Coding Schemes as Measurement Instruments? An Attempt to Assess the Psychometric Properties of a Coding Scheme

Andri Ioannou, Cyprus University of Technology, P.O. Box 50239, Limassol 3603, Cyprus,
andri.ioannou@gmail.com
Agni Georgiou, Elena Papanastasiou, University of Nicosia, P.O. Box 24005, Nicosia 1700, Cyprus,
Email: stylianou.a@unic.ac.cy, papanastasiou.e@unic.ac.cy

Abstract: CSCL researchers rely heavily on coding schemes to analyze collaborative discourse. As Rourke and Anderson (2004) explain, examples of qualitative content analysis (QCA) research in which a coding scheme is developed methodically and validated systematically are rare. In this study, we attempted to conduct a validation of a coding scheme using established quantitative methods; this was an effort to foster the replicability and validity of a coding scheme designed to measure online collaborative learning. Exploratory Factor Analysis suggested a different factor structure than the one conceptualized in the coding scheme. The authors raise questions related to what are the most appropriate and acceptable procedures in establishing the validity of coding schemes used in QCA.

Purpose of the Study

CSCL researchers place great emphasis on analyzing discourse (verbal data) -- and the collaborative learning processes captured in it -- in order to understand learning in groups (i.e., Dillenbourg, Baker, Nlaye, & O'Malley, 1996; Hmelo-Silver, 2003; Stahl, 2006). As Dillenbourg et al. (1996) stated, "deciding on the meaning of these expressions in a given dialogue context is quite complex, but necessary if we are to understand when students are really collaborating and co-constructing problem solutions" (p.18).

QCA is widely used as a methodology for the analysis of collaborative discourse, including discourse of asynchronous, computer-mediated discussion groups. The approach generally assumes the researcher will operationalize and unambiguously define the key themes (phenomena) to be coded, create coding categories, and establish their reliability by getting multiple raters to agree on how the categories should be applied to a sample of the data (see Chi 1997 for a detailed description of the steps involved in the creation of a coding scheme for QCA). Despite the large number of CSCL studies using QCA, the methodology has yet to achieve scientific status. Major critiques are concerned with the accuracy, objectivity, reliability, replicablity, and validity of the coding schemes, as discussed in the review by De Wever, Schellens, Valcke, & Van Keer (2005). Rourke and Anderson (2004) also pointed out that examples of QCA research in which a coding scheme is developed methodically and validated systematically are rare.

Our goal in this study was to foster the replicability and validity of a coding scheme designed to measure online collaborative learning, so that future studies could further benefit from using it for QCA. We followed a 2-step approach: First, we adapted an existing instrument developed and applied in previous research. Then, after coding and counting of a large dataset, we employed established methods in quantitative research to examine the psychometric properties of the coding scheme. In other words, we treaded a (qualitative) coding scheme as a (quantitative) measurement instrument. Exploratory Factor Analysis (EFA) followed by Reliability Analysis were used to examine whether the variety of themes (sub-categories) found in the online collaborative discourse could be reduced to invariant structures (factors) that matched the conceptualized categories in the coding scheme.

Method

The Coding Scheme

Researchers, such as Rourke and Anderson (2004) and De Wever et al. (2005) strongly encourage the re-use of coding schemes developed in previous research to foster their replicability and validity. Thus, the coding scheme for the study was adapted from Hmelo-Silver and Chernobilsky (2004). This scheme was developed using Activity Theory (e.g., Engeström & Miettinen, 1999) as the theoretical lens for understanding CSCL. Then, it was used for the study of collaborative learning during problem-based activities in the eSTEP and STELLAR environments (see Hmelo-Silver and associates).

The coding scheme was piloted with discourse data from 19 graduate students, in an Educational Psychology course, who collaborated online for the analysis of a case study with ultimate goal to co-construct a solution to the problem embedded in the case. During the piloting of the coding scheme, two content experts refined its operationalization using 10% of the data. The unit of analysis (segment) was determined to be a consistent "unit of meaning"; each theme was classified into a different coding category, as a separate unit of analysis. Segmentation and coding were carried out together (i.e., in one parse of the discourse). Using the refined coding scheme, the two coders coded the whole discourse independently with an acceptable inter-rater agreement (Ioannou, Bushey, & Artino, 2009). The coding scheme includes four main discourse categories (knowledge sharing, student collaboration, planning, and other content), broken into 17 subcategories, as presented in Table 1. Examples of the codes for each subcategory are omitted due to space limitation, but can be provided upon request.

