Final Report

Building ATN Institutional Capacity for Text Analytics

Commissioned Project, ATN Grants Scheme for Excellence in Learning and Teaching 2018

1 Project aim, rationale, need and sector readiness 3
1.1 Project Value Proposition 3
1.2 Project Focus: Text Analytics 3
1.3 Project Need 4
1.4 Sector Readiness 4
1.5 Capacity Building 5
1.6 Work Packages: Applications of Text Analytics 5
2 Project activities 7
3 Evaluation 7
3.1 Did the text analytics tools provide new insights, or accelerate analysis? 10
3.2 How well has the open source text analytics platform served its purpose as a catalyst for collaboration? 10
3.3 How did a university's organisational infrastructure need to adapt to make the most of text analytics? 10
3.4 How does the use of a writing analytics tool assist/impede students? 11
3.5 What are the key 'known unknowns' that the ATN should continue to investigate in future projects? 11
4 References 12
5 Appendices 15
5.1 Appendix 1: Investigators and Governance 15
5.2 Appendix 2: Text Analytics Pipeline 17

This page left blank intentionally

1 Project aim, rationale, need and sector readiness

Following consultation within the ATN Data Analytics Group (DAG), the *ATN Grants Scheme for Excellence in Learning and Teaching 2018* commissioned a 12-month (extended to 18-month) project focusing on *Text Analytics.* The project would seek to build the longer term technical and organisational capacity for ATN institutions to apply natural language processing techniques in order to collaboratively address strategic organisational priorities. The project was named *Higher Education Text Analytics (HETA)* with a website at http://heta.io.

1.1 Project Value Proposition

This was an opportunity for the ATN to accelerate the building of its text analytics capability by sharing expertise and technology across the institutions. As a short, applications focused project, the primary goal was to establish technical and operational processes to best leverage discrete areas of expertise for longer term mutual benefit. The primary sustainable output would be the creation of a collaborative group of experts and wider community of practice who could continue to work together, expanding beyond Text Analytics to other strategic foci beyond the project life span.

The project pooled resources across the ATN to address shared challenges and opportunities, to avoid parallel and duplicate efforts. The project aimed to maximise the benefits for the ATN through:

- A modular approach with a small, shared set of applications allowing flexibility for projects that are prototypic in ambition and well-defined in execution.
- Insight into analytics for all ATN members strengthened from the sharing of experience, findings and lessons across our individual contexts and implementations.
- A cooperative approach across partner institutions: contribution of expertise, resources and time from all on an equitable basis and as suitable, with funding distributed appropriately for the size and scope of the project.
- The project's recommendations would leverage the talents and resources of the ATN for a sector-wide collaborative approach to analytics to improve learning and teaching.

1.2 Project Focus and Rationale: Text Analytics

This project addressed the challenge of developing Text Analytics capacity in the ATN. The project aimed to build significant institutional capacity in the ability to provide automated analysis of three different kinds of text corpora. To do so requires a distinctive mix of expertise from Computing, Artificial Intelligence/Data Science, Learning Sciences, Academic Language & Literacy, Disciplinary Academics, and Writing Pedagogy.

This project addressed pressing needs on several fronts: pedagogical, research and capacity building, and understanding the commercial space:

The need to understand the marketplace. There is a rapidly growing commercial market in text analytics tools, and universities need to raise their literacy in order to ask probing questions about the scope and validity of such products. Vendors will often not disclose their algorithms, making informed decisions harder. This project will give the ATN first-hand experience of an open source tool developed at UTS, plus other tools that partners may have experience with. The project will therefore build a community to share expertise and experience about products and research/open source tools.

A shared, open source technical infrastructure will accelerate not only the rollout of end-user facing tools, but the underpinning development work, and the ability to test the replicability of findings. (See the Technical Appendix 3 for details of the *Text Analytics Pipeline*)

The need to build organisational capacity. Analyses of the state of the art in Australian higher education (Colvin et al. 2016; Siemens et al. 2013), and the partners' lived experience, emphasise that it takes time to build the expertise, and technical infrastructure, to gain value from learning analytics in any domain. Text Analytics is no exception: while university NLP research is commonplace (and an active field of inquiry in several of the ATNs), NLP tools are not yet commonplace in learning infrastructures, or the work practices of survey teams, or curriculum analysis teams. Analytics tools alone are not transformative, but in the longer term must be integrated into the relevant business processes and staff work practices. For WP1, tools should be integrated into the curriculum so that students and staff can see how the tools fit (e.g. aligned with the learning design, and the assessment regime). Students and staff must know how to access and interpret the data and feedback.

The need for common infrastructure to accelerate research and rollout. Collaboration around a common technical infrastructure will advance writing analytics in at least three ways:

1. *testing the current tools* in multiple contexts to verify their generalisability
2. *extending and refining them* for new subjects and contexts
3. *generating larger, shared datasets for researchers*, such as de-identified text corpora, human-annotated texts, and the output metadata from analytical tools. Some partners are also aims to introduce machine learning approaches, which improve with larger datasets.

