How should we measure online learning activity?

Abstract

The proliferation of Web-based learning objects makes finding and evaluating resources a considerable hurdle for learners to overcome. While established Learning Analytics methods provide feedback that can aid learner evaluation of learning resources, the adequacy and reliability of these methods is questioned. Because engagement with online learning is different from other Web activity, it is important to establish pedagogically relevant measures that can aid the development of distinct, automated analysis and visualization systems. Content analysis is often used to examine online discussion in educational settings, but these instruments are rarely compared with each other which leads to uncertainty regarding their validity and reliability. In this study, participation in MOOC comment forums was evaluated using four different analytical approaches: the DiAL-e framework, Bloom's Taxonomy, Structure of Observed Learning Outcomes (SOLO) and Community of Inquiry (CoI). Results from this study indicate that different approaches to measuring cognitive activity are closely correlated and are distinct from typical interaction measures. This suggests that computational approaches to pedagogical analysis may provide useful insights into learning processes.

Keywords: CMC, CSCL, content analysis, learning analytics, MOOCs, pedagogical frameworks.

Introduction

The usefulness of Computer Mediated Communication (CMC) in supporting teaching and learning, by increasing exposure to new ideas and building social capital through informal networks has been recognised for some time (Hiltz, 1981; Kovanovic et al., 2014), as has the distinctive nature of discourse in these settings (Cooper and Selfe, 1990; Leshed *et al.*, 2007). The automatic creation of transcripts of interactions, made possible by CMC technology, provides a unique and powerful tool for analysis especially with the large datasets offered by Massive Open Online Courses (MOOCs). However, while the content of CMC continues to be viewed as a ''gold mine of information concerning the psycho-social dynamics at work among students, the learning strategies adopted, and the acquisition of knowledge and skills'' (Henri, 1992:118), the reliability of analysis and evaluation measures is questioned (De Wever et al., 2006).

This study builds on earlier learning analytics work comparing pedagogical coding of comments associated with Web-based learning objects with typical interaction measures adopted by learning analytics research. These typical measures include intentional rating systems (e.g. 'like' buttons) (Ferguson and Sharples, 2014), 'opinion mining' techniques (e.g. sentiment analysis) (Ramesh *et al.*, 2013), and assessments of language complexity (e.g. words per sentence) (Walther, 2007). 'Likes' are a commonly used rating mechanism that are adopted to measure personal attitudes (Kosinski, Stillwell and Graepel, 2013); sentiment analysis has been used in social media research to explore people's mood and attitudes towards politics, business and a number of different variables, including evaluating satisfaction with online courses (Wen, Yang and Rosé, 2014); and the number of words per sentence has been identified as a signifier of language complexity in a number of studies (McLaughlin,

1974; Khawaja *et al*., 2009; Robinson, Navea and Ickes, 2013).

In this study we seek to explore the potential of content analysis (CA) methods that are founded on pedagogical theory, to test correlations - with each other and with typical interaction measures - with the aim of enhancing standard approaches to learning analytics. Specifically, we set out to test the hypotheses that CA methods, while ostensibly measuring different aspects of learning interactions, are closely associated, and that ratings derived from these methods are also correlated with the typical interaction measures discussed earlier. Potential correlations between these measurements have important implications for the development of automated analysis of online learning.

Related work

In recent years computer-supported collaborative learning (CSCL) research has witnessed a change in learning design focus, from instructor-led to learner-centred approaches. Concurrent with this there have been significant developments in content analysis to understand CSCL. Naccarato and Neuendorf (1998) describe CA as the "systematic, objective, quantitative analysis of message characteristics" (p. 20). However the importance of evaluating from a qualitative perspective was acknowledged nearly twenty years earlier in Hiltz's (1981) seminal paper on CMC, where combined qualitative and quantitative approaches are recommended to build a better picture. In addition, Gerbic and Stacey (2005) note that researchers tend to present qualitatively analysed units of meaning within discussions as numeric data in order to apply statistical analysis. This is similar to the method adopted in the present study, where comments have been evaluated in the context of different pedagogical coding models, rated appropriately, and analysed with statistical tools.

Weltzer-Ward's (2011) investigation of 56 content analysis methods used in studies of online asynchronous discussion identified Community of Inquiry (Garrison, Anderson and Archer, 2010), and analyses adopting Bloom's Taxonomy (Bloom et al., 1956) and the Structure of Observed Learning Outcomes (Biggs and Collis, 1982) as widely used methods with high citation counts, accounting for nearly 65% of the papers reviewed. In addition to these instruments, a novel content analysis method developed from the Digital Artefacts for Learning Engagement Framework (Atkinson, 2009) is employed in this study.

