Augmenting Pedagogic Writing Practice with Contextualizable Learning Analytics

Antonette Aileen Shibani Michael Xavier

B.E. (with Distinction) in Computer Science Engineering
Anna University (India), 2010

M.Sc. in Information Studies
Nanyang Technological University (Singapore), 2012

Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Technology Sydney

Connected Intelligence Centre

October 2019

Antonette Shibani (2019). Augmenting Pedagogic Writing Practice with Contextualizable Learning Analytics.

Supervisors:

Prof. Simon Buckingham Shum, University of Technology Sydney

Dr. Simon Knight, University of Technology Sydney

Certificate of Original Authorship

I, Antonette Aileen Shibani Michael Xavier declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Connected Intelligence Centre at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Shibani

Date: 16/10/2019

Dedicated to my parents

Prof. Dr. Leema Rose, my darling mother and confidante, who taught me kindness, hard work, and gratitude towards all things in life.

Mr. Michael Xavier, my loving father and pillar of support, who seeded confidence and trust in his little girl to conquer the big wide world.

Acknowledgements

I owe a ton of gratitude to people for their support in my doctoral learning journey. The journey has transformed me as a person and a researcher to become more critical, inquisitive, and reflective in my thinking.

Firstly, my sincere gratitude to my supervisors Prof. Simon Buckingham Shum and Dr. Simon Knight for their guidance throughout my PhD – I have definitely learnt from the best. Their sense of responsibility in going above and beyond to assist a student never ceases to amaze me.

I remember the warm welcome I received from Simon Buckingham Shum on the day I arrived at UTS, and the warmth has never faded since. With his ideological and innovative outlook, Simon has shaped my thinking and made me question my understandings. Thank you, Simon, I have been incredibly lucky to work with you, anrd learned so much from you. Simon Knight has been an amazing colleague and mentor, to whom I can turn to at any time. With his pragmatic thinking and way to question everything, he has taught me the value in constructive conflict, amongst many more. Thank you, Simon, for putting up with my seemingly naïve questions at times and guiding me to be a responsible researcher. My constant scramble having to address both Simon and Simon without confusing others hasn't been so bad after all, with the other perks that came with them!

Thanks to the enthusiastic instructors who brought life to the project in their classrooms and trusted me with their students. They are the true champions in making the innovations effective by making them practically feasible. Special thanks to Dr. Philippa Ryan who first implemented the project with her students in Law and the tutors. My sincere thanks go to Dr. Nicole Sutton, Raechel Wight and the tutors for their dedicated involvement in the project with Accounting. I'm also grateful to the students who participated in the research studies that form the basis of my empirical work. Their feedback and involvement have contributed to improvements for the future and helped us rethink what we were building.

My deep appreciation goes to my colleagues at the Connected Intelligence Centre for the conversations and feedback that have tuned my thinking in my research and beyond – in particular, Kirsty Kitto, Roberto Martinez, and Andrew

Gibson. My sincere thanks to CIC's web developer Radhika Mogarkar, who played a big role in bringing the tool into production use for large cohorts of students. Thanks also to Gabrielle, Emma, and Ratha for taking care of anything that needs fixing. Thanks to the #PhDGang for making the hard PhD life enjoyable with your presence and cheerfulness. To all CICers, thanks for being part of the wonderful family of lab colleagues.

I'm grateful for the support of Agnes Sandor and Claude Roux from Naver Labs Europe (previously, Xerox Research Centre Europe) for developing the Natural Language Processing Parser on which the writing feedback tool is built. My sincere thanks goes to my internal reviewers Lori Lockyer and Jo McKenzie, UTS, for reviewing my thesis and providing feedback on my initial draft. Thanks to Elizabeth Koh from the National institute of Education, Singapore, for paving the way for my research career. Thanks also to the reviewers and friends who have provided valuable feedback on my work.

I gratefully acknowledge the financial support and scholarships I received towards presenting my PhD work at research conferences. Thanks to ACM-W, Society for Learning Analytics Research, Asia-Pacific Society for Computers in Education, UTS Vice-Chancellor's conference fund, and the Connected Intelligence Centre, UTS for the sponsorship.

I have been tremendously supported by my family, to whom I'm very grateful. Huge thanks to my husband Vignesh who endured the best and worst of me during this challenging phase of life. He's been that caring, respectful, and loving partner everyone yearns to get, and I'm incredibly lucky to have him stand by me. Thanks to my parents, brother Sheen and his family, and relatives back in India for their encouragement, prayers and constant support throughout the years. Thanks to my dear friends who are always there for me. It's their emotional support that kept me sane and smiling during the long hours I put into this PhD. Last but not the least, I believe that nothing would have been possible if not for the blessings of the Almighty. Thank you, God, for the strength, opportunities, and protection you offer me every single day.

Preface

The following peer-reviewed publications produced during the PhD candidature contribute to my thesis.