The need to provide better writing feedback. Applied research and development at UTS, and internationally, provides preliminary evidence of positive student and academic responses to writing analytics, indicating the feasibility of scalable services (e.g. Buckingham Shum, et al. 2017; Gibson, et al. 2017; Knight, et al. In Press; International Writing Analytics Workshops 2016-18). Increasing university capacity to deliver formative, timely, personalised writing feedback to students, at scale, would be a significant achievement, and is simply not cost-effective using conventional methods. Thus, a unique opportunity exists for learning analytics, but only if it can be designed with pedagogical integrity to satisfy justified concerns around the limits of artificial intelligence, and academics' criteria for feedback.

1.3 Higher Education Sector Readiness

ATN readiness to share expertise and build capacity. A few leading US universities have begun to develop analytics infrastructures in recognition of the need for large scale text analysis (for both learning and research). University of Michigan, Carnegie Mellon and Arizona State University are leading examples, with whom we are in active dialogue. Strategically, therefore, the ATN needed to build its capacity to ensure that it has the academic and organisational depth of experience to shape the conversation with vendors and potential international partner universities. This project started with a working analytics infrastructure from UTS, which built on the international body of research into the effectiveness of writing analytics, plus know-how in participatory design processes and pedagogical practices for the use of these tools. Supported by UTS, partners would pilot the tools, and extend the technical and organisational infrastructure.

Readiness for analytics providing insight into writing quality. The higher education sector has now become familiar with the concept of making more effective use of student data, including several state of the art projects funded by the OLT (Colvin et al. 2016; Siemens et al. 2013; West et al. 2016) which this project's investigators have led/contributed to. However, the analytics that the majority of universities have access to on student writing is impoverished: beyond grades and plagiarism scores, learning platforms will typically show merely that a student has uploaded an assignment, posted to a

forum, or commented on a document. There is typically no indication regarding the quality, and certainly no formative feedback to the student on how to improve their writing. This project will equip the ATN to deploy more advanced analytics that should provide insights into the quality of writing, not just the quantity.

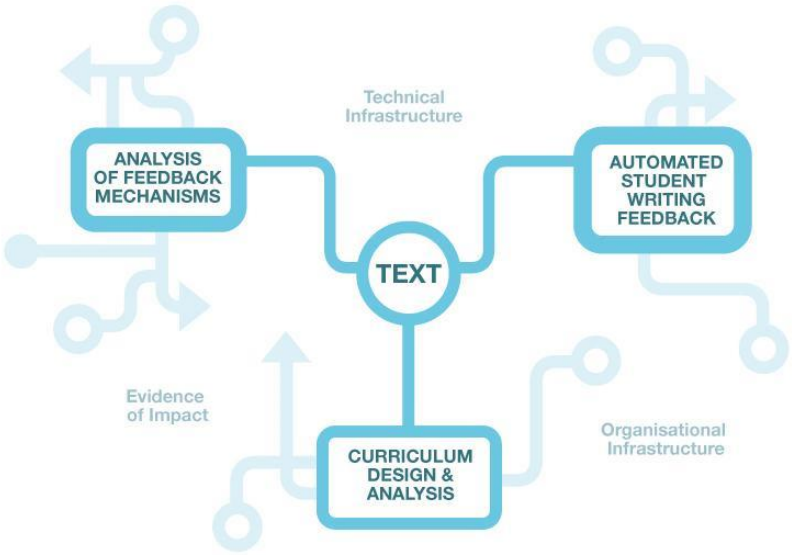
1.4 Capacity Building

The project sought to increase institutional capacity and commitments across the ATN network in areas such as:

1. **Technical Infrastructure.** A common technical platform will enable sharing, extending knowledge and techniques, integrating and delivering of code and algorithms as well as statistical solutions, end-user applications across the ATN.
2. **Organisational Infrastructure.** The power and potential of text analytics will begin to be embedded within ATN teaching and learning contexts, administrative processes and information systems, and a network will be created for sharing insights between ATNs.
3. **Evidence of Impact.** Initial measures of the impact of ATN-wide sharing of analytics knowledge and expertise on learning, teaching, and institutional capability will be established via evaluation of applications in strategic areas of common concern.

1.5 Work Packages: 3 Applications of Text Analytics

Three Work Packages (WPs) were created, applying text analytics to data that informs learning and/or teaching, in which two or more partners were interested to promote cross-ATN sharing and learning. In consultation with the respective ATN DVC's, the following WPs were chosen.



1.5.1 WP1: Automated Formative Feedback on Writing (UTS lead)

Writing, as both a process and a product, is fundamental to learning (Britton, 1970; Emig, 1971). Students use writing as a vehicle for learning disciplinary content by making notes from lectures and readings, reflecting on professional placements, summarising their thoughts, processing ideas, and formulating drafts. A key competency for students to build in their higher education, whether their career is academic or professional, is to write in a way that makes their thinking visible, through clear analytical and reflective texts.

However, this is a challenging skill to develop, and moreover, providing the quality feedback necessary for this development is demanding and time-consuming for educators. Despite the

evidence that rapid formative feedback is critical, providing timely feedback to intervene in and improve student writing prior to submission is prohibitively expensive at any kind of scale (consider for instance, large First Year classes). Appendix 1 provides a more detailed summary of related research.

