Scalable Use of Big Data Analytics in Small Group Learning

David Topps[1], Michelle Cullen[2]

Corresponding author: Dr David Topps 10-
Institution: 1. OHMES, University of Calgary, 2. University of Calgary
Categories: Learning Outcomes/Competency, Teachers/Trainers (including Faculty Development), Behavioural and Social Sciences, Research in Health Professions Education, Simulation and Virtual Reality

Abstract

The principles of big data analytics consist of more than Volume. Based on our approaches to the use of hybrid natural language processing and the underlying learning analytics afforded by these techniques, we examine how big data analytic approaches can be applied to small groups of learners. We describe a number of techniques, including the use of Sentiment Analysis and how this can be used to more richly inform the learning cycle.

Keywords: learning analytics; precision education; virtual scenarios; therapeutic language; sentiment analysis; cognitive computing

Background

The use of big data in Precision Medicine dominates the mindset of medical academic leadership at present. (Collins, 2015; Hawgood et al., 2015) The principles have also been extended to Precision Education. (Hart, 2016; Eagle and Dubyk, 2017; Topps, Ellaway and Greer, 2019) There are some important principles to bear in mind about big data analytics (Kobielus, 2013): Volume, Velocity, Variety and Veracity, but it is important not to focus too much on volume. Size does matter but more important is what you do with it.

Velocity (the ability to provide rapid, or near-real-time, feedback to learners), Variety (the collection of data of many different types from different sources in various formats), and Veracity (the uncertainty or objectivity of the data, compared to subjective and biased teacher assessments) are all crucial aspects of an educational system. (Kobielus, 2013)

In many areas of health professional education (HPE), we appreciate the importance of strong communication skills. Most medical schools in Canada spend lots of time and money on OSCEs and other approaches to teaching this. Other HPE institutions struggle to find the resources to support OSCEs, both in money and manpower. And these approaches are not at all scalable. With the rising interest in MOOCs and other approaches to scalable learning, the
challenges of individual feedback escalate hugely.

OSCEs in particular are resistant to such scalability, and also require the collocation of many teachers and learners. Because of these challenges, and how they affect the nursing program at our University, particularly in the area of therapeutic language skills, we have been exploring alternative approaches to teaching and assessing these skills.

**Methods**

Throughout the development of these various areas, we have adopted a Design-Based Research approach. (Amiel and Reeves, 2008; Barab, Squire and Barab, 2016) This has allowed us to refine our tools, platforms and analytics as we explored the complex interaction between teachers, learners and program leaders.

Our initial focus was on exploring the clinical reasoning and problem solving in our learners. For this, we adopted OpenLabyrinth, a virtual scenario platform that has been widely used in this area. (Ellaway, 2010; Topps and Ellaway, 2015) However, we also knew that OpenLabyrinth, like most of its ilk, is not strong when used to explore natural language processing (NLP) and communications. There have been other notable attempts to include NLP into virtual patients but these have been costly and complex. For example, the Maryland Virtual Patient project created a very powerful NLP module but at a cost of $11M over 6 years (for a single case). (Nirenburg *et al.*, 2009)

Although big data and cognitive computing platforms are making some promising progress in the area of NLP, the challenge remains that it is tedious and time-consuming to assemble the underlying decision tree, concept sets and language map for any particular case. (Abdul-Kader and Woods, 2015) Most clinical teachers who are tasked with creating the educational scenarios needed are faced with very limited budgets and time to do so.

Our project adopted the concept of the Mechanical Turk, where a human is used to conduct some of the cognitive tasks in a job stream. (Levitt, 2000; Quale, 2016)

*Figure 1: chat boxes in Turk Talk session*
Based on a mechanism that is similar to that seen in the chat-based customer-support software platforms, we adapted OpenLabyrinth to provide a chat-based interface (see Figure 1) between a teacher (or Turker) who could interface with (see Figure 2) up to 8 concurrent learners. (Cullen, Sharma and Topps, 2018)

*Figure 2: Turker view of chat channels*
As we developed this method of text oriented communications, we also found that in certain circumstances, it was easy to cognitively overload our Turkers with too many concurrent conversations. This was mitigated in a number of ways with text shortcuts or macros for commonly used phrases; voice to text conversion; and careful learning design in the associated scenario maps so that the arrival of designated learners was not absolutely concurrent.