Journal Papers

- **Antonette Shibani**, Simon Knight, and Simon Buckingham Shum (Under review) Educator Perspectives on Learning Analytics in Classroom Practice.
- Simon Knight, **Antonette Shibani**, Sophie Abel, Andrew Gibson, Philippa Ryan, Nicole Sutton, Raechel Wight, Cherie Lucas, Ágnes Sándor, Kirsty Kitto, Ming Liu, Radhika Vijay Mogarkar, and Simon Buckingham Shum (Under review) AcaWriter: A Learning Analytics Tool for Formative Feedback on Academic Writing.
- Simon Knight, Andrew Gibson, and **Antonette Shibani** (Under review) Implementing Learning Analytics for Learning Impact: Taking Tools to Task.

Conference Proceedings

- Antonette Shibani (Under review). Automated Revision Graphs: A novel visualization method to study written text as graph.
- **Antonette Shibani**, Simon Knight and Simon Buckingham Shum (2019). Contextualizable Learning Analytics Design: A Generic Model and Writing Analytics Evaluations. In Proceedings of the International Conference on Learning Analytics and Knowledge (LAK'19), Tempe, Arizona. ACM. https://doi.org/10.1145/3303772.3303785
- Simon Knight, **Antonette Shibani** and Simon Buckingham Shum (2018). Augmenting Formative Writing Assessment with Learning Analytics: A Design Abstraction Approach. Full Paper presented at the 13th International Conference of Learning Sciences (ICLS'18), Festival of Learning cross-over track, London.
- **Antonette Shibani**, Simon Knight and Simon Buckingham Shum (2018). Understanding Revisions in Student Writing through Revision Graphs. Poster presented at the 19th International Conference on Artificial Intelligence in Education (AIED'18), London.

- **Antonette Shibani** (2018). Developing a Learning Analytics Intervention Design and tool for Writing Instruction. In Companion Proceedings of the Eighth International Conference on Learning Analytics & Knowledge (LAK '18), Sydney, Australia
- **Antonette Shibani** (2018). AWA-Tutor: A Platform to Ground Automated Writing Feedback in Robust Learning Design (Demo). In Companion Proceedings of the Eighth International Conference on Learning Analytics & Knowledge (LAK '18), Sydney, Australia.
- **Antonette Shibani** (2017). Combining automated and peer feedback for effective learning design in writing practices. In DSC Proceedings of the 25th International Conference on Computers in Education, New Zealand
- **Antonette Shibani**, Simon Knight, Simon Buckingham Shum and Philippa Ryan (2017). Design and Implementation of a Pedagogic Intervention Using Writing Analytics. In Proceedings of the 25th International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education

Workshops

- **Antonette Shibani**, Ming Liu, Christian Rapp and Simon Knight (2019). Advances in Writing Analytics: Mapping the state of the field. Workshop chaired at the Ninth International Conference on Learning Analytics & Knowledge LAK'19, Tempe, Arizona.
- Ming Liu, Simon Knight, **Antonette Shibani** and Sophie Abel (2018, November). From features to feedback: Designing automated feedback for student writing. Workshop chaired in ALASI'18, Melbourne.
- **Antonette Shibani**, Sophie Abel, Andrew Gibson and Simon Knight (2018, March). Turning the TAP on Writing Analytics. Workshop chaired in the Eighth International Conference on Learning Analytics & Knowledge, Sydney.
- Andrew Gibson, **Antonette Shibani** and Sophie Abel (2017, November). An Introduction to Text Analysis for Learning Analytics. Workshop chaired in the Australian Learning Analytics Summer Institute ALASI'17, Brisbane.

Sources & Original Work

The thesis draws from original material of my own work in the publications listed above, to which I retain copyright permissions as an author of the work. Such prior publications when used in the thesis are explicitly cited where appropriate, and are not used in entirety. Publications of external authors are credited throughout the thesis with citations in text and references at the end of the thesis. Figures from external sources where authors granted permission for usage are cited in their captions.

Ethics

The research designs in the study are approved by the University of Technology Sydney's Human Research Ethics Committee, and are based on ETH16-675: Academic Essay Self-Assessment Project (AESA Project). The ethics applications subsequently revised for the studies in the thesis have the protocol numbers ETH17-1176, ETH18-2263, ETH18-3080, and ETH19-3475. The most recent participant information sheets and consent forms can be requested by email[1].

[1] antonette.shibani@gmail.com

Table of Contents

List of Figures

List of Design Representations

List of Tables

List of Resources

Several resources have been produced as part of this thesis, which are publicly available (Check individual resource for copyright details).

