and participation might be enhanced with a standard model as in the CONSORT and PRISMA statements.

Clearly the Concept System has played a major role in the dissertations studied, with a small number of studies completed with alternative software. Overall, the dissertations showed broad implementation of the analytic tools available in the program. However, it is also notable that despite Trochim's (1993) recommendation that reliability be routinely reported, only a very small number of studies examined reliability, perhaps because additional software was required. Increased study of reliability in concept mapping data might become possible with publication of software tutorials and/or addition of such capability to future programs, including the Concept System.

This review was limited in several respects. First, although I have significant amount of concept mapping experience, I was the sole reviewer, coder and analyst. The systematic coding scheme employed did not include every conceivable variable (e.g., bridging values, ceiling properties in rating scales), and another reviewer might prioritize the components of the review in a different manner. On the other hand, the availability of 104 of the 108 identified eligible dissertations may provide a measure of confidence in the pattern of data that were obtained.

Finally, the first concept mapping study was a doctoral dissertation. It is likely that the use of concept mapping in doctoral studies will continue to produce important findings in a wide variety of fields of study. In addition, the dissertation context is likely to remain fertile ground for continuing adaptation and innovation in the evolution of the method.

References

Adu, P. K. (2011). Conceptualizing doctoral advising from professors' and doctoral students' perspectives using concept mapping. *Doctoral dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 3531922).

Andrés, A. (2009). *Measuring academic research: how to undertake a bibliometric study*. Chandos: Oxford.

Augur, C. P. (1998). *Information sources in grey literature*. London: Bowker-Saur.

Bedi, R. P. (2004). Concept-mapping the client's perspective on counseling alliance formation. *Doctoral dissertation*. Retrieved from ProQuest Dissertation Database. (NQ93099).

Bedi, R. P. (2006). Concept mapping the client's perspective on counseling alliance formation. *Journal of Counseling Psychology*, *53*, 26–35.

Bickman, L. (1987). The functions of program theory. In L. Bickman (Ed.), San Francisco: Jossey-Bass.

Bosch, S. J. (2004). Identifying relevant variables for understanding how school facilities affect educational outcomes. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 3126217).

Cabrera, D. (2006). Systems thinking. *Doctoral dissertation*. . . Retrieved from Cornell University E-Commons http://hdl.handle.net/1813/2860.

Cacy, J. R. (1995). The reality of stakeholder groups: A study of the validity and reliability of concept maps. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 9608493).

Campbell, D. T. (1966). Pattern matching as an essential in distal knowing. In K. R. Hammond (Ed.), *The psychology of egon brunswik* (pp. 81–105).New York: Holt, Rinehart & Winston.

Chen, H., & Rossi, P. H. (1984). Evaluating with sense: The theory-driven approach. In R.F. Conner & Associates (Ed.), Beverly Hills, CA: Sage.

Florio, G. A., Donnelly, J. P., & Zevon, M. A. (1998). The structure of stress and coping in oncology nursing. *Journal of Occupational Health Psychology*, *3*, 227–242.

Florio, G. (1997). The structure of work-related stress and coping among oncology nurses on high stress units: A transactional analysis. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 9807292).

Heisler, K. W. (2014). Assessing the fit between child welfare information systems and frontline workers: The development of a task-technology fit instrument. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 98072892).

Heppner, P. P., & Heppner, M. J. (2004). *Writing and publish your thesis, dissertation and research*. Belmont, CA: Brooks/Cole-Thomson Learning.

Hopewell, S., Clarke, M., & Mallett, S. (2005). Grey literature and systematic reviews. In H. R. Rothstein, A. J. Simon, & M. Borenstein (Eds.), *Publication bias in meta-analysis: prevention assessment and adjustments*New York: Wiley.

Jackson, D. N. (2002). The constructs in people's heads. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement*Mahwah, NJ: Erlbaum.

Johnson, M. A. (2012). United States evaluation policy: A theoretical taxonomy. *Doctoral Dissertation*. . . Retrieved from Cornell University e-commons http://hdl.handle.net/1813/29402.

Kane, M., & Trochim, W. M. K. (2007). *Concept mapping for planning and evaluation*. Thousand Oaks, CA: Sage.

Lindstrom-Johnson (2009). Using concept mapping to structure students' views of the school environment's contribution to school violence. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI: 3356936).

Linton, R. (1985). Conceptualizing feminism: A structured method. *Doctoral Dissertation*. Obtained via Inter-Library Loan from Wayne State University library.

Messick, S. J. (1954). The Perception of Attitude Relationships: A Multidimensional Scaling Approach to The Structuring of Social Attitudes. *Doctoral dissertation*. Obtained via Inter-Library Loan from Princeton University.

Michalski, G., & Cousins, B. (2000). Differences in stakeholder perceptions about training evaluation: A concept mapping/pattern matching investigation. *Evaluation and Program Planning*, *23*, 211–230.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(6), e1000097. http://dx.doi.org/10.1371/journal.pmed1000097.

Pope, A. L. (2010). Moral commitment in intimate committed relationships: A conceptualization from cohabiting same-sex and opposite-sex partners. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 3434159).

Rosas, S. R., & Kane, M. (2012). Quality and rigor of the concept mapping methodology: A pooled study analysis. *Evaluation and Program Planning*, *35*, 236–245.

Torre, D. (1986). Empowerment: Structured conceptualization and instrument development. *Doctoral Dissertation*. Retrieved from ProQuest Dissertation Database. (UMI 8607315).

Trochim, W., & Linton, R. (1986). Conceptualization for planning and evaluation. *Evaluation and Program Planning*, *9*, 289–308.

Trochim, W., & Kane, M. (2005). Concept mapping: an introduction to structured conceptualization in health care. *International Journal for Quality in Health Care*, *17*, 187–191. http://dx.doi.org/10.1093/intqhc/mzi038.

Trochim, W., & McLinden, D. (2015). Hindsight is 20/20: The evolution of concept mapping over the past 25 years. *Evaluation and program planning*. (Special Issue: Concept Mapping at 25: Development, Application, and Future Directions).

Trochim, W., Cook, J., & Setze, R. (1994). Using concept mapping to develop a conceptual framework of staff's views of a supported employment program for persons with severe mental illness. *Consulting and Clinical Psychology*, *62*, 766–775.

Trochim, W., Milstein, B., Wood, B., Jackson, S., & Pressler, V. (2004). Setting objectives for community and systems change: An application of concept mapping for planning a statewide health improvement initiative. *Health Promotion Practice*, *5*, 8–19.

Trochim, W. (1985). Pattern matching, validity: And conceptualization in program evaluation. *Evaluation Review*, *9*, 575–604.

Trochim, W. (1989a). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, *12*, 1–16.

Trochim, W. (1989b). Concept mapping: Soft science or hard art? *Evaluation and Program Planning*, *12*, 87–110.

Trochim, W. (1989c). Outcome pattern matching and program theory. *Evaluation and Program Planning*, *12*, 355–366.

Trochim, W. (1993). The reliability of concept mapping. *Paper presented at the annual conference of the American Evaluation Association*.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection*. Newbury Park, CA: Sage.

James P. Donnelly is a licensed psychologist and professor at Canisius College in Buffalo, New York. He is also Director of Measurement and Statistics at the Institute for Autism Research at Canisius. His general research interests are in complex and chronic health conditions. He has been conducting concept mapping studies since he discovered the original 1989 article shortly after it was published.