At the same time as this Turk Talk project was in progress, OpenLabyrinth also benefited from other research projects that were exploring the capabilities afforded by incorporating activity metrics and xAPI-based reporting to a Learning Records Store (LRS). (Advanced Distributed Learning (ADL), 2014; Topps et al., 2016) xAPI activity statements can be summarized as consisting of simple triplets in the form of Actor-Verb-Object (or “Bob did this”). As with other datastores that have a triplet format, they can be additionally coded with Resource Descriptor Format (RDF) coding, which then in turn lends them to further semantic analysis, a technique typical of machine learning.

This even included the ability to extract xAPI statements from simple, cheap, Arduino-based physiologic sensors. (Meiselman, Topps and Albersworth, 2016) This $30 sensor was able to detect stress-related changes in participants, using heart rate and galvanic skin response sensors, which were remarkably sensitive. All of this data was also captured into the Learning Records Store, providing both a simple means to aggregate data across multiple platforms, in concordance with the Variety (Kobielus, 2013) principle of big data.

Exploring activity streams, as an alternative approach to survey-based data, is very much in keeping with the approach now taken by Google, Amazon et al, in the detailed analyses of their consumer base. Rather than ask the customer their opinion of a product, watch instead as to how they behave and their purchasing patterns. While this data-mining has been often eschewed in academic circles, it now predominates in the commercial world with a subtlety and depth that some would say has left traditional, hypothesis-driven research behind. (Baepler and
As we extended the abilities of the OpenLabyrinth research platform, and deployed Turk Talk in a variety of educational scenarios, some concerns were raised as to whether this approach was too dependent on the facilitating skills of the individual Turkers. We did indeed encounter a variety of facilitation styles and sought a scalable method to evaluate these.

We were able to analyze our learner and teacher performance using a number of cross-related data streams, from these various methods, again in concordance with the Variety principle of big data, but none of these looked at the actual conversations conducted via Turk Talk. Because of the volume of conversations, and limited resources, we discounted the use of more traditional methods such as Discourse Analysis and other qualitative assays of conversation.

We turned instead to Sentiment Analysis, (‘Sentiment analysis: A combined approach’, 2009; Liu, 2012) which is a technique that arose from evaluating the short text fragments associated with the comments on social media discussion boards around products and materials. When we first started exploring this area (see Figure 3), there were limited resources:

Figure 3: example of early JSON output statements provided by the Sentiment Analyzer

```json
{"utterances_tone":[{"utterance_id":0,"utterance_text":"Hello, can you help me","tones":[{"score":0.932314,"tone_id":"excited","tone_name":"excited"},{"score":0.874925,"tone_id":"polite","tone_name":"polite"}]},{"utterance_id":1,"utterance_text":"How are you?","tones":[]}],"utterance_id":2,"utterance_text":"Nothing is working :("},{"tones":[{"score":0.721634,"tone_id":"frustrated","tone_name":"frustrated"},{"score":0.963385,"tone_id":"sad","tone_name":"sad"}],"utterance_id":3,"utterance_text":"Sorry to hear this","tones":[{"score":0.992159,"tone_id":"polite","tone_name":"polite"},{"score":0.772394,"tone_id":"sympathetic","tone_name":"sympathetic"}]}
```

We used text fragments from the conversational streams generated during our training exercises, where we coached our early batches of Turkers in how to cope with the software, the multiple concurrent chat channels and the use of the shortcut macros.

**Results**

At a Kirkpatrick (Kirkpatrick and Kirkpatrick, 1998) level 1 analysis, student engagement and satisfaction was extremely high. As well as the usual ratings, reported elsewhere, there were two particularly strong indicators of how much they liked this approach compared to what was available before: (1) the helpful number of students who volunteered to come back into the program as Turkers, once they had graduated. They stated that they did this, partly because they enjoyed the process, but also because they found that it continued to consolidate their counselling skills after graduation; (2) when a logistical hiccup forced the cancellation of the Turk Talk cases for
one cohort, after going back to the previous questionnaire-based approach used before Turk Talk, the cohort revolted and insisted that the Turk Talk approach be reinstated.