UTS has over the last 3 years developed a prototype technical infrastructure for providing automated feedback to students on academic writing, grounded in the scholarship of writing education. The UTS *Academic Writing Analytics* (AWA) tool[1] is a web application for students to receive instant formative feedback on their writing, which calls NLP services in the UTS *Text Analytics Platform* (TAP: Appendix 3). Empirical evaluations show promise, but there remains much scope for improvement, which will be aided by the ATN collaboration which will provide more diverse test cases, models and user groups (Buckingham Shum et al, 2017; Gibson et al, 2017; Knight et al, In Press).

It was envisaged that TAP could also underpin all the WPs, with new user-facing applications being developed for WPs 2 and 3, and AWA being further refined.

1.5.2 WP2: Student Survey Feedback (QUT lead)

Qualitative survey results based in open ended text responses require a unique set of analytical tools. To date, this is largely a labour-intensive human process, which can of course yield insights, but is slow and costly to perform repeatedly at scale. Survey products are now offering (proprietary) text analysis algorithms, but with little validation evidence or scope for universities to modify, or share know-how.[2] In parallel, there is currently huge commercial interest in the use of text analysis to make sense of online customer feedback on products, opinions about companies or political parties, and so forth (e.g. Pang and Lee, 2008; Liu and Zhang, 2012).

In principle, such approaches should be able to yield insights in the specific context of higher education student feedback, but the ATN needs to develop its expertise in understanding specifically how to configure such tools to their needs.

The project will create pilot examples of the use of text analysis to create insights from open text responses and determine the potential for using similar approaches in other survey work of the university (e.g. surveys of staff, community, and businesses). This work package will produce insights into analysis (e.g. sentiment, topic), at different scales (e.g. courses, programs, schools and faculties).

1.5.3 WP3: Analysing Curriculum Materials (Curtin lead)

There is increasing interest and demand among higher education providers to understand the key concepts and pedagogical approaches adopted within a course. Work in curriculum mapping reveals a wealth of information related to program based assessment, scaffolding, support and communications, skills development and work integrated learning (Dawson and Hubball, 2014). However, to access such information is difficult given the complexity of curriculum offerings and sheer volume of curriculum materials (Sumsion & Goodfellow, 2004). This WP addresses these concerns by applying text analysis techniques to better automate and streamline how curriculum based information is extracted and reported.

---

1 UTS *Academic Writing Analytics*: https://utscic.edu.au/tools/awa

2 Product examples include: *Explorance* https://explorance.com/blue-text-analytics
Survey Analytics https://www.surveyanalytics.com/text-analysis.html
Ascribe https://goascribe.com/text-analytics/why-text-analytics

2 Project activities

Jan-Mar 2018: Project initiation and 2 day kick-off workshop (UTS)

Story: Kickoff Workshop, 6-7 Feb 2018

Apr-Sep 2018: Partners engaged their respective stakeholders, installed relevant text analytics software, and experimented. Partners shared their progress informally at face-to-face meetings (e.g. Australian Learning Analytics Summer Institute and Ascilite), and online via WP-specific videoconferences and the project blog

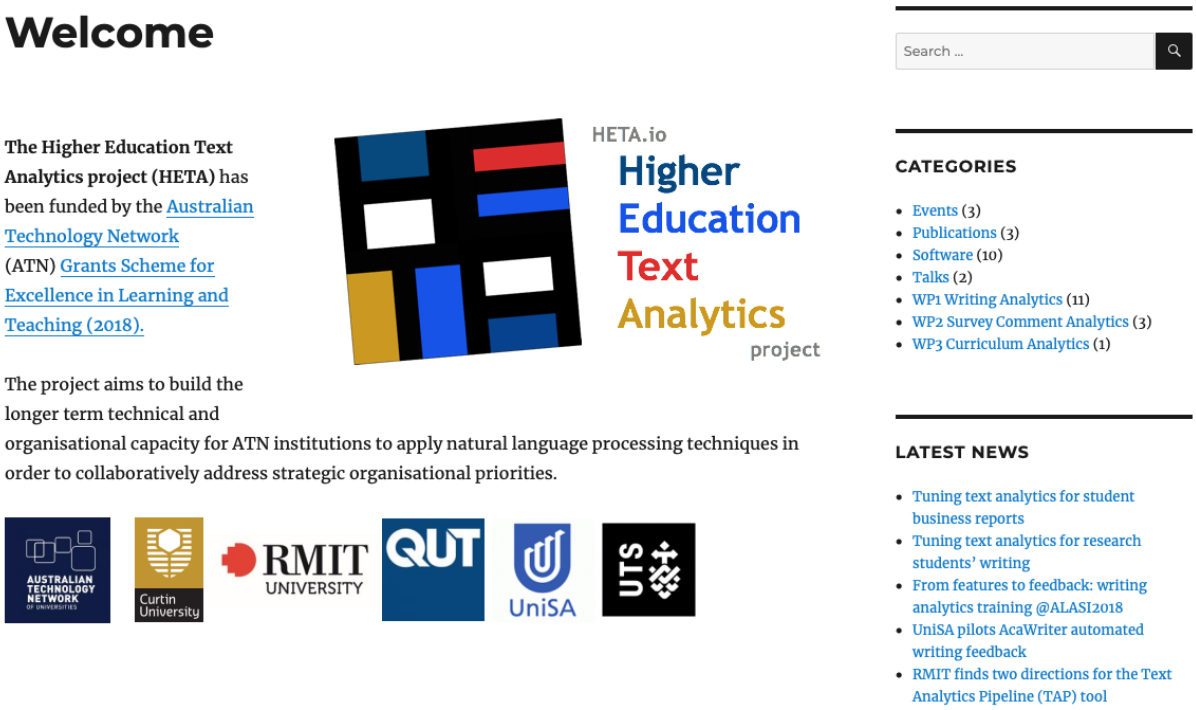
Dec 2018: Project wrap-up workshop (UTS)

