Journal of Capital Development in Behavioural Sciences Vol. 11 Issue 1 (March, 2023)
Faculty of Arts & Education, Lead City University, Ibadan, Nigeria
***ISSN Online:** 2449-0679 **ISSN Print:** 2354-3981*

Collaborative Learning Measures using Collaboration Analytics: Applications, Limitations, and Caveats Review

Lamidi Babatunde AUDU
s8laaudu@stud-uni.saarland.de; Lam_275@yahoo.com
+ 4917637369833, +2348101472001
Department of Educational Technology
University: Saarland University, Germany

Abstract
There is a widely held belief that social interaction among learners in groups enhances learning. However, for collaborative learning to be effective, it must meet certain requirements, for example, having appropriate goal structures and tasks that demand collective efforts to be solved. Collaboration analytics is a field of research that aims to measure collaborative learning or learning in groups by quantifying qualitative data. Collaboration analytics in research also provides a more reliable tool for handling massive amounts of data. However, for data to be reliable, there is a need for intercoder agreement. The endeavor to quantify qualitative data using collaboration analytics has proven to be useful, but with certain limitations and caveats.

Keywords: Collaboration analytics, Collaborative learning, Intercoder reliability, Content analysis.

Word Count: 112

Introduction
In recent centuries, the internet has activated a somatic occurrence in research on manlike way of acting. As thriving amount of people interrelate on a day-to-day ground in conversation rooms, web forums, email, on-the-spot messaging environs and the like, group technologist, and educator's visage to their way of acting to interpret the nature of disputatious comprehension, intake in Computer Supported Collaborative Learning how it can be modified textual matter of use. The constraints that internet researchers might encounter are how to

translate or transcribe and report online events in relevant turn of phrase, while at the same time working out their contrast in scientifically noticeable way of behaving. "Contentious cognition building is founded on the premise that individual take measures in distinct dialogue interaction and the ratio of the dialogue is correlated with interaction to comprehension gained. Trainees build difference of opinion in activity with their studying ally to gain proficiency of reasoning as well as regarding the gratified immersed cogitation.

(Weinberger, A., & Fischer 2006, P.5-6)". Meanwhile, researchers had asserted that "showing and tracing gestures are a main ingredient for individual divulgence (Krauss, Chen & Gottesman,2000) and in terms of modern technology, and for sorting out data using by contact screen apparatus (Agostinho, Ginns, Tindall-ford, Mavilid & Paas, 2016). The discussion will comprise of theory that apply to learners working in groups with or without gestures in computer supported collaborative learning.

Theoretical Background

The theory that will serve as a framework for this literature review is collaborative learning theory, which is based on Vygotsky's theory of proximal development. Collaborative learning theory states that learners can accomplish tasks in groups that they would not be able to accomplish individually. Another important assumption of collaborative learning theory is that collaborative learning plays a pivotal role in the development of critical thinking skills, since peer-to-peer tutoring fosters higher-level thinking, improves oral communication, and aids students in building their organizational skills.

The areas of both collaborative learning and collaboration analytics have gained importance in the last decades, as solving tasks has become more sophisticated and often requires the knowledge and the skills of interdisciplinary teams. Another important development is the endeavor to quantify qualitative data to make it more reliable, which, as will be seen below, is not always an easy task. The application of collaboration analytics can be useful in collaborative learning skills, but it is not without caveats and limitations.

The research questions dealt with in this literature review are the following:

1. How do individual learners benefit best from collaborating with other learners?

2. How can collaborative learning be measured in a reliable manner?
3. What are the caveats of machine-based learning analytics?
4. How reliable are the methods employed for the quantification of qualitative data?

The method employed for this review is a comprehensive, qualitative content analysis of the Empirical Process Analysis Method 3 seminar literature.