The Turk Talk approach was not intended to completely replace the OSCE approach to learning communications skills. It was intended to improve the base skill levels of the learners, prior to attempting the very limited number of OSCEs inherent in their degree syllabus. The effectiveness of the approach was confirmed in the marked improvement in performance of the cohorts when they did challenge the OSCE exams, which will be reported shortly.

A cost analysis was performed, comparing the Turk Talk approach with the traditional OSCE exam. The Turk Talk method was much cheaper (Cullen and Topps, 2019) but also benefits from being scalable. We have used it in both small group sessions (6-12 learners) and larger cohorts of up to 130 concurrent learners. The OpenLabyrinth software (Topps, Ellaway and Rud, 2019) is readily adaptable to having tiers of Turkers, with front-line facilitators being assisted by senior Turkers. All of this is supported within the standard web-based interface of OpenLabyrinth.

OpenLabyrinth itself has also proven to be remarkably scalable, having been used in MOOCs of up to 30,000 participants, without the need for specialised hardware or high performance computing infrastructure. (Stathakarou, Zary and Kononowicz, 2014)

The simple web interface and low bandwidth needs also solved a practical problem found in traditional OSCEs. Turkers could work from anywhere. Since the school of nursing had a number of part-time teachers, it was often difficult for them to travel to dedicated sessions. The ability to work from home or another clinical site made it much easier to find and sustain a cadre of Turkers.

We made great use of a train-the-trainee effect. As noted above, many students who went through the program came back as teachers and Turkers. This neatly solved the sustainability challenge of sufficient Turkers, especially as the program expanded into an increasing number of small and large groups.

However, it was also noted that there was marked variation in the styles and skill levels of facilitation seen amongst our cadre of Turkers. We initially spent much effort on balancing out the learner experience so that all learners would encounter Turkers who were felt to have different facilitation styles. After the first few iterations, we were able to relax this approach as we found that learners strongly benefited from the Turk Talk method, independent of who they had as a Turker. We also noted that the variety of communication styles was more reflective of the variety that they would see in practice.

There were Turkers who were less strong in their skills. We were also initially concerned that these experienced teachers and practitioners were not always receptive to having their counselling skills criticized. This is where the variety of data streams helped. Rather than providing them with opinion-based subject-matter-expert feedback, which was sometimes met with skepticism, we instead simply provided them with data-informed activity metrics, along with peer-based reference criteria, with no hint of implied criticism. This more neutral approach promises to have the advantage of being less threatening to those in need of improvement.

Our explorations of sentiment analysis are incomplete. We explored a number of platforms. Our attempts were limited by the lack of a data scientist on our team so we were restricted to the use of quite simplistic approaches. We also noted that this is a rapidly changing area and we encountered 3 different software engines when working with a single cognitive computing platform (IBM Watson) over a period of 18 months. While it is good benefit from the
latest improvements, wholesale changes (see Figure 4) in methods of data manipulation taxed our abilities to be nimble.

**Figure 4: IBM Watson AlchemyLanguage, the first cognitive NLP engine we tried, now deprecated**

The general approach is very scalable however. The cognitive computing platforms are designed to accept millions of text fragments per second, which far exceeds the thousands of fragments that we generate per day. The costs are also trivial for a program our size, since they are aimed at huge social media platforms, rather than small educational research platforms. This very much plays to our advantage and we were able to get all our sentiment analysis done for free.

**Figure 5: some of the free analytics available now from IBM Watson sentiment analysis**
At first glance, the outputs (see Figure 5) from the sentiment analytics were useful and reasonably intuitive to grasp. More study is needed to validate how they relate to Turker and learner performance but the initial indications are very promising. It will also be interesting to explore whether simple color and sentiment representations (see Figure 6) are useful in real-time feedback during a Turker session.

**Figure 6: simple color coded sentiment analytics from Google's cognitive computing platform**
Our initial explorations suggest that this would create further cognitive overload if presented to the Turker and learner during their chat session, but may serve as a very useful indicator to the Scenario Director who is monitoring the performance of her Turkers from a second tier of monitoring.

The variety of data that could be incorporated into the feedback given during a Turker session fits with big data principles. As a simple example, we used a color coded bar to represent time in waiting. (see Figure 7)

*Figure 7: Scenario Director view of multiple channels with colored bars for wait times*
The learner who had been waiting longest was shown as a red bar, with a progression through orange, yellow, green and blue. This simple indicator was effective at showing both the Turker and the Scenario Director whether the channels were being responsive in a timely manner.