Jan-May 2019: An extension was approved to use unspent funds to provide information resources and software documentation

3 Deliverables

3.1 Project website

The HETA project website is published at http://heta.io



3.2 Open source software and documentation

The software developed in the project has been documented, tested, and much of it already released open source: http://heta.io/technologies

- Announcement: Open source release of writing analytics infrastructure
- Progress with the Text Analytics Pipeline (TAP)
- Jupyter Python notebooks, available as part of workshop training resources

3.3 Staff development events and resources

3.3.1 Broad readership news stories

News stories were written for general audiences to help them understand how different universities were experimenting with text analytics, e.g.:

- Tuning text analytics for student business reports
- Tuning text analytics for research students' writing
- UniSA pilots AcaWriter automated writing feedback
- RMIT finds two directions for the Text Analytics Pipeline (TAP) tool
- The 'verbology' of learning (Curtin U)
- Building a bespoke language model for student feedback (UTS)
- A new tool for navigating feedback from student surveys (UTS)
- Contextualizable learning analytics for writing support

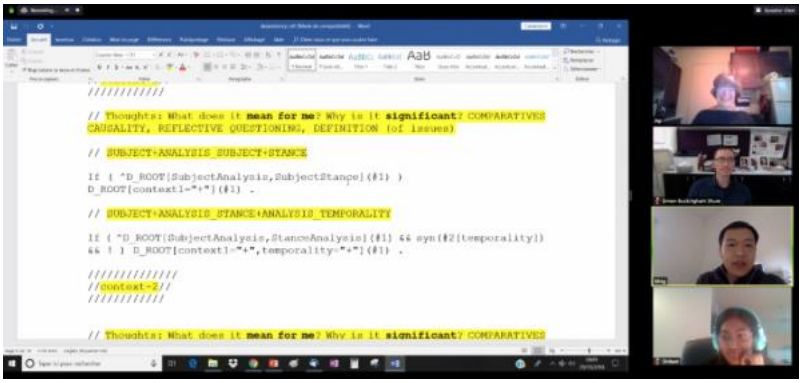
3.3.2 Writing analytics orientation for staff and students

The *AcaWriter* web application and its underlying text analytics platform *TAP* are the most mature infrastructures from the project, already deployed in teaching practice with positive outcomes. Educator resources are published on the HETA website, documenting how AcaWriter have been integrated into a range of student activities: http://heta.io/resources

The project also developed a set of more detailed introductory videos and more comprehensive website, tested with students, e.g. https://uts.edu.au/acawriter. These resources (including the editable master files) can be adapted by any partner.

Thirdly, the project developed technical training through a series of webinars which take text analytics and computational linguists into the details of the rhetorical parsing technology in TAP, used by AcaWriter:

- Writing analytics: online training in rhetorical parsing

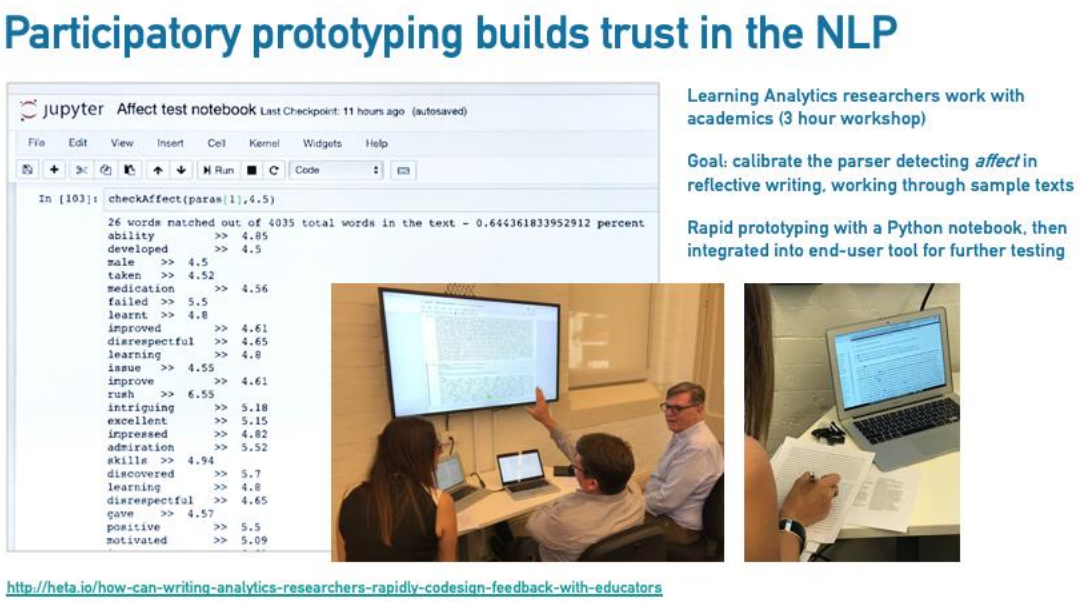


The project also ran face-to-face tutorials that introduced participants to bridging the gap between educational outcomes in writing, and low level textual features:

- TAP/AWA tutorial @LAK18 (Sydney, Mar. 2018)
- From features to feedback: writing analytics training @ALASI (Melbourne, Nov.2018)

Finally, the project documented the process of co-designing text analytics with academics, to help reveal what takes place 'backstage' in order to deliver tools that educators and students trust:

- Co-designing automated feedback on reflective writing with the teacher
- How can writing analytics researchers rapidly co-design feedback with educators?



3.4 Final report

Beyond these material deliverables, the remainder of this report discusses the wider outcomes from the collaboration, reflecting on the successes and failures.

4 Results

The questions that this project set out to investigate were as follows, introduced with the expectations we started the project with.

4.1 Did the text analytics tools provide new insights, or accelerate analysis?

This is the fundamental question across all work packages. Analytics tools may 'simply' make it possible to conduct labour-intensive analyses orders of magnitude faster, but they may also yield qualitatively different kinds of insights. To what extent can each work package report such progress after one year?

4.2 How well has the open source text analytics platform served its purpose as a catalyst for collaboration?

There are a number of specific criteria that can be used to evaluate its effectiveness for ATN collaboration and adoption:

- *Were the anticipated benefits of choosing a common platform realised, and from whose perspective (students; teachers; researchers; professional staff; IT divisions)?*
- *Did partners extend the infrastructure and in what ways?*
- *Did the ready availability of text analytics services lead to unexpected uses?*

4.3 How did a university's organisational infrastructure need to adapt to make the most of text analytics?

Does the adoption of text analytics in a given work package impact the associated administrative processes and information systems?

4.4 How does the use of a writing analytics tool assist/impede students?

The project provides limited resources to conduct in depth research into writing, but the UTS research team will assist ATN partners in planning their writing analytics pilots. These could investigate questions ranging from the relationship of analytics to grades as an indicator of writing quality, and other indicators such as student attitudes to writing (self-report via interview; writing attitudes survey instrument), and changes in the effectiveness of their writing habits (e.g. number of revisions made, assessed by system logs and self-report). We are also interested in student attitudes to automated feedback including any concerns (self-report via interview; survey; online feedback tools). In the course of working with writing analytics, the UTS team has received ethics approval for working with student texts and students themselves, and will be happy to share their experience with partners.

4.5 What are the key 'known unknowns' that the ATN should continue to investigate in future projects?

It is not necessarily the case that all work packages will report successes. It may be the case that some are deemed not worthy of continued exploration, while others are demonstrating exciting new advances that merit additional resources to exploit. The project will identify the most promising avenues for further work.

5 References

Allen, L. K., & Perret, C. A. (2016). Commercialized Writing Systems. In Crossley, S. and McNamara, D. (Eds.), *Adaptive Educational Technologies for Literacy Instruction*. London: Taylor and Francis, p.145.

Andrews, R. (2009). *Argumentation in Higher Education: Improving practice through theory and research*. Routledge.

Bazerman, C. & Russell, D. (1994). (Eds.) *Landmark Essays on Writing Across the Curriculum*. Davis, CA: Hermagoras.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, *38*(8), 941–956.

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: disruptive effects of assessment patterns. *Higher Education Research & Development*, *34*(1), 45–59.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698–712.

Britton, J. (1970). *Language and learning.* London: Allen Lane.

Buckingham Shum, S. (2016). *Algorithmic Accountability for Learning Analytics.* Invited Talk, Institute of Education, University College London (20 April 2016). http://bit.ly/aala2016

Buckingham Shum, S., Á. Sándor, R. Goldsmith, R. Bass and M. McWilliams (2017). Towards Reflective Writing Analytics: Rationale, Methodology and Preliminary Results. *Journal of Learning Analytics,* 4, (1), 58–84. http://dx.doi.org/10.18608/jla.2017.41.5

Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., K. ., Betik, D., & Crossley, S. (2016). *Critical Perspectives on Writing Analytics.* Workshop, 6th International Conference on Learning Analytics and Knowledge, Edinburgh, UK.

Calvo, R., Aditomo, A., Southavilay, V., Yacef, K. (2012). The use of text and process mining techniques to study the impact of feedback on students' writing processes. *ICLS 2012: 10th International Conference of the Learning Sciences*, Sydney: International Society of the Learning Sciences.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, *36*(4), 395–407.

Carroll, D. (2013). *Graduate Course Experience 2013*. Melbourne, Australia: Graduate Careers Australia.