Videos:

- AWA-Tutor demo: https://www.youtube.com/watch?v=K212XabCL5w
- AcaWriter demo: https://www.youtube.com/watch?v=P8_WuzQUs8s
- Introductory videos by the Law instructor: https://www.youtube.com/watch?list=PLcS9QDvS_uS6V-KknW098LQ5ru4wftgJ5&v=zUBcEJABCB0
- Introductory video by the Accounting instructor: https://www.youtube.com/watch?v=yAfu1fWNCNg

User Guides

- Appendix A: Hand out on rhetorical moves provided to Law students
- Appendix B: Self-Evaluation Exercise (SEE) prompts for rhetorical moves using AcaWriter
- Learning Design for Law (Includes a downloadable guide for students): http://heta.io/resources/wawa-improve-sample-text-plus-peer-discussion-civil-law/
- Learning Design for Accounting (Includes a downloadable guide for students): http://heta.io/resources/wawa-improve-business-report-writing-accounting/

Questionnaires

Appendix C: Instructor Interview Guide

Appendix D: Tutor Feedback Questionnaire

Abbreviations

AES – Automated Essay Scoring

AI – Artificial intelligence

API – Application programming interface

AWA – Academic Writing Analytics (Tool)

AWE – Automated Writing Evaluation

AZ – Argumentative Zoning

CARS – Create a Research Space (Model)

CLAD – Contextualizable Learning Analytics Design

DBR – Design Based Research

EAP – English for Academic Purposes

ECD – Evidence Centred Design

EFL – English as a Foreign Language

ICT – Information and Communication Technology

ITS – Intelligent Tutoring Systems

LA – Learning Analytics

LD – Learning Design

NLP – Natural Language Processing

OECD – Organisation for Economic Co-operation and Development

RQ – Research Question

TAP – Text Analytics Pipeline

TEL – Technology Enhanced Learning

UI – User Interface

WA – Writing Analytics

XIP – Xerox Incremental Parser

Abstract

Academic writing is a key skill that contributes to essential learning outcomes for higher education students. Despite its importance, students often lack proficiency in writing and find it challenging to learn. While previous research suggests that students' writing skills are enhanced through formative feedback, the time-consuming nature of providing formative feedback on individual student drafts, especially in large cohorts, makes it impractical for educators to provide detailed writing support in this way. A promising approach, therefore, is the use of *writing analytics* to provide automated formative feedback on writing. This particular form of *learning analytics*, using computational techniques and natural language processing, provides timely, immediate, and consistent automated feedback to help students improve their writing. However, for such tools to work effectively in pedagogic settings, and be adopted by practitioners, academics need to feel a sense of ownership over how the tool fits into their practice. This recognition motivates an increased emphasis on aligning learning analytics applications with learning design, so that analytics-driven feedback is congruent with the pedagogy and assessment regime.

The thesis investigates how writing practice can be augmented with a writing analytics tool called 'AcaWriter' by aligning it with learning design. The approach is evaluated across two disciplines in authentic higher educational settings using a design-based research approach. Mixed methods and multiple data sources are used to examine how students perceive and interact with automated feedback, and revise their writing. Based on this analysis, the thesis provides empirical evidence that students found the writing intervention and automated feedback from AcaWriter useful, and improved their subject-related writing skills, thus validating its applicability in writing contexts. It identifies varied levels of student engagement with automated feedback and ways to scaffold its application for effective use. Cross-fertilizing research and practice, the key insights gained from these design iterations are formalised as the *Contextualizable Learning Analytics Design* model. The model clarifies how the features, feedback and learning activities around AcaWriter can be tuned for different pedagogical contexts and assessment regimes, by co-designing them

with educators. The thesis also studies the perspectives of educators, who play a key role in implementing such learning analytics innovations in their classrooms. The thesis advances theory and practice in the development of flexible learning analytics applications, capable of providing meaningful, contextualized support that enhances learning, and adoption by practitioners in authentic practice.

Chapter 1: Introduction

This chapter outlines the objectives of the research and sets the context for the thesis. Section 1.1 describes the context and motivation of the research[2] and explains the research goals of the thesis. Section 1.2 then states the significance and contributions of this research and defines the scope of the study. Finally, Section 1.3 provides an outline of the remaining chapters of the thesis.

1.1 Context and Motivation

Writing is an essential skill that university students should develop for succeeding in their academic and professional lives. However, many students are not proficient in the kinds of writing required in academic settings, particularly lacking expertise in the creation of coherent and engaging writing that develops argumentative claims (Horstmanshof & Brownie, 2013; Krause, 2001; National Commission on Writing, 2003). Limited support for academic writing makes it challenging to learn for students in higher education. There is a recognised need to help students with their academic writing in an ongoing and integrated way (Graham & Perin, 2007; Wingate, 2012). This support can be provided to students with *formative feedback* on their writing. Through such formative feedback, students can close the feedback loop by applying the feedback that they receive to improve their work, by addressing the gap between their current performance and instructor expectations. For instance, formative feedback on students' drafts can help them to improve their final writing through the application of the feedback in subsequent revisions of their drafts. This approach arguably results in greater impact on students' learning than summative assessment, which by contrast, focuses on evaluating the students' achievement after submission at the end of their learning, not allowing room for further improvements (Sadler, 1989).