Bloom's Taxonomy of the Cognitive Domain

'Bloom's Taxonomy' (Bloom et al., 1956) has become a popular and well-respected aid to curriculum development and means of classifying degrees of learning. As amended by Krathwohl (2002), Bloom consists of a hierarchy that maps learning to six categories of knowledge acquisition (Table 1) each indicating the achievement of understanding that is deeper than the preceding category. In CA studies Yang et al. (2011) aligns Bloom with Henri (1992), a precursor of CoI. In addition, Kember's (1999) association of Bloom's dimensions with Mezirow's (1991) 'thoughtful action' category (e.g. writing), and the utility of mapping word types to Bloom's levels of cognition (Gibson, Kitto and Willis, 2014) are supportive of the use of the Taxonomy in this study.

Structure of Observed Learning Outcome Taxonomy (SOLO)

Similar to Bloom, SOLO (Biggs and Collis, 1982) is a hierarchical classification system that describes levels of complexity in a learner's knowledge acquisition as evidenced in their responses (including writing). SOLO adopts five categories (Table 2) to distinguish levels of comprehension. SOLO-based studies include Gibson, Kitto

and Willis (2014) who map the taxonomy to their proposed learning analytics system, and Karaksha et al. (2014) use Bloom and SOLO to evaluate the impact of e-learning tools in a higher education setting. In addition, Shea et al. (2011) adopt CoI and SOLO to evaluate discussion within online courses, and Campbell (2015) and Ginat and Menashe (2015) apply SOLO to the assessment of writing.

Community of Inquiry (CoI)

The structure of CoI is based on the interaction of cognitive presence, social presence, and teaching presence, through which knowledge acquisition takes place within learning communities (Garrison, Anderson and Archer, 2001). As the current study is concerned with identifying evidence of critical thinking associated with learning objects, the focus is on the categorisation of the cognitive presence dimension which attends to the processes of higher-order thinking within four types of dialogue (Table 3) – starting with a initiating event and concluding with statements that resolve the issues under discussion.

Dringus (2012) suggests that the CoI provides "an array of meaningful and measurable qualities of productive learning and communication…" (p. 96). The model has been adopted in a variety of contexts; Shea et al. (2013) extends CoI with quantitative CA and social network analysis methods, Joksimovic et al. (2014) correlates language use with CoI dimensions, Kovanovic et al. (2014) adopt the framework in their study into learners' social capital, and Kitto et al. (2015) cite CoI in support of their 'Connected Learning Analytics' toolkit.

Digital Artefacts for Learning Engagement framework (DiAL-e)

The DiAL-e framework (Atkinson, 2009), was devised to support the creation of pedagogically effective learning interventions using Web-based digital content. It

adopts ten overlapping learning design categories (Table 4) to describe engagement with learning activities, and is pragmatically grounded "in terms of what the learner does, actively, cognitively, with a digital artefact" (Atkinson, 2009). Case study research indicates that practitioners gain value from using the framework (Burden and Atkinson, 2008), and DiAL-e has been adopted as a pedagogical model in studies that evaluate Web-based learning environments (Kobayashi, 2013; Author, 2015).

Learning Analytics (LA)

The underlying assumptions of LA are based on the understanding that Web-based proxies for behaviour can be used as evidence of knowledge, competence and learning. Through the collection and analysis of interaction data (e.g. learners' search profiles and their website selections) learning analysts explore "how students interact with information, make sense of it in their context and co-construct meaning in shared contexts" (Knight, Buckingham Shum and Littleton, 2014 p. 31). LA methods that focus on discussion forums include learner activity, sentiment analysis, and interaction between learners within forums (Ferguson, 2012) which may be used to predict likely course completion.

Methodology

Comments posted on MOOC forums were manually rated by reference to the selected content analysis methods. Ratings for Bloom, SOLO and CoI were based on the application of values to whole comments, whereas the adapted DiAL-e model used aggregated scores derived from the number of DiAL-e category examples observed in each comment. For example, a comment coded using DiAL-e may contain a question, an indication of research activity (e.g. a hyperlink to a relevant resource), and a statement supporting a previous comment. This would result in an aggregated score of 3 – one for inquiry, one for research and one for collaboration. As with the other three methods, this score is referred to in the rest of this study as the Pedagogic Value (PV). The results of this coding were assessed for intra-rater reliability. Following earlier work suggesting the usefulness of further research into coding based on pedagogical frameworks (Author, 2015), analysis in this study seeks to test the hypothesis that Pedagogical Values for each content analysis method are closely correlated, and explores the implications of this to the development of the automated analysis of online learning. In order to explore possible correlations between inferred learning activity (from scores derived from content analysis), with typical measures of engagement, the level of positive user-feedback ('likes' per comment and comment sentiment) and language complexity (words per sentence) were also evaluated.

Data collection and analysis

An anonymised dataset derived from comment fields associated with 'The Archaeology of Portus' Massive Open Online Course (MOOC) offered on the FutureLearn platform in June 2014 was used in this study. More than 20,000 asynchronous comments, generated by nearly 1,850 contributors (both learners and

educators) occurred within the comment fields of each of the 110 learning 'steps' offered during the 6 weeks of the course.