Concept of Collaborative Learning

Collaborative learning can be defined as a situation in which 2 or more individuals learn or attempt to learn something together (Schneider, B., Dowell, N., & Thompson, K. (2021). The idea of collaborative learning rests on the assumption that learners benefit from learning in group settings through knowledge exchange (Weinberger, A., Stegmann, K., & Fischer, F. (2007). Social interaction facilitates learning through the exchange of knowledge. Thus, when learners interact, their knowledge becomes similar in what Weinberger et al. (2007) call *knowledge convergence*. Knowledge is constructed when a group of learners works jointly to solve a complex task. However, there is a relevant distinction to be made between the concepts of *knowledge equivalence*, which refers to a similar level of knowledge in the members of a group of learners, and *shared knowledge*, which refers to the idea that learners possess knowledge on the same concepts as their partners. Moreover, collaborative learning implies both learning while collaborating with other learners as well as learning to collaborate (Schneider et al., 2021). Another important assumption is that collaborative learning aids individual learners in a group due to their different learning resources and unshared prior knowledge. Knowledge convergence is a concept that describes the processes that take place when learners work in groups and can be both conceptualized and measured. Furthermore, knowledge convergence can be considered a learning outcome of collaborative learning. Citing Cohen (1994), Weinberger et al. (2007) state that the knowledge contribution of each learner in a group can be measured by counting the number of times learners participate and classifying their participation as related or unrelated to the task at hand.

Measuring Collaborative Learning Processes

Vogel and Weinberger (2018) point out that the learning gains achieved by learners when helped by other learners were higher than when learners worked individually. However, these researchers state certain requirements for collaborative learning to be effective- namely, the right goal structures, and tasks that require collective work. Vogel and Weinberger (2018) point out that collaborative learning can be analyzed and measured through observing processes and activities conducive to learning. Nonetheless, these researchers are skeptical regarding the use of qualitative approaches, as in their opinion they lace reliability and validity, thus making predictions difficult. These researchers propose a theoretical frame of reference for the analysis and operationalization of learning processes. Their frame of reference is based on Vygotsky's theory of collaborative learning which rests on the belief that learners build upon each other's contributions and thus achieve higher levels of development. Moreover, based on Piaget's principle of resolution of socio-cognitive conflicts, they state that peer tutoring aids learning in settings where students from different disciplines are brought together to solve interdisciplinary problems, or when learners are asked to reflect on others' arguments. Variables that measure learning can be operationalized by using segmentation and coding. Vogel and Weinberger (2018) describe segments as the smallest units of analysis on which the coding scheme is applied. These segments can be words, phrases, or messages.

Collaboration Analytics and Computer-supported Collaborative Learning (CSCL)

Learning analytics is defined as the measurement, collection, analysis, and reporting of data about learners and their contexts. The purpose of learning analytics is understanding and optimizing both the learning process and the environment in which takes place (Siemens, 2013 as cited in Schneider et al., 2021). Collaboration analytics combines the definition of collaborative learning and learning analytics to focus on the collection, measurement, and analysis of data from groups of learners with the aim of providing them support. One of the main areas of collaboration analytics is the development of data collection tools using IT. However, computational models for collaboration analytics entails challenges: One of these is predicting social constructs that are not easily identifiable, and another one is providing coding schemes with high validity and reliability (Schneider et al., 2021). Some researchers are of the opinion that collaboration analysis can have positive effects on collaborative learning both in theory and in practice.

Weinberger and Fischer (2006), for example, state that written learning interactions can be analyzed quantitatively using computer-supported collaboration scripts. These researchers propose a framework for the analysis of 4 dimensions of knowledge construction using CSCL, namely: 1. the participation dimension, 2. the epistemic dimension, 3. the argument dimension, and 4. the dimension of social modes of construction (Weinberger and Fischer, 2006). The participation dimension allows the researcher to find out whether learners participate, and the frequency of their participation, and on epistemic dimension sheds light on the content of learners' contributions. Weinberger and Fischer (2006) differentiate between on-task participation, when learners attempt to solve the given learning task, and off-task participation, when learners do not attempt to solve the task at hand; the argument dimension refers to learners' ability to construct arguments and counter arguments with the aim of providing solutions to complex problems, and the dimension of social modes of construction refers to the degree to which learners refer to the knowledge provided by their peers. Social modes of construction can take the form of externalization of the knowledge provided by their peers, elicitation by asking questions, or conflict-oriented consensus-building, when learners are encouraged to seek different perspectives or to find better arguments when their perspectives face criticism.