To provide such continued support to students in the form of formative feedback on writing, many approaches exist as described in Section 2.2, each with

[2] A part of this Section 1.1 has been published in the International Conference Proceedings of LAK19 (Shibani, Knight, & Buckingham Shum, 2019).

their own set of benefits and limitations. Instructor feedback is most commonly used in traditional classrooms, but it is time consuming and requires considerable effort by the instructor, especially in large cohorts. An alternative is the use of peer feedback, where students make judgements about the performance of each other and provide feedback or marking. Formative peer assessments can provide greater immediacy, timeliness, and individualization of feedback via corrective, confirmatory and suggestive feedback (Topping, Smith, Swanson, & Elliot, 2000). Peer feedback and discussion on students' writing also enable students to learn from each other (Allal, Lopez, Lehraus, & Forget, 2005). However, the usefulness of peer feedback depends on the quality of feedback provided and might vary among students leading to inconsistencies in its application. As a result, peer feedback is often implemented alongside other scaffolding that supports its efficacy.

Another strategy to providing feedback is through automated assessment of writing using data analytics and processing techniques. With the growing quantity of data generated from the Internet of Things (IoT) and digitization, and new developments in advanced technologies, there is increasing potential in optimising education to impact learners at scale (Merceron, Blikstein, & Siemens, 2016), and this is enabled by the field of 'Learning Analytics' (LA). A sub-domain of LA called 'Writing Analytics' focuses on supporting learners in their writing, and can be used to develop tools that provide automated formative feedback on students' drafts based on different text features. Such automated writing tools can provide timely and consistent feedback on students' writing (reviewed in Section 2.4). One example is *the Academic Writing Analytics (AWA)* tool, subsequently named *AcaWriter* (referred to simply as AcaWriter for brevity henceforth). This tool provides formative feedback on students' academic writing (Knight, Buckingham Shum, Ryan, Sándor, & Wang, 2018; Knight, Shibani, et al., In submission). It uses natural language processing techniques to identify sentences in a text that match specific rhetorical functions, like emphasizing an important point or summarizing, by using linguistic markers that indicate these rhetorical moves. These kinds of moves are a key component in good academic writing and preliminary evidence reported that they were correlated with writing quality (Simsek et al., 2015), an evidence base to which this thesis further contributes. Feedback from the tool on the presence of these moves can help students reflect

on and improve the rhetorical structure of their writing, and forms the basis of my thesis.

Regardless of the quality of technology, a concern for learning analytics tools like AcaWriter is that they may not be used effectively unless they are embedded in the curriculum (Wingate, 2012). These technologies have to be embedded in the curriculum by implementing them in well-designed contexts for better uptake by students, and for solving existing pedagogical issues. This move from developing learning analytics technologies to integrating them as part of a larger educational context can be accomplished using 'pedagogical learning analytics intervention design' which refers to "systematic efforts to incorporate the use of analytics as a productive part of teaching and learning practices in a given educational context" (Wise, 2014). A pedagogic approach to embedding such automated tools in authentic subject designs has been emphasized as a means to align learning analytics with learning design, and can provide a contextual framework to document the pedagogic intent of analytics applications and to collect data for its evidence (Bakharia et al., 2016; Lockyer, Heathcote, & Dawson, 2013). This requires the inclusion of educators in the development of tools and pedagogic interventions, by understanding their perspectives and co-designing with them, to implement learning analytics innovations in their classrooms for long term adoption. These design approaches add value to learning by closing the gap between the potential, and actual use, of technologies.

Furthermore, an inherent problem in the application of such technologies is that educational systems are contextual, meaning different factors like educators and their instruction, the classroom environment, the specific subject matter, and task design, all influence learning in the particular pedagogical setting. Thus, is unclear if generalized solutions can cater to all learning contexts in the same way. For instance, a predictive model trained on data from one discipline may not be generalizable to another discipline if there are contextual factors that affect the model performance. The move from *big* data to *meaningful* data is hence important for learning analytics researchers so that LA does not promote a one-size-fits-all approach (Gašević, Dawson, & Siemens, 2015; Merceron et al., 2016). While general productivity and management tools like writing editors (MSWord, Google Docs etc.), and learning management systems (Blackboard, Canvas etc.) work well in many cases, the context is particularly

important for student-facing tools that provide feedback on higher-order constructs like writing. A recognized challenge in the field of learning analytics is the uncertainty around "LA's pedagogical relevance and value-add in contextualized learning and teaching settings across different disciplinary domains" (Tan & Koh, 2017). While theories connecting LA applications with pedagogy are receiving increasing attention, the challenge of developing discipline-specific and contextualized LA systems remains. Most systems are developed to be generalizable and open for wider contexts, and very few learning analytics systems can provide contextualized support (Liu, Bartimote-Aufflick, Pardo, & Bridgeman, 2017).

Scalability and contextualization in LA seem to contradict each other in the sense that they prioritize different objectives. This arises from the tension between quantitative approaches that often identify generalizable regularities in learning, and qualitative approaches that tend to understand the particularities tied to specific contexts (Shaffer, 2017; Wise & Cui, 2018). There is a need to strike a balance between the two by developing scalable learning analytics applications that can cater to large numbers of students, while also catering for specific contexts by considering the nuances that make them distinctive. To achieve such a balance, we require an understanding of practical considerations and implications for pedagogical innovations that impact learning (student writing in this case) in authentic practice. Hence, the overall aim of my research is *"to investigate the implementation, and impact on student writing, of an automated feedback tool in higher education teaching practice"*. Based on this aim and the gaps in literature the research addresses, three main research questions that guide the study are as follows, each refined into two sub-questions:

<u>RQ1 - Writing Products:</u> What is the impact of automated feedback on rhetorical moves in student writing?