Qualitative and quantitative content analyses were undertaken manually by the main author using four different methods, on a sample of 600 comments (the MOOC2014 corpus). Qualitative analysis comprised of assessing and coding the MOOC2014 corpus using a content analysis scheme developed from the Community of Inquiry (CoI) cognitive presence dimension (Table 3), as well as thematic analysis schemes developed from Bloom's Taxonomy (Table 1), SOLO Taxonomy (Table 2) and the DiAL-e Framework (Table 4). In total, each entire comment was coded four times using all methods, with a 7-day interval between the application of methods.

De Wever, et al. (2006) argue that reliability is the primary test of objectivity in content studies, where establishing high replicability is important. Intra-rater reliability (IRR) tests were undertaken on 120 comments randomly selected from the MOOC2014 corpus. The results of this coding were assessed for intra-rater reliability. Similar to inter-rater reliability, which measures the degree of agreement between two or more coders, *intra*-rater reliability quantifies the level of agreement achieved with one coder assessing a sample more than once, after a period of time has elapsed. While not as robust as methods employing multiple coders, testing for intra-rater reliability is viewed as an early stage in establishing replicability, provides and indication of coder stability (Rourke et al., 2001, p. 13), and is an appropriate measure for the small-scale, exploratory study reported here.

Many different indicators are used to report IRR (e.g. percent agreement, Krippendorff's alpha, and Cohen's kappa), in this case the intra-class correlations (ICC) method was adopted as it is appropriate for use with the ordinal data under analysis (Hallgren, 2012). ICC was calculated using SPSS software and produced

results of between 0.9 and 0.951 (Table 5), suggesting that that IRR is substantial for this sample (Landis and Koch, 1977).

Procedure and analysis

Quantitative analysis comprised a number of coding activities leading to statistical analysis. Comment data was manually coded and appropriate software was used to automate those parts of the procedure that required consistent and repeatable approaches to data search and numerical calculation.

1: Data consolidation.

In an effort to find typical comment streams, 6 steps were selected based on their closeness to average word and comment count, and where less than 5% of comments were made by the most frequently posting contributor. A further 6 steps were selected: 3 with the highest number of comments and highest word count and 3 with the lowest number of comments and lowest word count.

2: Count and categorise pedagogic activity.

The first 50 comments from each of these 12 steps were then coded, amounting to 600 out of the total 20,253 comments – a 3% sample. Coding for Bloom, SOLO and CoI was based on the application of values to whole comments, whereas the adapted DiAL-e model used aggregated scores derived from the number of DiAL-e category examples observed in each comment. For example, a comment coded using DiAL-e may contain a question, an indication of research activity (e.g. a hyper-link to a relevant resource), and a statement supporting a previous comment. This would result in an aggregated Pedagogic Value (PV) of 3 –- one for inquiry, one for research and one for collaboration.

3: Collect and analyse typical learning measures.

The number of 'likes' per comment were counted, and words per sentence and sentiment data for each comment (calculated using LIWC2007 software) were aligned with each comment and analysed using SPSS predictive analytics software.

4: Correlate PV for each instrument.

The Pedagogical Values for each CA method were analysed using SPSS predictive analytics software.

Results

SPSS software was used to conduct statistical analysis across the comments. Normal distribution frequencies for all content analysis variables were produced, and scatter plots with fitted lines were generated to identify the existence and intensity of simple linear regression.

Hypothesis 1: PVs for each CA method are closely correlated. Positive linear associations were made between all CA methods and each other (Table 6), all of which were highly correlated. The variables with the strongest statistically significant correlation across all three dimensions were the methods based on Bloom and SOLO taxonomies (Figure 1), but correlation with scores based on the Community of Inquiry model was also high (Figure 2).

Hypothesis 2: PVs for each CA method are correlated with typical interaction measures (sentiment, words per sentence, likes). All comparisons produced graphs that indicated approximate linear relationship between the 4 PV dependent variables and 3 explanatory variables (Figures 3 to 5). There was a low positive correlation between all CA methods and words per sentence, a low negative correlation between all CA methods and positive sentiment, and no statistically significant relationship between the methods and 'likes' (Table 7).

Analysis shows that PV is not related to the number of 'likes' awarded to comments by users, which may indicate that within this learning environment, issues other than those strictly related to learning attention received positive feedback. However, there were statistically significant correlations between language complexity and sentiment with PV scores. Language complexity has been associated with critical thinking (Carroll, 2007), and the positive association between longer sentences and higher PV scores suggest that learners' in the context of the 'steps'

analysed in this MOOC, tended to use longer sentences when engaged in in-depth

learning. The negative association of positive sentiment with higher PV scores indicate

that learners may employ a more formal approach to writing comments that indicate

critical thinking, than when writing at a more surface level.

Discussion

Improved discoverability, enhanced personalisation of learning and timely feedback are useful to learners, but only when the outcome is meaningful and adds real value to the acquisition of knowledge. This study has established that attention to learning has taken place within comments associated with learning objects. Simple measurements of this attention have been made (Pedagogical Value) using four different CA methods, which are distinct from measuring simple 'likes' and automated analysis of language complexity.