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385–405.

Coates, H. B. (2009). Engaging students for success: Australasian Student Engagement Report. Australian Council for Educational Research (ACER). http://research.acer.edu.au/cgi/viewcontent.cgi?article=1017&context=higher_education

Crisp, G. (2011). *Rethinking assessment in the participatory digital world – Assessment 2.0.* Australian Learning & Teaching Council, National Teaching Fellowship: Project & Final Report: http://www.transformingassessment.com

Colbran, S. (2011). Evaluation of the usefulness of self-assessment, peer assessment and academic feedback mechanisms. *Proceedings ATN Assessment Conference,* Perth.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Buckingham Shum, S., Nelson, K., Alexander, S., Lockyer, L., Kennedy, G., Corrin, L., & Fisher, J. (2016). *Student retention and learning analytics: a snapshot of Australian practices and a framework for advancement.* Canberra, ACT: Australian Government Office for Learning and Teaching. Deliverables & Final Report: http://he-analytics.com

Dawson, S., & Hubball, H. (2014). Curriculum Analytics: Application of Social Network Analysis for Improving Strategic Curriculum Decision-Making in a Research-Intensive University. *Teaching and Learning Inquiry: The ISSOTL Journal*, 2(2), 59-74.

Derntl, M., & Calvo, R. A. (2011). E-learning frameworks: facilitating the implementation of educational design patterns. *International Journal of Technology Enhanced Learning*, 3(3), 284–296.

Dikli, S. (2006). An Overview of Automated Scoring of Essays. *The Journal of Technology, Learning and Assessment*, *5*(1).

Drury,, H. (2013). *An online writing centre for undergraduate engineering students: a one stop shop (iWrite).* Canberra, ACT: Australian Government Office for Learning and Teaching. Final Report CG10-1713. http://www.olt.gov.au/system/files/resources/CG10_1713_Drury_Report_2013.pdf

Emig, J. (1971). *The Composing Processes of Twelfth Graders*. Illinois: NCTE.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Ericsson, P. F., & Haswell, R. H. (2006). Machine scoring of student essays: Truth and consequences. Utah State University Press.

Ganobcsik-Williams, L. (Ed.). (2006). *Teaching Academic Writing in UK Higher Education*. Basingstoke, UK: Palgrave Macmillan.

Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C. and Knight, S. (2017). Reflective Writing Analytics for Actionable Feedback. Proceedings of LAK17: 7th International Conference on Learning Analytics & Knowledge, March 13-17, 2017, Vancouver, BC, Canada. (ACM Press). DOI: http://dx.doi.org/10.1145/3027385.3027436.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

International Writing Analytics Workshops 2016-17: http://wa.utscic.edu.au/events/lak16wa and http://wa.utscic.edu.au/events/lak17wa

Johnstone, K. M., Ashbaugh, H., & Warfield, T. D. (2002). Effects of repeated practice and contextual-writing experiences on college students' writing skills. *Journal of Educational Psychology*, *94*(2), 305.

Kellogg, R. T., & Raulerson, B. A. (2007). Improving the writing skills of college students. *Psychonomic Bulletin & Review*, *14*(2), 237–242.

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. (In Press). Designing Academic Writing Analytics for Civil Law Student Self-Assessment. *International Journal of Artificial Intelligence in Education,* (Special Issue on Multidisciplinary Approaches to Reading and Writing Integrated with Disciplinary Education, Eds. D. McNamara, S. Muresan, R. Passonneau & D. Perin). DOI: http://dx.doi.org/doi:10.1007/s40593-016-0121-0

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, *23*(2), 157–172.

Li, J., & Luca, R. D. (2014). Review of assessment feedback. *Studies in Higher Education*, *39*(2), 378–393. https://doi.org/10.1080/03075079.2012.709494

Lillis, T., & Turner, J. (2001). Student writing in higher education: contemporary confusion, traditional concerns. *Teaching in Higher Education*, *6*(1), 57–68.

Liu B., Zhang L. (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) *Mining Text Data*. Springer, Boston, MA

McGee, P., Drury, H., Jones, J., O'Carroll, P., et al., (2009). *Creating a student centred online learning environment for report writing in the sciences and engineering.* Canberra, ACT: Australian Government Office for Learning and Teaching. Final Report CG6-30.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.

National Commission on Writing. (2003). *Report of the National Commission on Writing in America's Schools and Colleges: The Neglected "R," The Need for a Writing Revolution*. College Board.

Norton, L. S. (1990). Essay-writing: what really counts? *Higher Education*, *20*(4), 411–442.

OECD. (2013). PISA 2015: Draft reading literacy framework. OECD Publishing.

Bo Pang and Lillian Lee (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, Vol. 2: No. 1–2, pp 1-135. http://dx.doi.org/10.1561/1500000011

Rowe, A. D., & Wood, L. N. (2009). Student perceptions and preferences for feedback. *Asian Social Science*, *4*(3), 78.