RQ1a. What are students' perceptions of the writing task with/ without automated feedback?

RQ1b. What is the impact of automated feedback on student revisions?

RQ2 - Writing Processes: How do students engage with automated writing feedback?

RQ2a. How can we study students' interactions with automated feedback?

RQ2b. How does scaffolding (using peer feedback and additional instruction) impact student engagement with automated feedback?

RQ3 - Educator Perspectives: What are practitioner perspectives on automated writing feedback in authentic practice?

RQ3a. What factors influence adoption in authentic classrooms?

RQ3b. How do the practitioners engage in implementation across disciplines, and what are the outcomes?

These research questions are motivated in detail by the gaps identified in the existing literature, in Section 3.1 Research Questions.

1.2 Research Contributions and Significance

The main contribution of this thesis is the design, implementation, and evaluation of a learning analytics tool (AcaWriter) and the development of related social and technical infrastructure, for students to improve their writing in authentic higher educational contexts. The approach has led to significant improvements in student writing with the use of automated feedback from AcaWriter and the pedagogical design embedding its usage. This impact and implementation has been empirically validated across two disciplines and evaluated using various metrics including student perceptions, actual improvements made in the writing in terms of the revisions made and the processes involved, and instructor perspectives. By understanding how students engage with automated feedback and the effects it has on their writing in authentic contexts, we can effectively improve educational practice.

By developing new ways to improve writing and exploring the processes involved in doing so, the thesis has contributed to the theoretical and practical understanding of wider learning analytics applications to support their adoption in authentic practice. It introduces a conceptual model in learning analytics called the 'Contextualizable Learning Analytics Design (CLAD)' model by formalising the key concepts that emerged from the design-based research cycles. The model

emphasizes the importance of connecting learning analytics and learning design to balance generalizable scalability and contextualized support, clarifies the nature of that interdependency in terms of key factors (assessment, features, task design and feedback), and exemplifies CLAD in the context of the AcaWriter tool in different writing contexts. In taking this contextualized approach, the research brings together stakeholders using *co-design* by understanding multiple perspectives and designing suitably with those considerations in account. It exemplifies a multidisciplinary approach to learning analytics by bringing together fields like instructional design and learning analytics with core disciplines like Law and Accounting. The study highlights the practical considerations and consequences in conducting research in authentic classrooms using AcaWriter, which has been evaluated across disciplines in authentic higher educational settings in a design-based research approach over multiple iterations. It thus emphasizes authenticity in LA and demonstrates a design-based research methodology for long term task-oriented design of LA applications.

Furthermore, the thesis contributes to the building of socio-technical infrastructure that can enable the development of enterprise-level tools that go beyond research prototypes. In addition to developing new functionality that enabled AcaWriter to provide contextualized automated feedback on writing, the thesis has also facilitated the development of other technical elements, described below:

- An online platform written in PHP called AWA-Tutor (detailed in Section 4.2.2), that structures the student experience in a writing assignment, and facilitates data collection and the use of automated feedback from AWA/ AcaWriter for students embedded in a robust learning design.
- A new method of studying students' revision patterns in writing and their engagement with automated feedback, called 'Revision Graphs', implemented using text analytics and network graphs in Python (Section 4.4.5.4).
- A suite of R programs that analyse student data from AWA-Tutor to look for insights using statistical inferences and visual plots.

Finally, the thesis has contributed to the building of social infrastructure around the use of AcaWriter for academics internal and external to the university. Open resources from research and practice are disseminated to educators and

general public in a shared repository, and methods are devised to adopt the approach to new writing contexts. By sharing the approaches taken to successfully implement such learning analytics innovations across contexts, a community of users can build upon each other's work. This includes practices employed and resources used like design principles, task designs with their rationale, guides and video resources to help students and educators, and questionnaires for evaluating research. A consolidated list of such materials is available in the List of Resources.

1.3 Thesis Organisation

The current chapter has introduced the context and goals of this research, which sits in the "middle space" (Knight, Buckingham Shum, & Littleton, 2014; Suthers & Verbert, 2013) between pedagogic writing support and learning analytics. It explained the importance of developing new ways to improve writing, and provided the big picture of how the thesis contributes to existing research and practice. The remainder of the thesis is outlined as follows:

Chapter 2: Literature Review presents the existing body of knowledge and locates the thesis within the literature on writing, learning analytics, and theory behind technology integrated educational practice. It describes research on the types of support for student writing, and how automated writing feedback can provide efficient technology enhanced support.

Chapter 3: Methodology describes the research method, study context, data collection, data analysis, and procedures involved in the study at an overall level. It explains how the Design Based Research (DBR) methodology chosen for the study best suits the current research by making use of iterative designs to support writing in authentic practice. It also outlines the ethical considerations of the study and provides a timeline of the iterations (stages) involved in the study.