Conclusions

Big data analytics and their principles can be applied at a small group level. One does not need a data corpus, reminiscent of the Large Hadron Collider, in order to be precise. Our approach provides hundreds of thousands of data points per session, which can be analyzed post hoc or in real-time, and which provides far better richness than the simple pass/fail scores that were inherent in the SCORM days of educational software.

The volume of learners that can be accommodated with this approach is also very scalable and at greatly reduced costs, compared to OSCEs and questionnaires.

The velocity of the data, providing much more timely analyses to learners, Turkers and Directors, is greatly accelerated compared to using standard questionnaire-based cases.

The variety of data sources is much broader, looking at various aspects of professional practice including decision making and fact synthesis. Because the data format employed in the LRS is triplet-based and amenable to coding in RDF format, semantic analysis and knowledge mapping with the use of graph databases becomes feasible.

Because we are measuring who does what when in the activity streams, our analyses are not just dependent on content as a metric of performance. In certain clinically sensitive areas such as psychological counselling, being able
to abstract performance data separately from clinical data has significant benefits in the current concerns about data privacy.

In the world of rich data from the Internet of Things, sensor data from an even broader variety of instruments provides an enhanced capability to explore all aspects of physiological parameters. In certain contexts, such as research labs who need cheap and accessible methods to measure stress responses, or high-stakes examination environments who need objective methods of authentication (such as face and keyboard fist identification), all of which can now be incorporated into an open-standards educational research platform that supports precision metrics.

**Take Home Messages**

- Big data analytic principles can be applied to small groups.
- Cognitive computing techniques such as sentiment analysis are cheap to apply for education scenarios.

**Notes On Contributors**

Prof David Topps is Medical Director of OHMES (Office of Health & Medical Education Scholarship) for the Cumming School of Medicine at the University of Calgary. [https://orcid.org/0000-0003-3593-544X](https://orcid.org/0000-0003-3593-544X)

Ms Michelle Cullen, MScN, BN, is program lead in Behavioural Health Education for the School of Nursing at the University of Calgary. [https://orcid.org/0000-0002-4075-1445](https://orcid.org/0000-0002-4075-1445)

**Acknowledgements**

None.

**Bibliography/References**

Abdul-Kader, S. and Woods, J. (2015) ‘Survey on Chatbot Design Techniques in Speech Conversation Systems’, *International Journal of Advanced Computer Science and Applications*, 6(7). [https://doi.org/10.14569/IJACSA.2015.060712](https://doi.org/10.14569/IJACSA.2015.060712)

Advanced Distributed Learning (ADL) (2014) *Experience API v1.0.1*, GitHub. Available at: https://github.com/adlnet/xAPI-Spec/blob/master/xAPI.md (Accessed: 13 June 2019).

Amiel, T. and Reeves, T. (2008) ‘Design-Based Research and Educational Technology: Rethinking Technology and the Research Agenda’, *Journal of Educational Technology & Society*, 11(4), pp. 29–40.

Baepler, P. and Murdoch, C. (2010) ‘Academic Analytics and Data Mining in Higher Education’, *International Journal for the Scholarship of Teaching and Learning*, 4(2). [https://doi.org/10.20429/ijsotl.2010.040217](https://doi.org/10.20429/ijsotl.2010.040217)

Barab, S., Squire, K. and Barab, S. A. (2016) ‘Design-Based Research: Putting a Stake in the Ground’, *THE*
Collins, F. (2015) *Precision Medicine Initiative | National Institutes of Health (NIH), National Institutes of Health.* Available at: https://www.nih.gov/precision-medicine-initiative-cohort-program (Accessed: 1 February 2016).