Multimodal learning analytics (MMLA) combines data from different sources and, as its name suggests, analyzes data produced using different communication channels. Multimodality in didactics has been subjected to research in different areas in the last 2 decades; however, multimodal tracking, or analytics, is a much more recent field of study (Di Mitri, D., Schneider, J., Specht, M. & Drachsler, H. (2018). With the 4th industrial revolution, the boundaries between the physical, the biological and the digital realms are blurred. One application is the internet of things, which involves connecting sensors to objects of the physical world or bodies of living organisms. The aim is to produce data that can be interpreted by machines. Di Mitri et al. (2018) stress that there is a "general call for multimodality", meaning that there is a need to link digital and physical interactions using multimodal data systems with the aim of producing data about collaborative learning and collective sense-making. These researchers point out that communication between people takes place through different modalities other than the verbal one, e.g., voice cadence, facial expression, and body language.

Thus, it is possible to MMLA to provide a connection between complex learning behavior and learning theories (Di Mitri *et al.,* 2018).

Learning analytics enables educators to provide quality education to underprivileged students by developing methods that can both examine and quantify non-standardized forms of learning (Blikstein and Worsley, 2016). Assessment and feedback, two difficult areas for constructivist learning, could be enhanced by the employment of "fine grained" data collection and analysis by providing educators with novel assessment techniques. MMLA could provide sensing and assessment modalities in 3 different areas: Student knowledge assessment, student affect and physiology assessment, and the assessment of students' intentions and beliefs (Blikstein and Worsley, 2016).

However, the application of MMLA poses great challenges: the first one is that there is no consensus regarding the specific ways in which MMLA could aid students in their learning processes; the second one is that a way of combining human and IT interpretations of multimodal data is still lacking, and the third one is the gap between the meaning of learning assigned by learning sciences on the one hand and machine learning on the other hand. Furthermore, there are significant ethical issues to be considered, as participants in groups in many countries must give their formal consent to being recorded and/or filmed while participating in learning activities. So, even if MMLA could provide a link between observable learning behaviors and learning theories, as Di Mitri et al. (2018) claim, human participants in research experiments must be fully informed about the aims of research and must explicitly give their consent before MMLA can be employed in collaboration analytics.

Content Analysis-the Endeavor of Quantifying Qualitative Data

Content analysis of written messages is of pivotal importance in the field of mass communication research. Analysis of communication transcripts is carried out using both the quantitative and the qualitative approaches. In the quantitative approach, content is coded, summarized, and expressed in terms of frequencies. In contrast, the qualitative approach uses methods like participant observation and ethnography, without computing frequencies. In the same manner, reliability is expressed differently in the 2 approaches: while reliability is expressed as a numeric value in the quantitative approach, in the qualitative approach reliability is established through triangulation with quantitative data sources (Strijbos et al., 2006).

When massive amounts of messages are analyzed by multiple coders, intercoder reliability becomes indispensable for the data to be useful. High levels of disagreement among researchers suggest that their methods are weak, including the possibilities of unclear operational definitions, categories, and poor judge training (Lombard et al., 2002). Furthermore, in cases involving high volumes of data, dividing the coding work among different coders becomes necessary. Intercoder reliability is determined when at least two coders categorize units of research and subsequently use the categorizations to numerically calculate the extent of agreement among the different coders (Lombard et al., 2002). This calculation can be done using different formulae, e.g., the percent agreement, the Holsti method, Cohen's kappa index of Krippendorf's alpha index. Cohen (1986) states that the percentage of agreement between 2 judges who assign cases to a set of κ categories, mutually exclusive and exhaustive necessarily contains a proportion that is attributable to chance.