Table 1: Coding scheme.

Category/ Factor	Sub-category/ Variable	Operationalizion
1. Knowledge Sharing	1.Telling	This contribution corresponds to a discussion of facts, concepts, theories, and experiences with little or no clear connection to the case or earlier ideas contributed to the discussion.
	2.Elaborated Telling	This contribution corresponds to a discussion of facts, concepts, theories, and experiences in relation to earlier ideas contributed to the discussion and with clear connection to the case.
	3.Transforming	This contribution corresponds to a *critical* discussion of facts, concepts, theories, and experiences in relation to earlier ideas contributed to the discussion and with clear connection to the case; this discussion reflects argumentation and reasoning, beyond simple elaboration.
2. Student Collaboration	4.New Ideas	Learner presents an idea or opinion that is new in the context of the discussion. Includes ideas about defining the problem and identifying a solution.
	5.Modification	Learner only partially agrees with a presented idea or opinion, but makes a modification to it.
	6.Elaboration/ Expansion	Learner presents statements, clues, or comments that further refine or expand an earlier idea/opinion; this contribution does not include a critical discussion of concepts and theory to make this complex.
	7.Agreement/ Consensus	Learner presents statements of agreement between and amongst group members.
	8.Disagreement	Learner presents statements of disagreement between and amongst group members.
	9.Syntheses/ Summaries	Learner presents summary previous ideas; this contribution does not include a critical discussion to make this statement complex.
	10.Acknowledgment	Acknowledgment of a contribution.
	11.Explanatory Questions	Asking for clarification or justification of some idea or opinion.
	12.Meta and Prompts	Questions, comments, or suggestions that prompt further discussion and meta-thinking; these statements are idea-oriented, rather than task-oriented.
	13.Justification/ Clarification	Elaboration, justification, clarification as a result of an elaborative question.
3. Planning	14.Monitoring and Planning	Includes self-monitoring and group-monitoring utterances and decisions about how to move forward with the task.
4. Other Content	15.Tool-related talk	Questions, answers, comments, and clarifications related to the technology.
	16.Process-related talk	Questions, answers, comments, and clarifications related to the task or process.
	17.Social Talk	Talk amongst group members that does not add directly to the task.

Data Collection and Procedure

In a subsequent larger scale-study, data were collected from 175 graduate students, working in groups of 3-4 members. The data were collected over a period of two years, in sections of two different courses (Learning Theory and Educational Psychology), in two different universities (a public University in the Northeast USA and a private University in Cyprus), while students collaborated on the same online case study activity.

The coding scheme of Table 1 was used to analyze the online discourse of all groups. Like with the pilot, the unit of analysis was the “unit of meaning" and segmentation and coding were carried out together. Two coders coded approximately 50% of the discourse independently, with an acceptable inter-rater agreement (close to 80%). Disagreements were resolved with discussion between the coders until 100% agreement was achieved. Then, one of the coders finished coding the rest of the discourse. A total of 3385 units were coded (i.e., N=175 graduate students x n° units of analysis = total # of units coded).

Analysis and Results

Following the coding and counting in QCA, the data were organized into an SPSS file -- the 17 variables (sub-categories) were organized through the 175 subjects. An EFA was carried out to examine whether the variety of themes (variables) found in the online collaborative discourse could be reduced to invariant structures (factors) that matched the conceptualized categories in the coding scheme. Ideally, the factor structure of the data would match the main categories of the coding scheme of Table 1. The oblique, Promax rotation with Kaiser Normalization was used to allow the factors to be correlated with each other. In addition, the Maximum Likelihood extraction method was used; this is a true factor analysis extraction method that analyzes only the common variance in the variables and produces optimal results that reflect the nature of the population (Costello & Osborne, 2005). Finally, the eigenvalues cut-off point was set to 1.1 to ease in the interpretation of the factors. Four factors were extracted that accounted for 61.27% of the total variance in the data. The first factor explained 27.19% of the variance and included 6 variables (7, 4, 3, 2, 6, 10). The second factor explained 14.9% of the variance and included 5 variables (15, 16, 14, 17, 13). The third factor explained 10.88% of the variance and included 4 variables (12, 1, 9, 11). The fourth factor explained 8.31% of the variance and included 2 variables (5, 8). The results of the rotated structure matrix appear on Table 2.

Table 2: Rotated structure matrix of EFA.