Chapter 4: Design Iterations in Learning Context 1 details the implementation of design iterations of the study in the first learning context: an undergraduate Civil Law subject. It includes empirical implementations with descriptions on the tools developed, analysis, and evaluations that led to changes in each design iteration from the previous iteration. It also examines the research questions by applying mixed methods to analyse the empirical data.

Chapter 5: Conceptual Model and Design Transfer to Learning Context 2 describes the Contextualizable Learning Analytics Design (CLAD) model that emerged from the implementations discussed earlier, and applies it to a second learning context: an undergraduate Accounting subject. It exemplifies the transfer of formalised design principles to this new learning context through implementation and evaluation. It also addresses research questions using the empirical data collected.

Chapter 6: Practitioner Perspectives explores the factors influencing practitioners in implementing such learning analytics applications in their classrooms and their perspectives on the outcomes gained. It uses a qualitative research methodology to interview instructors, and groups the themes that emerged from their responses inductively.

Chapter 7: Discussion, Future work, and Conclusion provides a summary of findings for all the research questions by consolidating insights from the different design iterations and interviews. It summarises the key contributions that this thesis makes, recognises the limitations, and outlines implications for future research.

Chapter 2: Literature Review

This review begins with a focus on writing as a key skill for students and describes how writing is taught in different instructional settings[3]. The chapter then provides a brief overview of the field of learning analytics and its affordances to optimise learning environments, particularly around writing in its sub-domain called writing analytics. Automated tools developed to improve writing are then discussed in detail to provide an insight into prior work done in this research area. The writing tool used in the current research is then explained with rationale and background behind its development. Finally, evaluation methods utilised to study the effectiveness of such tools are discussed. The chapter concludes with a summary of the literature, particularly highlighting gaps in this background and implications for the current research.

2.1 Writing

Writing is a key skill that contributes to essential learning outcomes in higher education students. It is found to be a high-impact educational practice that correlates to information literacy and student engagement (Kuh, 2008). By writing effectively, students develop their academic language proficiency and master their subject matter from the curriculum by elaborating their mental representations of knowledge. They not only learn from writing by gathering, preserving and transmitting information, but are also able to influence others using persuasion, thus proving writing to be a powerful tool for accomplishing a number of goals (Graham, Gillespie, & McKeown, 2013). Writing is hence seen as a key academic literacy which is socially situated for student learning (Lea, 2004; Lillis, 2003). Here, a key part of that socially situated nature is genre, and rhetorical function or metadiscourse help convey the intent of the writing to the reader (discussed in detail in Section 2.4.3). Writing is also a professional skill for students to develop in order to communicate well in their workplace (OECD, 2000).

[3] Sections of this chapter draw from published work: Section 2.5 & Section 2.4.1.2 - International Conference Proceedings of AIED (Shibani, Knight, & Buckingham Shum, 2018a), Section 2.4.4.2 - Journal of Writing Research (Knight, Shibani, et al., In submission), and Section 2.7.4 - The Internet and Higher Education (Shibani, Knight, & Buckingham Shum, In Submission).

Despite its importance, students are seen to lack proficiency in writing. Strong writing skills are hard to develop as it requires involvement in cognitive processes like planning, translating, and reviewing by interacting with the task environment and the writer's long-term memory (Flower & Hayes, 1981). Even though students have good basic skills of writing to convey their thoughts in terms of form, content and language, they are particularly weak in advanced skills to create coherent and engaging writing by making argumentative claims (National Commission on Writing, 2003). When transitioning from school to university, students find essay writing to be the most challenging academic demand, often finding it difficult to learn (Krause, 2001). Teaching academic writing is a challenge in higher education as students must be inducted into both the prerequisite skills, and professional for good writing (Lillis & Turner, 2001). Moreover, teachers find it challenging to provide writing support as a result of the diversity of students in their classrooms, in terms of demographics and students' preparedness to engage with the curriculum, pedagogy and the conventions of the courses in which they enrol (Horstmanshof & Brownie, 2013). Provision of student support for academic writing has been limited, mostly taking the form of English for Academic Purposes (EAP) for non-native speakers, or remedial action to improve writing skills in an ad-hoc manner (Wingate, 2012). There is thus a need to help students in learning to write in an ongoing and integrated way, to support the ongoing development of their writing.

2.2 Writing Support

Explicitly teaching learners how to write using strategy instruction, providing opportunities to practise writing, and supporting their writing with personalized feedback are shown to be effective approaches to support writing (Graham & Perin, 2007). To provide continued support for improving writing, students should also be provided with formative feedback on their writing. A formative practice provides evidence about student achievement to teachers, learners or peers that helps them make better decisions for the next steps (Black & Wiliam, 2009). Formative assessment and feedback on the current state of students aids them to gain awareness on where they stand in terms of their goals and how to improve their progress. This kind of ongoing feedback enables adaptation of learning based on evidence from previous episodes (Bloom, 1971). This formative feedback is different from the summative assessment approach, which mostly focuses on evaluating and recording the students' achievement

at the end of their learning for the purpose of grading (Testing, 1987). Thus, formative and summative assessments are mainly distinguished by their intended purpose and timing (Harlen & James, 1997). Formative assessment and feedback complete the feedback loop by leading to closure of gap between the expected standard and actual performance, and hence have been found to have more impact on students' learning than summative assessments (Sadler, 1989). The importance of formative assessment in student learning has been re-iterated in the higher education context by calling for reconstruction of existing curricula and formalizing the processes involved in formative assessments (Yorke, 2003). Recent work that explored student and staff perspectives on effective feedback also demonstrated the significance of timely and iterative feedback for improvements to be made (P. Dawson et al., 2019).