With regard to social media 'likes', this study accords with Kelly (2012) who argues that these measures suggest a variety of ambiguous meanings, and Ringelhan, Wollersheim and Welpe (2015) who suggest that Facebook 'likes' are not reliable predictors of traditional academic impact measures.

Results indicate clear and statistically significant correlations between the four CA methods. It is perhaps unsurprising that the instruments derived from taxonomies designed to describe cumulative levels of understanding (Bloom and SOLO) should show the closest correlations. However, methods developed to explain the development of reflective discussion (CoI) and measure engagement in pedagogic activities (DiAL-e) are also closely associated; with each other and the other two instruments. If these instruments are measuring different things, why are they closely aligned?

There are three possible explanations for this. Firstly, they may be measuring very similar behaviours related to the depth and intensity with which people write about what they are thinking. If we agree that there is an approximate connection between complexity of writing and depth of understanding, it makes sense that someone who has applied greater attention to their learning, and wishes to share this

with others, will use more elaborate arguments ('Create' in Bloom, or 'Relational' in SOLO), or attempt to sum up theirs and others ideas ('Resolution' in CoI), or demonstrate that they are engaging in a variety of activities (DiAL-e); all of which appear similar to all CA methods, and suggests that comments evidencing these types of focus will tend to be ranked in a similar manner.

The practice of coding comments revealed styles of writing that are typical of this environment but which are not accounted for in all 4 content analysis instruments. Because of the succinct nature of many comments in the sample, these are particularly evident at the lower end of the CoI, SOLO and Bloom scales. While the 'Triggering' and 'Exploration' dimensions of CoI explicitly facilitate coding for questions and some of the social dynamics characteristic of CMC (aspects which are also classifiable using the 'Inquiry' and 'Collaborate' dimensions of DiAL-e), neither SOLO nor Bloom explicitly account for these features.

SOLO and Bloom have their origins in efforts to assess the quality of student work and evaluate the success or otherwise of instructional design in achieving educational goals in formal settings. Within the context of our study this focus tends to favour the evaluation of lengthy and complete texts (e.g. essays and 'extended abstracts'). Whereas CoI, and the adaption of DiAL-e used in this study, are specifically designed to measure and comprehend the characteristics and value of critical discourse in online discussions, which affords their application to the ephemeral, fragmented styles characteristic of this environment (Herring, 2012). These differences in focus inevitably lead to some instruments identifying some activities better than others, for example in the MOOC2014 corpus, examples of all dimensions in all instruments were identified with the exception of 'Figurative' in DiAL-e and 'Extended Abstract' in SOLO. While it should be recognised that

cognition is a complex phenomena that is not fully understood, and the simple metrics used in this study cannot explain the richness of human behaviour in this setting, Shea et al. (2011) suggest combining different CA methods as a method to achieve greater accuracy.

Finally, although intra-rater reliability tests demonstrated high levels of accuracy, Garrison, Anderson and Archer (2001) suggest that coding processes are unavoidably flawed, because of the highly subjective nature of the activity. Since one judge was involved, and coding decisions tend to reflect subjective understandings of what constitutes effective learning, there is a strong likelihood of bias entering the process. In addition, as the selected sample is biased, this may have adversely affected results for some methods. For example, while two of the steps analysed had little more than 50 comments, others contained many hundreds of comments. By coding only the first 50 of each step the opportunity to find comments aimed at resolving discussions was reduced, which may have affected results for the Community of Inquiry method more than others.

Conclusions

The aims of this study were to compare relevant and established content analysis methods with each other, and with typical interaction methods, with the aim of formulating recommendations on the use of these methods in future studies. Results suggest that while the content analysis instruments are designed to evaluate different aspects of online discussions, they are closely aligned with each other in terms evidencing very similar behaviours related to the depth and intensity of cognition. The implications of this finding have importance for instructors and learners.

Nearly 25 years ago Henri (1992) identified content analysis as a vital tool for educators to understand and improve learning interactions within CSCL - an issue as important now as it was then. However, despite progress in codifying content analysis methods, the process of coding by hand cannot successfully manage the increasing volume of data generated by online courses (Chen, Vorvoreanu and Madhavan, 2014) and requires the development of appropriate automatic methods. In this study we have identified strong correlations between all content analysis methods, and between these methods and learners' use of language, which suggests potential for developing real-time, automated feedback systems that can identify areas in need of intervention.

While this small scale study contributes to understanding the limits of the use of 'likes' as indicators of on-topic engagement, and establishes links between learners' language use and their depth of learning, we believe that further content analysis of different datasets from many other courses with contributions made by different participants, covering diverse subjects, and analysed by multiple raters is required to establish widely applicable methods.