Relevance of Intercoder Reliability

Coding is a method typically used in qualitative data analysis. Coding is used to divide raw data into distinct categories. The first step in coding is defining what the data are about and classifying them. It is pivotal to start coding as soon as the data come in to provide researchers with a better understanding of the data. Liamputtong (2009) recommends researchers to read through their documents first without coding them, taking notes about what aspects seem particularly interesting. They should read the data a second time, and at this time produce an index of key terms or categories that aid in interpreting and theorizing the data. Subsequently researchers should start with axial coding, or finding linkages between the different categories, such as cause-and-effect relations. While nowadays there are computer-assisted qualitative data analysis programs like for example ATLAS, MAXQA and NVivo, these should be used only as aids to find, categorize and retrieve data, as these programs cannot analyze the data for researchers (Liamputtong, 2009). One of the reasons why AI cannot carry out analysis is that context is crucial for understanding qualitative research findings, and computer programs may fragment information into pieces thus de-contextualizing the data (Liamputtong, 2009).

One of the problems of transcribing oral data, for example from recorded interviews, is that the transcription of oral data carried out by individual researchers cannot be considered objective. As Bucholtz (2000) stresses, transcription is a process embedded in power relations. Transcription of oral data involves both interpretative and representational decisions, namely, *what* is to be transcribed, and *how* it is to be transcribed. Another decision of great importance is whether the transcription will be naturalized, that is, adapted to written language conventions, or denaturalized, that is, the text will retain the oral discourse forms. For Bucholtz (2000), the transcriber must assume a reflexive role when doing a transcription of oral data, as he or she must be aware of the importance of what is transcribed and how it is transcribed and of its ideological implications. Bucholtz (2000) is of the opinion that in every transcription there is a purpose, an audience, and the role of the transcriber toward the text. While due to the exigencies of accuracy in academic transcription there is probably less bias than in non-academic transcriptions, where sociopolitical issues are also present. Bucholtz (2000) provides an example of two transcriptions of the same conversation- one of them forensic and the other one academic. The differences regarding both omissions and utterances are significant, and the attorney of the defendant concluded that the confession had been coerced. This is a situation in which more than one coder would be needed to provide reliability to a transcription. However, it is possible that Bucholtz (2000) was not familiar with NVivo software for dialogue transcription, since the software was created in 1997.

Another study that illustrates how fallible observers are as individuals, is the one published by Landis and Koch (1977). These researchers conducted a study about the lack of convergence between the patient evaluations of multiple sclerosis carried out by a neurologist in Winnipeg and a neurologist in New Orleans. The physicians only agreed on the diagnoses of 43% of the patients, even though both neurologists evaluated the same records. Furthermore, their diagnostic criteria proved not to be very different.

For DuBois (1991) the process of discourse transcription is far from a mechanical task, as it relies on interpretation within a theoretical frame of reference. Otherwise, instead of arriving at functionally significant categories, the researcher is left with "raw acoustic facts" (p.72). Thus, discourse transcription creates a representation in writing making it available to discourse research. Furthermore, for a transcription to be useful, it must present the needed information and present it in a form that is easily understood. For this purpose, DuBois (1999) recommends the use of standardized symbols, so that data can maintain its integrity through different contexts of use.

Chi (1997) also assumes that researchers of qualitative data are biased. This researcher proposes a methodology of verbal analysis, in which the contents of verbal utterances are quantified. As Chi (1997, p.2) puts it, "…one tabulates, counts, and draws relations between the occurrences of different kinds of utterances to reduce the subjectiveness of qualitative coding." The aim of quantifying utterances is to make qualitative data less subjective by quantifying it in some manner.

In studies based on qualitative methods, usually, the coder and the researcher are the same people. However, in studies that require large samples, multiple coders might be needed. While some scholars like Lombard *et al.* (2002) argue that multiple coders are needed to establish reliability, it is dubious that a coder who has not carried out research can understand the context in which it took place. Lombard et al. (2002) state that intercoder agreement is of crucial importance for content analysis. They define intercoder reliability as the extent to which independent coders evaluate units of analysis and reach the same conclusions. This term is interchanged with the terms of intercoder -or interrater- agreement.