Variable/ Sub-category	Factor 1: Task-related activities	Factor 2: Regulation of the process	Factor 3: Regulation of content	Factor 4: Conflict management
7.Agreement/ Consensus	**0.83**	~~0.17~~	~~-0.07~~	~~0.56~~
4.New Ideas	**0.70**	~~0.22~~	~~0.32~~	~~0.05~~
3.Transforming	**0.61**	~~0.11~~	~~0.02~~	~~0.22~~
2.Elaborated Telling	**0.59**	~~0.35~~	~~0.18~~	~~0.37~~
6.Elaboration/ Expansion	**0.53**	~~0.27~~	~~0.33~~	~~0.10~~
10.Acknowledgment	**0.39**	~~0.21~~	~~0.03~~	~~0.22~~
15.Tool-related Talk	~~0.23~~	**0.80**	~~0.31~~	~~0.22~~
16.Process-related Talk	~~0.07~~	**0.75**	~~0.22~~	~~0.18~~
14.Monitoring and Planning	~~0.33~~	**0.65**	~~0.25~~	~~0.02~~
17.Social Talk	~~0.33~~	**0.51**	~~0.19~~	~~0.11~~
13.Justification/ Clarification	~~0.18~~	**0.49**	~~0.21~~	~~0.27~~
12.Meta and Prompts	~~0.19~~	~~0.20~~	**0.73**	~~-0.19~~
1.Telling	~~0.25~~	~~0.23~~	**0.73**	~~-0.08~~
9.Syntheses/ Summaries	~~0.01~~	~~0.32~~	**0.65**	~~0.13~~
11.Explanatory Questions	~~0.10~~	~~0.30~~	**0.62**	~~0.22~~
5.Modification	~~0.28~~	~~0.25~~	~~0.14~~	**0.73**
8.Disagreement	~~0.43~~	~~0.13~~	~~-0.11~~	**0.64**

As evident from Tables 1 and 2, the EFA suggested a different factor structure than the one conceptualized in the coding scheme. As a next step, we made an effort to verbally interpret the resulting factors and understand how the variables (sub-categories) may fit together in the collaborative learning process.

Factor 1: The six variables in this factor seem to measure the "*task-related activities*" (i.e., efforts aimed at solving the problem at hand, such as exchange of ideas, agreements, and contributions of supporting information and theoretical perspectives).

Factor 2: Four of the variables in this factor seem to measure the "*regulation of the process*" (i.e., efforts aimed at planning the task); however, the "Justification/Clarification" variable in this factor is not justifiable.

Factor 3: At a first glance, the four variables in this factor do not seem to fit well conceptually. However, assuming the operationalizion of variable 1 (telling), it would be reasonable to think that such statements may have triggered questions and prompted further discussion (variable 12 and 13). That is, if one brought into the discussion a piece of information that was not related to the case, s/he was likely to receive a question. In addition, questions and prompts were likely to encourage a synthesis or summary of ideas (variable 9). A closer examination of student discourse provided support for these arguments. As such, the four variables in Factor 3 can, in fact, make sense conceptually as a measure of "*regulation of content*" (i.e., efforts aimed at understanding all the contributions).

Factor 4: The two variables in this factor seem to measure "*conflict management*" (i.e., presentation of different views and options and modification of ideas).

Table 3: Factor correlation matrix.

	Factor 1: Task-related activities	Factor 2: Regulation of the process	Factor 3: Regulation of content	Factor 4: Conflict management
Factor 1	1.00			
Factor 2	0.28	1.00		
Factor 3	0.17	0.37	1.00	
Factor 4	0.30	0.24	-0.06	1.00

According to the factor correlation matrix presented in Table 3, Factors 2 and 3 were the most highly correlated between each other (r=0.37, with a medium effect size), indicating that these factors tend to occur concurrently. Indeed, solving a problem collaboratively seems to involve *regulation of content* (e.g., asking questions or making summaries of ideas) and *planning* (e.g., talk about the process), both occurring concurrently in the collaboration space. Additionally, Factor 1 was correlated with Factor 4 (r=0.30) and Factor 2 (r=0.28), indicating some sort of similarity among them. This is not surprising assuming all factors measured some aspect of the online collaborative learning process.

Table 4: Cronbach's coefficient alpha reliabilities.