Looking further, our future work will endeavor to build on this broader content analysis and the linear regressions presented in this paper, and engage with Machine Learning (ML) techniques to develop real-time, automated feedback systems. The potential of ML's computational processes is in the combination of multiple metrics (e.g. content analysis scores, word counts and sentiment) in order to predict pedagogical values and provide meaningful feedback. Our expectation is that ML-based tools may be employed to support self-directed learning, enable educators to identify those learners who are experiencing difficulties as well as those who are doing well, and help learning technologists design more effective CSCL environments. In order for educators to successfully manage the high volume of diverse learning interactions inherent in massive courses, the development of this type of software is becoming increasingly important.

References

Atkinson, S. (2009) *What is the DiAL-e Framework?* Available at: http://dial-e.net/what-is-the-dial-e/ (Accessed: 3 June 2015).

Biggs, J. B. and Collis, K. F. (1982) *Evaluating the Quality of Learning: Structure of the Observed Learning Outcome Taxonomy*. New York, NY: Academic Press.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. and Krathwohl, D. R. (1956) *Taxonomy of Educational Objectives. The Classification of Educational Goals. Handbook 1*. Edited by B. S. Bloom. New York, NY: McKay.

Burden, K. and Atkinson, S. (2008) 'Beyond Content: Developing Transferable Learning Designs with Digital Video Archives', in *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 4041–4050.

Campbell, R. J. (2015) 'Constructing Learning Environments Using the SOLO Taxonomy', in *Scholarship of Learning and Teaching (SoTL) Commons Conference*. Savannah, GA: Georgia Southern University.

Carroll, D. W. (2007) 'Patterns of student writing in a critical thinking course: A quantitative analysis', *Assessing Writing*, 12(3), pp. 213–227.

Chen, X., Vorvoreanu, M. and Madhavan, K. (2014) 'Mining Social Media Data for Understanding Students' Learning Experiences', *IEEE Transactions on Learning Technologies*, 7(3), pp. 246–259.

Cooper, M. M. and Selfe, C. L. (1990) 'Computer Authority, Persuasive Learning : Internally', *College English*, 52(8), pp. 847–869.

Dringus, L. (2012) 'Learning Analytics Considered Harmful.', *Journal of Asynchronous Learning Networks*, 16(3), pp. 87–100.

Ferguson, R. (2012) 'Learning analytics: drivers, developments and challenges', *International Journal of Technology Enhanced Learning*, 4(5/6), pp. 304–317.

Ferguson, R. and Sharples, M. (2014) 'Innovative Pedagogy at Massive Scale: Teaching and Learning in MOOCs', in de Freitas, S. I., Ley, T., and Muñoz-Merino, P. J. (eds) *Open Learning and Teaching in Educational Communities*. Springer International Publishing, pp. 98–111.

Garrison, D. R., Anderson, T. and Archer, W. (2001) 'Critical thinking, cognitive presence, and computer conferencing in distance education', *American Journal of Distance Education*, 15(1), pp. 7–23.

Garrison, D. R., Anderson, T. and Archer, W. (2010) 'The first decade of the community of inquiry framework: A retrospective', *Internet and Higher Education*. Elsevier Inc., 13(1-2), pp. 5–9.

Gerbic, P. and Stacey, E. (2005) 'A purposive approach to content analysis: Designing analytical frameworks', *Internet and Higher Education*, 8, pp. 45–59.

Gibson, A., Kitto, K. and Willis, J. (2014) 'A Cognitive Processing Framework for Learning Analytics', in *4th International Conference on Learning Analytics and Knowledge (LAK 14)*. Indianapolis, Indiana. 24-28 March 2014, pp. 1–5.

Ginat, D. and Menashe, E. (2015) 'SOLO Taxonomy for Assessing Novices' Algorithmic Design', in *SIGCSE '15 Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. New York, NY, pp. 452–457.

Hallgren, K. A. (2012) 'Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial', *Tutorials in Quantitative Methods for Psychology*, 8(1), pp. 23–34.

Henri, F. (1992) 'Computer conferencing and content analysis', in Kaye, A. R. (ed.) Collaborative learning through computer conferencing. *The Najadan Papers*. Berlin Heidelberg: Springer, pp. 117–136.

Herring, S. C. (2012) 'Grammar and electronic communication', in Chapelle, C. (ed.) *The Encyclopedia of Applied Linguistics*. Hoboken, NJ: Wiley-Blackwell, pp. 1–9.

Hiltz, S. R. (1981) The Impact of a Computerized Conferencing System on Scientific Research Communities. *Final Report to the National Science Foundation*. Research Report No. 15. Newark, NJ.

Joksimovic, S., Gasevic, D., Kovanovic, V., Adesope, O. and Hatala, M. (2014) 'Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions', *The Internet and Higher Education*. Elsevier Inc., 22, pp. 1–10.

Kelly, B. (2012) What's next, as Facebook use in UK universities continues to grow? *Impact of Social Sciences, LSE Blog*. Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2012/06/19/facebook-universities-continue-grow/ (Accessed: 19 September 2015).

Kember, D. (1999) 'Determining the level of reflective thinking from students' written journals using a coding scheme based on the work of Mezirow', *International Journal of Lifelong Education*, 18 (January 2015), pp. 18–30.