Ryan, M., Michael Ryan, Nan Bahr, Dean Brough, Suzanne Carrington, Tina Cockburn, Christy Collis, Ingrid Larkin, Robyn Nash, Melinda Shirley, Georgina Barton, Louise Bow, Natalie Gamble, Jill Ryan (2013). *Developing a Systematic, Cross-Faculty Approach to Teaching and Assessing Reflection in Higher Education.* Canberra, ACT: Australian Government Office for Learning and Teaching. Final Report: http://www.olt.gov.au/system/files/resources/PP9_1327_Ryan_report_2012.pdf

Shermis, M. D., & Burstein, J. (2013). *Handbook of Automated Essay Evaluation: Current Applications and New Directions*. Routledge.

Siemens, G., Dawson, S. and Lynch, G. (2013). *Improving the Quality and Productivity of the Higher Education Sector.* Canberra, ACT: Australian Government Office for Learning and Teaching. Final Report: http://www.olt.gov.au/system/files/resources/SoLAR_Report_2014.pdf

Sumsion, J., & Goodfellow, J. (2004). Identifying generic skills through curriculum mapping: a critical evaluation. *Higher Education Research & Development*, 23(3), 329-346.

West, Deborah, Henk Huijser, David Heath, Alfred Lizzio, Danny Toohey, Carol Miles, Bill Searle, & Jürg Bronnimann (2016). *Learning Analytics: Assisting Universities with Student Retention.* Canberra, ACT: Australian Government Office for Learning and Teaching. Deliverables & Final Report: http://www.letstalklearninganalytics.edu.au

Winters, N., & Mor, Y. (2008). IDR: A participatory methodology for interdisciplinary design in technology enhanced learning. *Computers & Education*, 50(2), 579–600.

Winters, N., & Mor, Y. (2009). Dealing with abstraction: Case study generalisation as a method for eliciting design patterns. *Computers in Human Behavior*, 25(5), 1079–1088.

6 Appendices

6.1 Appendix 1: Investigators and Governance

This project was submitted by all of the ATN universities in recognition of the strategic importance of the role of natural language processing in the proposed pilot projects. Table 2 summarises the team structure.

PI: Simon Buckingham Shum (UTS) Technical Lead: Andrew Gibson (QUT) Project Manager: Gabrielle Gardiner (UTS)				
Curtin	QUT	RMIT	UniSA	UTS
Co-I: David Gibson	Co-I: Sam Nielson Andrew Gibson	Co-I: Pablo Munguia David McLay	Co-I: Shane Dawson Jing Gao	Co-I: Simon Knight
Other staff as required for each WP	*Other staff as required for each WP*	*Other staff as required for each WP*	*Other staff as required for each WP*	*Other staff as required for each WP*

PI: Prof. Simon Buckingham Shum holds a Chair in Learning Informatics at the University of Technology Sydney, where he is Director of the Connected Intelligence Centre. He leads the university's learning analytics, a topic on which he is a regular keynote speaker. He was a co-founder and a Vice-President of the Society for Learning Analytics Research (SoLAR). He coordinated the international network of Learning Analytics Summer Institutes (2013-16), Program Co-Chair for the 2012 & 2018 *International Conferences on Learning Analytics & Knowledge (LAK)*. He has been a PI and Co-I on numerous projects ($10.8M total), including three OLT Projects on Learning Analytics.

Co-I (Project Technical Lead): Andrew Gibson is a Research Fellow in Writing Analytics at the UTS Connected Intelligence Centre, and from 2018 a Lecturer at QUT Information Systems School. His research focuses on reflective writing analytics for psychosocial meaning, and he has written software that utilise a range of natural language processing and machine learning techniques for this purpose. With an additional interest in transdisciplinarity, he works across both educational and computational domains. He has been developing Java applications for educational organisations for over 20 years, and has been working with Scala, AWS and learning analytics for the past 4 years.

Co-I (UniSA lead): Prof. Shane Dawson is the Director of the Teaching Innovation Unit and Professor of Learning Analytics at the University of South Australia. The TIU functions to lead and transform learning and teaching practices at UniSA through collaborative capacity building processes and researching the application of digital technologies in higher education.

Co-I (Curtin Lead): Prof. David Gibson is Director of Learning Futures at Curtin University, a team focusing on strategic innovations that advance the mission of the university. The team shapes the future of learning and teaching at the university through human and technological capacity building and promotes faculty-based research as well as continuous improvement using learning analytics. The team's programs influence and impact students, staff and the broader community by leading and managing early stage innovation projects that range across formal and informal learning innovations, pathways & partnerships, the UniReady enabling program, and learning analytics. The team's data