Formative feedback should be in a format that can be easily understood by students to help them improve their future learning processes. To make effective use of this feedback, student self-regulation is important as effective learning involves not only transmission of knowledge from teachers to students, but also students developing skills by themselves including setting up goals, managing resources and regulating their motivations and thinking (Nicol & Macfarlane-Dick, 2006). Such self-regulated learning can be supported by providing formative feedback. There are different ways in which formative feedback can be provided to help students in their writing. These kinds of formative feedback can be distinguished based on the feedback providers and the feedback content. The most commonly used feedback types in the learning context of writing are instructor feedback, self-assessments, peer assessments, exemplars and automated tools, discussed next.

2.2.1 Instructor feedback

Feedback provided by the instructor/ teacher is most common in first language (L1) and second language (L2) contexts, where teacher evaluates the work of students and provides them feedback for improvements. A substantial body of literature has discussed the types of feedback provided those are effective for improving writing (Bitchener, Young, & Cameron, 2005; Muncie, 2000; Truscott, 1996). One common approach which is performed in one form or another is 'error correction feedback', provided on grammar and spelling. However, this approach is seen to be an ineffective practice for helping students improve their accuracy of writing in a holistic way (Truscott, 1996). While grammar error correction might improve quality of the current

writing to some extent, it has been argued that it added little value to future accuracy in writing in a number of studies and distracts students from more productive aspects in their writing. Importantly, feedback that can be internalized by students by acting on it fosters more development in the long run than corrective feedback that does not involve them in critical engagement (Muncie, 2000). Also, relying entirely on instructor feedback can disadvantage students' intellectual development because students may feel compelled to incorporate their suggestions in subsequent revisions even if they disagree or do not understand them because of the instructor's power to mark assignments (Curry & Hewings, 2003). Dialogue and conference accompanied by feedback are seen to facilitate better understanding as the process includes clarifications and instructions for the given feedback (Bitchener et al., 2005; F. Hyland, 1998).

2.2.2 Self-Assessments

To transition from feedback to self-monitoring, it is necessary to make provisions for students to acquire evaluative expertise (Sadler, 1989). This could be done by allowing students to explicitly engage in the process of assessing their own writing. Self-assessments enhance students' capacity to make judgements and self-regulate their work for sustainable learning. Students' assessment of their own writing will help them identify problems in their writing and think of ways to improve them. They produce higher learning outcomes and increased responsibility of own learning over time through reflection and problem-solving (Dochy, Segers, & Sluijsmans, 1999). Such sustainable assessments aid for convergence between self-assessed grades and subject-expert grades by improving their evaluative expertise for future learning (Boud, Lawson, & Thompson, 2015). Developing evaluative judgement to be able to assess the quality of work has hence been recognised as a core skill needed for students (Boud, Ajjawi, Dawson, & Tai, 2018).

To support that, the need for explicit provision of involving with self-assessment activities has been emphasized (Sadler, 1989). Further, it has been observed that feedback plays a big role in developing evaluative expertise, as students have to know if their judgements are correct by calibrating them against known standards. For such feedback to be effective, it has to clearly mention the desired goal standard, the gap between the actual and expected standard and how to alter this gap (Ramaprasad, 1983; Sadler, 1989). Students also improved their accuracy of self-assessments over time, which emphasizes the need for continual practice. Further improved accuracy of

students' self-assessments were noted when teachers provided feedback on their assessments, which resonates with the argument of providing feedback on self-assessments (Dochy et al., 1999).

2.2.3 Peer feedback

Assessment as a tool for learning has a great impact on students' learning and development into reflective practitioners and can often involve a combination of self and peer assessments (Dochy et al., 1999). Peer assessment and feedback is another type of feedback that has been effective in a number of studies to enhance students' learning and evaluative judgement (Boud, Cohen, & Sampson, 2014). In peer assessments, students make judgements about the work or performance of other students and provide feedback or marking. This enables peer learning where students simultaneously learn through the provision of feedback, and contribute to other students' learning. It should not be seen as a substitute for teaching and activities of staff members, but rather to complement the quality of education. In addition to the learning process, students also learn other skills like working with others, critical enquiry and reflection, communication of knowledge and skills, self and peer assessment (Boud et al., 2014). Formative peer assessment when combined with more levels of feedback like corrective, confirmatory and suggestive feedback provides greater immediacy, timeliness, and individualization of feedback (Topping et al., 2000). It is well suited for large classes, where it is not practically possible for the instructor to provide formative feedback to all students since the process is time-consuming. Peer assessment makes the marking process less expensive compared to marking by expert graders in a large scale. It also minimizes the time taken to give feedback for more meaningful learning. However, for peer assessments to be effective, it should have clear guidelines to train students for marking and giving useful feedback to students (Williams, 1992). This is because not all students know how to give useful feedback to peers for improving their work.