Khawaja, M. A., Chen, F., Owen, C. and Hickey, G. (2009) 'Cognitive load measurement from user's linguistic speech features for adaptive interaction design', *Lecture Notes in Computer Science*, 5726 (PART 1), pp. 485–489.

Kitto, K., Cross, S., Waters, Z. and Lupton, M. (2015) 'Learning Analytics beyond the LMS: The Connected Learning Analytics Toolkit', in *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge - LAK '15*. Poughkeepsie, New York: ACM, pp. 11–15.

Knight, S., Buckingham Shum, S. and Littleton, K. (2014) 'Epistemology, assessment, pedagogy: where learning meets analytics in the middle space', *Journal of Learning Analytics*, 1(2), pp. 23–47.

Kobayashi, M. (2013) 'Using Web 2.0 in Online Learning: What Students Said About VoiceThread', in *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 234–235.

Kosinski, M., Stillwell, D. and Graepel, T. (2013) 'Private traits and attributes are predictable from digital records of human behavior', *Proceedings of the National Academy of Sciences*, 110(15), pp. 5802–5805.

Kovanovic, V., Joksimovic, S., Gašević, D. and Hatala, M. (2014) 'What is the Source of Social Capital? The Association Between Social Network Position and Social Presence in Communities of Inquiry', in *Proceeding of Workshop on Graph-based Educational Data Mining at Educational Data Mining Conference (G-EDM 2014)*. London, pp. 1 – 8.

Krathwohl, D. R. (2002) 'A Revision of Bloom's Taxonomy: An Overview', *Theory Into Practice*, 41(4), pp. 212–218.

Landis, J. R. and Koch, G. G. (1977) 'The measurement of observer agreement for categorical data.', *Biometrics*, 33(1), pp. 159–174.

Leshed, G., Hancock, J. T., Cosley, D., McLeod, P. L. and Gay, G. (2007) 'Feedback for guiding reflection on teamwork practices', in *Proceedings of the 2007 International ACM Conference on Supporting Group Work - GROUP '07*. Sanibel Island, Florida, USA: ACM, pp. 217–220.

McLaughlin, G. H. (1974) 'Temptations of the flesch', *Instructional Science*, 2, pp. 367–383.

Mezirow, J. (1991) 'A critical theory of adult learning and education', in Boud, D. and Walker, D. (eds) *Experience and Learning: Reflection at Work*. Geelong, Victoria: Deakin University, pp. 61–82.

Naccarato, J. L. and Neuendorf, K. a. (1998) 'Content analysis as a predictive methodology: Recall, readership, and evaluations of business-to-business print advertising', *Journal of Advertising Research*, 38(3), pp. 19–33.

O'Riordan, T., Millard, D. E. and Schulz, J. (2015) 'Can you tell if they're learning?: Using a pedagogical framework to measure pedagogical activity', in *ICALT 2015: 15th IEEE International Conference on Advanced Learning Technologies*. Hualien, TW: IEEE, pp. 1–3.

Ramesh, A., Goldwasser, D., Huang, B., Daum, H. and Getoor, L. (2013) 'Modeling Learner Engagement in MOOCs using Probabilistic Soft Logic', *NIPS Workshop on Data Driven Education*, pp. 1–7.

Ringelhan, S., Wollersheim, J. and Welpe, I. M. (2015) 'I Like, I Cite? Do Facebook Likes Predict the Impact of Scientific Work?', *Plos One*, 10(8), pp. 1–21.

Robinson, R. L., Navea, R. and Ickes, W. (2013) 'Predicting Final Course Performance From Students' Written Self-Introductions: A LIWC Analysis', *Journal of Language and Social Psychology*, 32(4), pp. 469–479.

Rourke, L., Anderson, T., Garrison, D. R. and Archer, W. (2001) 'Methodological Issues in the Content Analysis of Computer Conference Transcripts', *International Journal of Artificial Intelligence in Education*, 12, pp. 8–22.

Shea, P., Gozza-Cohen, M., Uzuner, S., Mehta, R., Valtcheva, A. V., Hayes, S. and Vickers, J. (2011) 'The Community of Inquiry framework meets the SOLO taxonomy: a process-product model of online learning', *Educational Media International*, 48(2), pp. 101–113.

Shea, P., Hayes, S., Uzuner Smith, S., Vickers, J., Bidjerano, T., Gozza-Cohen, M., Jian, S. B., Pickett, A. M., Wilde, J. and Tseng, C. H. (2013) 'Online learner self-regulation: Learning presence viewed through quantitative content- and social network analysis', *International Review of Research in Open and Distance Learning*, 14, pp. 427–461.

Walther, J. B. (2007) 'Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition', *Computers in Human Behavior*, 23(5), pp. 2538–2557.

Weltzer-Ward, L. (2011) 'Content analysis coding schemes for online asynchronous discussion', *Campus-Wide Information Systems*, 28, pp. 56–74.

Wen, M., Yang, D. and Rosé, C. (2014) 'Sentiment Analysis in MOOC Discussion Forums: What does it tell us?', *Proceedings of Educational Data Mining EDM2014*, pp. 1–8.