	# Items	alpha
Factor 1: Task-related activities	6	.76
Factor 2: Regulation of the process	5	.76
Factor 3: Regulation of content	4	.70
Factor 4: Conflict management	2	.62

Following the EFA, for each factor we run Cronbach's coefficient alpha reliabilities -- a widely used measure of internal consistency. The resulting Cronbach's alphas for the four factors appear in Table 4. Researchers recommend coefficient alpha levels of at least .70 (see guidelines in Gable & Wolfe, 1993). Thus, alphas for Factors 1-3 were deemed acceptable; alpha for Factor 4 was lower than recommended, but this should be expected considering this factor includes two items only (Gable & Wolfe, 1993).

Discussion

In this study our goal was to foster the replicability and validity of a coding scheme for QCA. We followed a 2-step approach: First, we adapted an existing instrument developed and used in previous research like Rourke and Anderson (2004) and De Wever et al. (2005) recommend. Then, after coding and counting of a large dataset, we employed established methods in quantitative research to examine the psychometric properties of the coding scheme as a measurement instrument.

Unlike we would have expected, the EFA on the coded data suggested a different factor structure than the one conceptualized in the coding scheme. Reliability Analysis showed the internal consistency within the resulting factors was acceptable for the most part. Although we interpreted the resulting factors and understood (at least for the most part) how the variables fitted together, we do not have a good explanation for the inconsistency between the initial structure of the coding scheme and the emerging factors.

We believe this discrepancy is linked to the fact that the theoretical base for the initial development of the coding scheme (Activity Theory, Engeström & Miettinen, 1999) is fundamentally different from the traditional psychology and its psychometric model. In general, CSCL researchers work with epistemological assumptions that are inconsistent with the psychometric model. CSCL studies (and thus the coding schemes they develop for QCA) often emphasize the importance of context in cognition (e.g., Lave and Wenger, 1991) and the socially situated and culturally-mediated nature of CSCL (e.g., Engeström & Miettinen, 1999). Thus, it might be meaningless or even inappropriate to treat a coding scheme as a (quantitative) measurement instrument with psychometric properties such as invariant factor structure, subscale reliability, and instrument validity. On the other hand, however, CSCL studies often report statistical tests on data produced from QCA. In this sense, one would think that validating the coding scheme as a (quantitative) measurement instrument would be meaningful and appropriate before further statistical testing is conducted.

This discussion is by no means complete. Validating coding schemes using well-established quantitative methods, such as EFA and Reliability Analysis, may be a way to achieve rigor in QCA studies and to make coding schemes re-usable and easier to travel across studies. Yet, several questions remain unanswered: Is it appropriate to tread a QCA coding scheme as a (quantitative) measurement instrument to validate it systematically? How the theoretical base for the coding scheme should to be considered if a quantitative validation in undertaken? Is it adequate to establish the validity of a coding scheme first theoretically (during its development) and then empirically as Rourke and Anderson (2004) suggest? What are the best procedures to follow in order to foster validity and reliability of coding schemes in QCA?

References

Chi, M.T.H. (1997). Quantifying qualitative analysis of verbal data: A practical guide, *The Journal of the Learning Sciences*, *6*, 271–315.

Costello, A.B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research and evaluation*, *10* (7), Retrieved November 4, 2010 from http://pareonline.net/genpare.asp?wh=0&abt=10

deWever, B., Schellens T., Valcke, M., & Van Keer H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers and Education, 46*, 6-28.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In H. Spada & P. Reiman (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier.

Engestrom, Y., & Miettinen, R. (1999). Introduction: Activity theory, a well-kept secret. In Y. Engestrom (Ed.), *Perspectives on Activity Theory* (pp. 1-16). Cambridge: Cambridge University Press.

Gable, R. K., & Wolfe, M. B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. Boston: Kluwer Academic Publishers.

Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers and Education, 41*, 397-420.

Hmelo-Silver, C. E., & Chernobilsky, E. (2004). Understanding collaborative activity systems: The relation of tools and discourse in mediating learning. In Y. Kafai, W. Sandoval, N. Enyedy, A. Nixon, & F. Herrera (Eds.), *Proceedings of the International Society of the Learning Sciences* (pp. 254-261). Mahwah NJ: Erlbaum.

Ioannou, A., Bushey, C., & Artino, A. R. (2009, April). *Supporting collaborative learning in online environments: Web 2.0 in an online educational psychology class*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Print.

Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis. *Educational Technology Research and Development, 52*(1), 5-18.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Acknowledgments

We thank the anonymous reviewers of the first draft of this short paper for their thoughtful comments and suggestions.