Cullen, M., Sharma, N. and Topps, D. (2018) ‘Turk Talk: human-machine hybrid virtual scenarios for professional education’, *MedEdPublish*, 7(4). https://doi.org/10.15694/mep.2018.0000266.1

Cullen, M. and Topps, D. (2019) *OSCE vs TTalk cost analysis - Office of Health & Medical Education Scholarship.* Calgary. https://doi.org/10.5683/SP2/RJXRCW

Eagle, C. and Dubyk, A. (2017) *Precision Health for Alberta—A Prospectus.* Edmonton. https://doi.org/10.5683/SP2/RUAWJP

Ellaway, R. H. (2010) ‘OpenLabyrinth: An abstract pathway-based serious game engine for professional education’, in *Digital Information Management (ICDIM), 2010 Fifth International Conference on*, pp. 490–495. https://doi.org/10.1109/ICDIM.2010.5664241

Hart, S. (2016) ‘Precision Education Initiative: Moving Toward Personalized Education’, *Mind, Brain and Education.* https://doi.org/10.1111/mbe.12109

Hawgood, S., Hook-Barnard, I. G., O’Brien, T. C. and Yamamoto, K. R. (2015) ‘Precision medicine: Beyond the inflection point.’, *Science translational medicine.* American Association for the Advancement of Science, 7(300), p. 300ps17. https://doi.org/10.1126/scitranslmed.aaa9970

Kirkpatrick, D. L. and Kirkpatrick, J. D. (1998) *Evaluating training programs: the four levels.* Berrett. Koehler Publisher, Berkley, USA.

Kobielus, J. (2013) *The Four V’s of Big Data, IBM Big Data & Analytics Hub.* Available at: http://www.ibmbigdatahub.com/infographic/four-vs-big-data (Accessed: 18 June 2016).

Levitt, G. M. (2000) ‘The Turk, Chess Automaton’. McFarland & Company, Incorporated Publishers. Available at: http://dl.acm.org/citation.cfm?id=517712 (Accessed: 11 January 2016).

Liu, B. (2012) *Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies.* Morgan & Claypool Publishers. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Meiselman, E., Topps, D. and Albersworth, C. (2016) *Medbiq xAPI workshop, Medbiq Annual Conference.* Baltimore, MD. Available at: http://www.slideshare.net/topps/medbiq-xapi-workshop2b (Accessed: 13 June 2019).

Nirenburg, S., McShane, M., Beale, S., Jarrell, B., et al. (2009) ‘Integrating cognitive simulation into the Maryland virtual patient.’, *Studies In Health Technology And Informatics*, 142, pp. 224–229. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19377155.

Prabowo, R. and Thelwall, M. (2009) ‘Sentiment analysis: A combined approach’, *Journal of Informetrics.* Elsevier, 3(2), pp. 143–157. https://doi.org/10.1016/j.joi.2009.01.003
Quale, M. (2016) *Wikipedia, Wikipedia*. Available at: https://en.wikipedia.org/wiki/The_Turk (Accessed: 13 June 2019).

Stathakarou, N., Zary, N. and Kononowicz, A. A. (2014) ‘Beyond xMOOCs in healthcare education: study of the feasibility in integrating virtual patient systems and MOOC platforms’, *PeerJ*. PeerJ Inc., 2, p. e672. https://doi.org/10.7717/peerj.672

Topps, D., Ellaway, R. and Greer, G. (2019) *PiHPES Project: broadened perspectives on learning analytics*. Calgary. https://doi.org/10.5683/SP2/VDGGG3

Topps, D. and Ellaway, R. H. (2015) *Annotated Bibliography for OpenLabyrinth*. Calgary. https://doi.org/10.13140/RG.2.1.3897.8642

Topps, D., Ellaway, R. and Rud, S. (2019) 'OpenLabyrinth v3.4.1 source code - OLab Dataverse'. Calgary: OHMES, University of Calgary. https://doi.org/10.5683/SP2/7OFRIJ

Topps, D., Meiselman, E., Ellaway, R. and Downes, A. (2016) *Aggregating Ambient Student Tracking Data for Assessment, Ottawa Conference*. Perth, WA: SlideShare. Available at: http://www.slideshare.net/topps/aggregating-ambient-student-tracking-data-for-assessment (Accessed: 11 June 2019).

### Appendices

None.

### Declarations

*The author has declared that there are no conflicts of interest.*

*This has been published under Creative Commons “CC BY 4.0”* (https://creativecommons.org/licenses/by-sa/4.0/)

### Ethics Statement

The Conjoint Health Research Ethics Board (CHREB), University of Calgary has granted ethical approval to carry out this study. (Certificate REB17-1950.)

### External Funding

This paper has not had any External Funding

AMEE MedEdPublish: rapid, post-publication, peer-reviewed papers on healthcare professions’ education. For more information please visit www.mededpublish.org or contact mededpublish@dundee.ac.uk.