De Wever, B., Schellens, T., Valcke, M. and Van Keer, H. (2006) 'Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review', *Computers and Education*, 46, pp. 6–28.

Yang, D., Richardson, J. C., French, B. F. and Lehman, J. D. (2011) 'The development of a content analysis model for assessing students' cognitive learning in asynchronous online discussions', *Educational Technology Research and Development*, 59, pp. 43–70.

Tables and Figures

Bloom score	Descriptor
0 - Off-topic	There is written content, but not relevant to the subject under discussion.
1 - Remember	Recall of specific learned content, including facts, methods, and theories.
2 - Understand	Perception of meaning and being able to make use of knowledge, without understanding full implications.
3 - Apply	Tangible application of learned material in new settings.
4 - Analyse	Deconstruct learned content into its constituent elements in order to clarify concepts and relationships between ideas.
5 - Evaluate	Assess the significance of material and value in specific settings.
6 - Create	Judge the usefulness of different parts of content, and producing a new arrangement.

Table 1: Bloom's Taxonomy (Bloom et al., 1956; Chan et al., 2002; Krathwohl, 2002).

SOLO score	Descriptor
0 - Off-topic	There is written content, but not relevant to the subject under discussion.
1 - Prestructural	No evidence any kind of understanding but irrelevant information is used, the topic is misunderstood, or arguments are unorganised.
2 - Unistructural	A single aspect is explored and obvious inferences drawn. Evidence of recall of terms, methods and names.
3 - Multistructural	Several facets are explored, but are not connected. Evidence of descriptions, classifications, use of methods and structured arguments.
4 - Relational	Evidence of understanding of relationships between several aspects and how they may combine to create a fuller understanding. Evidence of comparisons, analysis, explanations of cause and effect, evaluations and theoretical considerations.
5 – Extended abstract	Arguments are structured from different standpoints and ideas transferred in novel ways. Evidence of generalisation, hypothesis formation, theorising and critiquing.

Table 2: SOLO Taxonomy (Karaksha et al., 2014).

CoI score	Descriptor
0 - Off-topic	There is written content, but not relevant to the subject under discussion.
1 – Triggering event	A contribution that exhibits a sense of puzzlement deriving from an issue, dilemma or problem. Includes contributions that present background information, ask questions or move the discussion in a new direction.
2 - Exploration	A comment that is seeking a fuller explanation of relevant information. This can include brainstorming, questioning and exchanging information. Contributions are unstructured and may include: unsubstantiated contradictions of previous contributions, different unsupported ideas or themes, and personal stories.
3 - Integration	Previously developed ideas are connected. Contributions include: references to previous messages followed by substantiated agreements or disagreements; developing and justifying established themes; cautious hypotheses providing tentative solutions to an issue.
4 - Resolution	New ideas are applied, tested and defended with real world examples. Involves methodically testing hypotheses, critiquing content in a systematic manner, and expressing supported intuition and insight.

Table 3: Community of Inquiry, cognitive presence (Garrison, Anderson and Archer, 2001).

DiAL-e category	Descriptor
Narrative	A contribution that includes a story or narrative based on relevant themes or required task.
Author	A concrete example of applied learning.
Empathise	A contribution that evidences understanding of other perspectives.
Collaborate	A contribution that encourages on-topic interaction and collaboration.
Conceptualise	A comment that evidences reflection, explorations of "what if" scenarios, theorising, and making comparisons.
Inquiry	Contributions that attempt to solve a real world issue, including questions and comments aimed at developing enquiry.
Research	Indications of attempts at research as well as presentation of evidence.
Representation	Reflections on the presentation of course information and supporting media.
Figurative	Using content as an allegory or metaphor for other purposes (e.g. parody).
Off-topic	There is written content, but not relevant to the subject under discussion.

Table 4: Adapted DiAL-e Framework (Atkinson, 2009).

Instrument	Interclass Correlation Coefficient
Community of Inquiry	0.900
SOLO	0.928
Bloom	0.951
DiAL-e	0.918

Table 5: Intra-rater reliability values.

Correlations	CoI	SOLO	Bloom	DiAL-e
CoI		R = 0.811 ***	R = 0.83 ***	R = 0.673 ***
SOLO	R = 0.811 ***		R = 0.868 ***	R = 0.693 ***
Bloom	R = 0.83 ***	R = 0.868 ***		R = 0.711 ***
DiAL-e	R = 0.673 ***	R = 0.693 ***	R = 0.711 ***	

Table 6: Content analysis methods correlations (***: p<0.001).