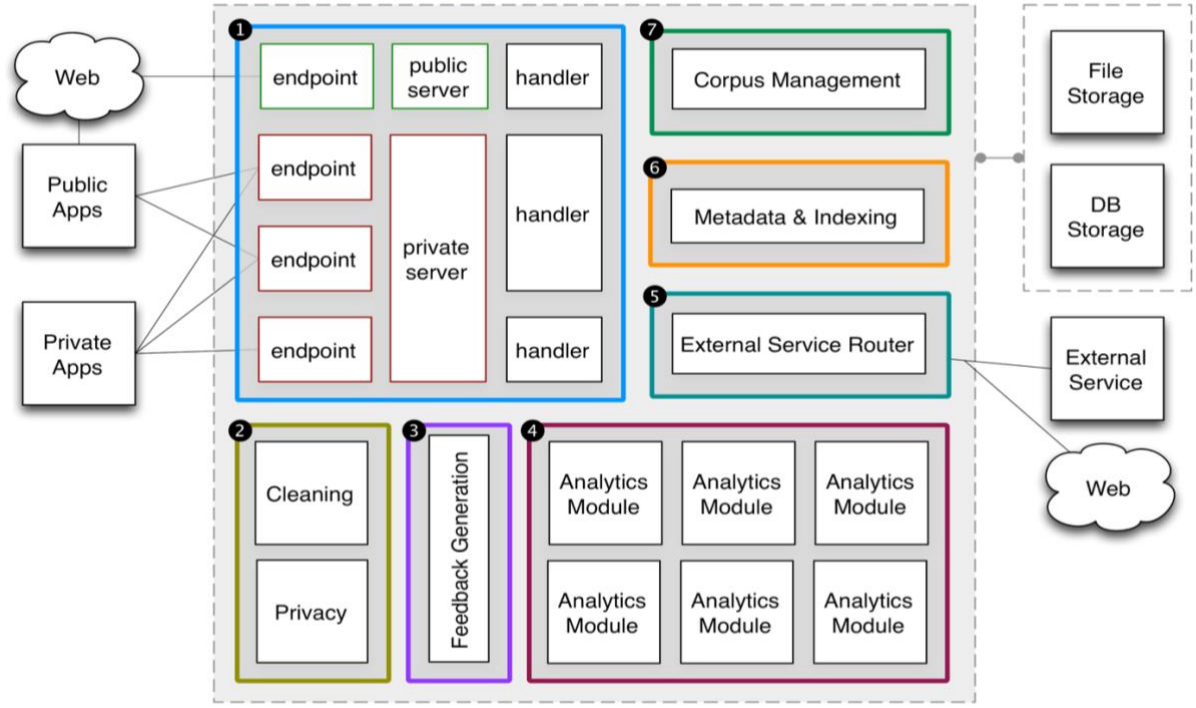
scientist helps lead the university's growth in capacity in analytics-based insights through a community of practice, which has participation from five senior executive areas of the university. David is currently in the final stages of establishing the UNESCO Chair for Data Science in Higher Education Learning and Teaching.

Co-I (UTS Lead): Simon Knight is a Lecturer in Writing Analytics at the UTS Connected Intelligence Centre. His research focuses on the relationship of analytics to epistemology, pedagogy and assessment, discourse analytics, and epistemic cognition, particularly around information seeking. He is active in the Learning Analytics and Learning Sciences communities, co-chairing the *ICLS14 Workshop on Learning Analytics for Learning and Becoming in Practice, LAK15 Workshop on Temporal Analyses of Learning Data,* and *LAK16 Workshop on Critical Perspectives in Writing Analytics.*

Co-I (RMIT Lead): Dr. Pablo Munguia is an Associate Professor and Director of Learning Environments and Analytics within the Education Portfolio at RMIT. The Learning Analytics unit at RMIT is in charge of student surveys, collaborating with academics through communities of practice and providing student-driven mechanisms through models and tools. Work from this unit is presented at LASI and LAK conferences. Learning analytics is coupled with the Learning Environments uni by providing insights into course design, correct pedagogical use of the LMS and digital tools therefore closing the loop from insights to action in the learning and teaching space. Pablo also maintains an active research profile and is editor in chief of the *Journal for Experimental Marine Biology and Ecology*, one of the world's top 3 marine biology journals.

6.2 Appendix 2: Text Analytics Pipeline

FUNCTIONALITY: The project will utilise as a starting point, existing technical infrastructure developed at UTS called the Text Analytics Pipeline (TAP). TAP is a web application, written in Scala using the Play framework and GraphQL. It provides text analytics services to multiple clients via a GraphQL API [see diagram module 1], and facilitates the connection to multiple external text analytics services [module 5]. It provides text cleaning, de-identification, indexing, metadata management, and the ability to work with multiple corpora as well as single documents [modules 2, 6, 7]. It can analyse documents in near real time or be used to batch process documents [module 4]. TAP can utilise external storage such as S3 for file storage or PostgreSQL for database storage.



SUSTAINABILITY: With the exception of 3rd party analytics services, all software used by the project is Open Source and is currently available for download via GitHub. The current TAP software runs on the JVM and can be provided as a docker image, providing a wide range of deployment options (most enterprise platforms). The modular architecture of the software allows for key services to be run as microservices allowing easy interfacing with other technologies and other software written in other languages.

SCALABILITY: A reactive functional design approach allows for vertical scalability, and deployment on Amazon Web Services (AWS), allows for scalability of compute capacity. A streaming architecture allows for horizontal scaling and running of TAP on a cluster, and for direct communication with highly scalable big data analytics platforms such as Apache Spark.

EXPERTISE: The technical lead is Andrew Gibson (see bio section). His expertise will be supported by developers at each site. University specific support will be provided through key IT personnel at each site.

OPEN SOURCE: All development technologies are or will be open source and freely available. TAP is licensed under Apache 2.0.