Lombard et al. (2002) carried out a study about the assessment of intercoder reliability in published mass media research, and the result was disappointing, as only 69% of the research analyzed reports contained information about intercoder reliability. Lombard et al. (2002) conclude that there are significant problems with assessing and reporting intercoder reliability, which translates into a low validity of mass communication research. Two of the most important issues seem to be the determination of units of analysis and rules for segmentation and coding. As Strijbos *et al.* (2006) put it, intercoder reliability of both segmentation and coding are pivotal to guarantee objectivity, reliability and replication of findings. These researchers recommend that the procedures employed to determine the units of analysis as well as the rules that guide coding should be clearly explained. Furthermore, coding procedures should be published for cross-validation and for secondary analysis. Strijbos et al. (2006) recommend 5 steps to that end. These are: 1. A clear determination of the units of analysis, 2. The development of a segmentation procedure, 3. Testing the reliability of the segmentation procedure, 4. Re-using or re-adapting coding categories, and 5. Testing the reliability of the coding categories.

Limitations

Both collaborative learning and collaboration analytics are fields of study which are dynamic and rapidly changing, especially with the advances in IT and AI. 11 of the articles analyzed in this review were published before 2012- some of them published even before the turn of the century. While some of the findings of the literature here analyzed hold up to this date, others were written before or immediately after software programs like NVivo were developed. Thus, the problems described by Bucholtz (2000) and DuBois (1991) might well have been overcome today, as software does not hide content that it considers "irrelevant". A useful way of proving the efficacy of NVivo and other software packages which aid researchers in the quantification of qualitative data would be to compare the results of qualitative studies both using and not using NVivo or other software and to triangulate them with studies employing different research methods.

Conclusions

The cornerstone of collaborative learning is that learning in groups potentiates learning processes in individuals working in groups. While this might be supported with data analyzing the dynamics of highly motivated, homogeneous groups with knowledge convergence and equivalence, this assumption does not hold for heterogeneous groups with little or no knowledge convergence and equivalence. Teachers who work with heterogeneous groups, like the author, are aware of the great challenge teamwork represents in heterogenous groups. For the author, who taught English for five years at Nigerian secondary schools, it was a great challenge to get students with significantly different levels of knowledge, cultural capital, and motivation to work together. One of the outcomes of teamwork among students who do not share knowledge convergence and equivalence was that the more advanced students got bored and became demotivated, while the slow learners could not follow the discussion thread.

Regarding collaboration analytics, and especially CSCL, it can be stated that these new fields provide researchers with valuable tools to measure collaborative learning. However, as has been stated above, it is not always easy to measure motivation and interest in a quantitative manner. In the field of research, as shown by Bucholtz (2000), Landis and Koch (1977) and DuBois (1991), collaboration analytics are of great advantage, especially when massive amounts of data are to be coded and interpreted. Intercoder reliability is crucial for the studies to be reliable and transferable. However, as Liamputtong (2009) points out, AI cannot perform the same interpretative reasoning human intelligence can, nor can it draw cause and effect relations or associate different concepts. Thus, the applicability of AI is limited to routine repetitive tasks that involve no creativity, associative thinking, or reasoning- it is doubtful that a software program could ever perform a thick description in qualitative research.

Another important caveat is that a higher quantitative participation is not necessarily related to higher learning outcomes, that is, the more verbose students are not necessarily the ones who learn more in a group. Thus, learning outcomes cannot be accurately measured with instances of participation. Furthermore, the perception of learning of a student can differ widely from that of his or her instructor, as very often grades do not reflect learning outcomes. Students who obtain a poor grade may have learned more from their mistakes than students who have obtained satisfactory grades. Furthermore, attitudes that cannot be measured using collaboration analytics, for example, motivation and interest. Learning dynamics vary from group to group, and while there is a multiplier effect in highly motivated, homogeneous groups with comparable levels of knowledge equivalence and shared knowledge, the same cannot be said of heterogenous groups where there are huge differences among the members. Thus, an individual can be highly motivated to learn in one group but not in another. This would make the accuracy of collaboration analytics measurements for individual learners doubtful.