Correlations	WPS	Positive words	Likes
CoI	R = 0.38 ***	R = -0.32 ***	R = 0.0 P = 0.996
SOLO	R = 0.325 ***	R = -0.288 ***	R = 0.02 P = 0.623
Bloom	R = 0.358 ***	R = -0.289 ***	R = 0.028 P = 0.494
DiAL-e	R = 0.217 ***	R = -0.182 ***	R = 0.024 P = 0.559

Table 7: Content analysis methods correlations with typical measures (***: p<0.001).
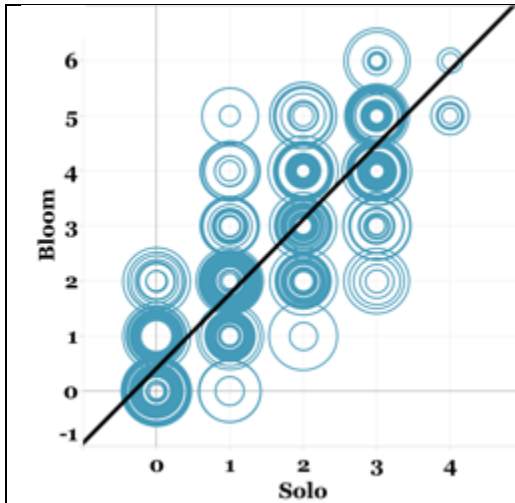
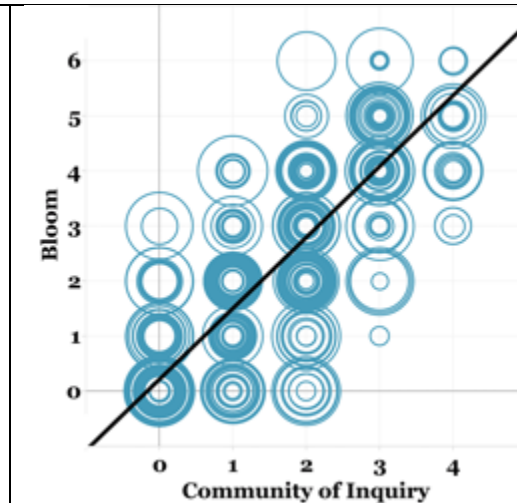Figure 1: Correlation between Bloom and SOLO

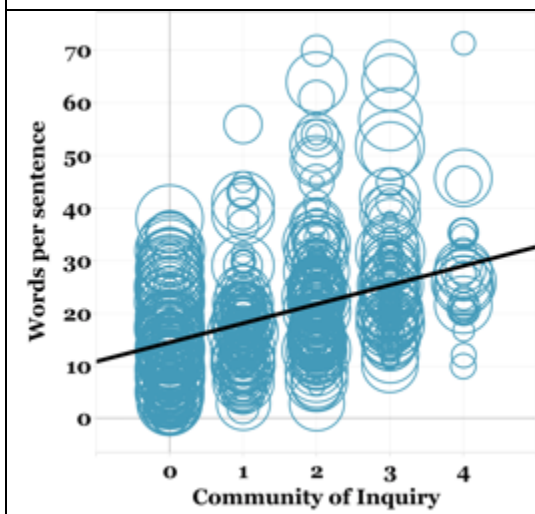Figure 2: Correlation between Bloom and CoI
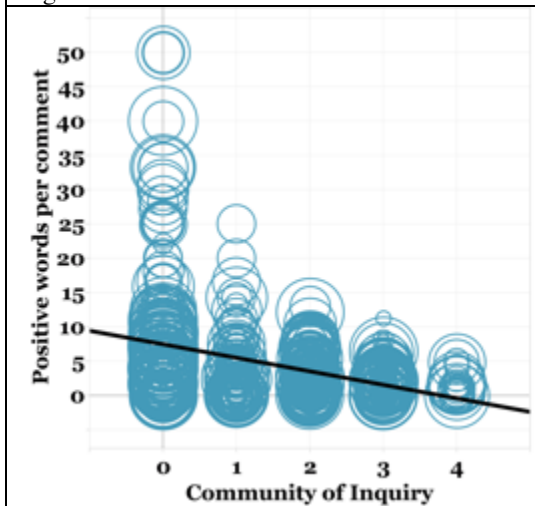

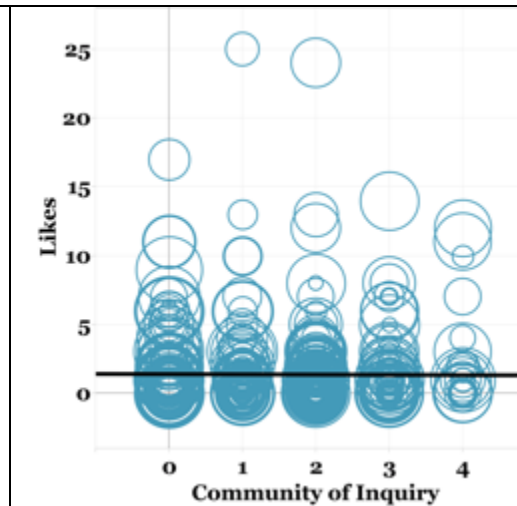Figure 3: Correlation between CoI and WPS

Figure 4: Correlation between CoI and positive words

Figure 5: Correlation between CoI and Likes

How should we measure and show online learning activity?   28