While collaboration analytics is a tool that facilitates learning measurements, its measurements must be taken with a pinch of salt. Moreover, there are other, ethical concerns regarding the use of MMLA. One of these is related to the privacy of the learners, who might not want to be filmed or recorded. In fact, in some EU countries it is forbidden to record or film participants without their explicit consent. Thus, collaboration analytics can be used as one method for measuring collaborative learning, combined with other methods, e.g., learners' perceptions. However, the findings of collaboration analytics would have to be triangulated to prove their reliability and validity.

Recommendation

Presently, there are numerous studies that have examined the extent of the exert influence of working in group with the aid of technology to find out how much learners can learn under various conditions. Contributors to computer supported collaborative learning studies had made significant contributions about collaborative environment under several conditions. As stated herein the limitations and conclusion of the peer-review had shown that there is knowledge gap in a collaborative learning and it is imperative that contributors considered the rapid changes in a collaborative learning, as the theory of connectivism had proven the importance of knowledge sharing through technology. There is also need for educators or scientists to review the available literature and suggest alternative on how collaborative learning could be improved.

References

Agostinho, S., Ginns, P., Tindall-Ford, S., Mavilidi, M., & Paas, F. (2016). Touch the screen: Linking touch-based educational technology with learning- A synthesis of current research in L.Lin. & R Atkinson (Eds). Educational technologies: Challenges, applications and learning outcomes.

Agostinho, S., Tindall-Ford, S. Ginns, P. Howard. S.J., Leahy, W., & Paas, F. (2015). Giving learning a helping hand: finger tracing of temperature graphs on an iPad Educational Psychology Review 27

Blikstein, P. & Worsley, M. (2016) Multimodal Learning Analytics and Education Data Mining: Using Computational Technologies to Measure Complex Learning Tasks, *Journal of Learning Analytics, 3(2),* 220-238.

Bucholtz, M. (2000) The politics of transcription, *Journal of Pragmatics 32,* 1439-1465.

Chi, M.T. (1997) Quantifying Qualitative Analyses of Verbal Data: A Practical Guide, *The Journal of the Learning Sciences, 6(3),* 271-315.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups, *Review of Educational Research*, *64*, 1-35.

Cohen, J. (1986, January 20) A coefficient of agreement for nominal scales, *Current Contents*, p.18.

Di Mitri, D., Schneider, J., Specht, M. & Drachsler, H. (2018) From signals to knowledge: A conceptual model for multimodal learning analytics, *Journal of Computer Assisted Learning, 34,* 338-349.

DuBois, J. (1991, March) Transcription Design Principles for Spoken Discourse Research, *Pragmatics, 1(1),* 71-106.

Krauss, R.M., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process Model. In D. McNeil (Ed). Language and gesture.

Landis, R. & Koch, G. (1977, March) The Measurement of Observer Agreement for Categorical Data, *Biometrics, Vol. 33, Nr.1,* 159-174.

Liamputtong, P. (2009) Qualitative data analysis: conceptual and practical considerations, Health Promotion Journal of Australia 20(2) 133-139.

Lombard, M., Snyder-Duch, J.; & Bracken, C.C. (2002) Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability, *Human Communication Research, 28(4)* 587-604.

Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. Medical Teacher, 36

Schneider, B., Dowell, N., & Thompson, K. (2021) Collaboration Analytics- Current State and Potential Futures*, Journal of Learning Analytics, Vol. 8 (1),* 1-12.

Siemens, G. (2013) Learning analytics: The emergence of a discipline, *American Behavioral Scientist, 57(10),* 1380-1400.

Strijbos, J.W., Martens, R., Prins, F. & Jochems, W. (2006) Content analysis: What are they talking about? *Computers and Education*, Elsevier, 46, pp.29-48.

Vogel, F., & Weinberger, A. (2018) Quantifying Qualities of Collaborative Learning Processes, *International Handbook of Collaborative Learning Processes, 1st Edition,* Routledge.

Weinberger, A. & Fischer, F. (2006) A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46(1), 71-95.*

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction, 17(4),* 416-426.