Past studies comparing the effects of teacher and peer feedback have observed that student perceptions of the value of feedback are higher for teacher feedback which is paid more attention to than peer feedback (Connor & Asenavage, 1994; Paulus, 1999). This effect was seen because students consider teachers as experts and hence make more revisions in drafts based on their feedback. Teacher feedback was also seen to have a higher uptake compared to peer feedback by students (Ruegg, 2015). However, it was observed that frequent teacher feedback led to misunderstandings or unsuccessful

revisions. In contrast, frequent peer feedback led to successful revisions (Ruegg, 2015). With respect to the types of feedback, specific feedback and general comments had higher uptake by students. The impact of peer feedback particularly on students' writing in higher education has been examined in a recent meta-analysis in more detail (Huisman, Saab, van den Broek, & van Driel, 2019). It found that peer feedback resulted in greater writing improvements compared to controls with no feedback or self-assessment. Furthermore, it found similar improvements in writing from peer feedback and teacher feedback, making them comparable. Thus, the benefits of peer feedback are well recorded in the literature, with emphasis on providing guidance to students to offer good feedback to peers.

2.2.4 Exemplars and Benchmarking

Another method to expose students to, and encourage them to evaluate, the work of others is through the provision of exemplars. Exemplars are concrete examples of written works that can be provided to support students in understanding high and low quality texts. Engagement with exemplars is seen as an effective method for students to understand the assessment criteria. It helps them identify and reflect on features of good and bad writing that can be applied to their own writing. Exemplars make the assessment process transparent and the criteria for assessment clear to all participants (Hendry, Armstrong, & Bromberger, 2012; Rust, Price, & O'donovan, 2003). They are often found to be more effective when accompanied by discussion and socialization in order for explicit and tacit knowledge transfer to occur (Rust et al., 2003). One criticism is that they inhibit student creativity by encouraging students to mimic examples. However, countercriticisms argue that standards are often expressed using multiple exemplars, which facilitate adequate understanding on the expected quality. Also, knowing the criterion and characteristics of a discipline aids creativity and not vice-versa (Sadler, 1989). A recent study has highlighted the benefits of calibration tasks like 'benchmarking', where students assess exemplar texts of varying quality and develop their evaluative judgement (Knight, Leigh, Davila, Martin, & Krix, 2019). Such tasks aid in deeper discussion on the features in exemplar texts that characterise quality, for students to build their ability to assess and provide feedback on other texts from their own writing or peers.

2.3 Learning Analytics

To provide support for writing and other forms of learning, education is increasingly making use of the affordances of technology and data. With the growing quantity of data generated from the Internet of Things (IoT) and digitization, many fields are interested in the application of analytics and big data. The emergence of new technologies and artificial intelligence have enabled novel forms of feedback that never existed before, and more fine-grained, time-efficient analysis. In education, high profile initiatives like massive open online courses (MOOCS), online video-based learning, educational apps, and personal and portable computing devices, have provided access to increased quantities of learner data (Merceron et al., 2016). Thus, analytics and big data in education holds the potential to develop scalable methods that can be employed widely across large numbers of students and institutions; to obtain big impact from big data (Merceron et al., 2016). This potential to impact teaching and learning practices led to the emergence of 'Learning Analytics'.

As defined in the 1st International Conference on Learning Analytics and Knowledge (LAK), Learning Analytics is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and environments in which it occurs". Since the emergence of Learning Analytics (LA) as a field, research and practice in LA have taken many forms to cater to specific interests. Several fields within educational research have contributed to the development of learning analytics including: Citation analysis, Social network analysis, User modeling, Education/ Cognitive modeling, Tutors, Knowledge discovery in databases, Adaptive hypermedia, and E-learning (Siemens, 2013), which bring forth a diverse spectrum of LA and related Educational Data Mining applications.

In higher education, Learning Analytics is envisioned as a method to help allocate resources, develop competitive advantages, and improve the quality and value of the learning experience (Siemens & Long, 2011). While data was previously available in educational research, 'learning analytics' holds the promise of providing insight for learning research through the longitudinal collection of data on various levels of granularity from multiple sources in authentic learning environments, and potential impact at scale (Reimann, 2016). Some applications include student behaviour modeling, prediction of performance, increasing reflection and awareness, prediction of drop-out and retention services, improving feedback and assessment services, and

recommendation of resources (Papamitsiou & Economides, 2014). As Learning Analytics becomes more mainstream and widely applied in practice, a particular interest for higher education is to focus on pedagogically salient applications of LA that can be impactful in supporting users in their learning. This includes the provision of personalized support for learners using tools that provide feedback on learning aspects enabled by LA, and will be the focus of my work on writing using 'Writing Analytics'.

2.4 Writing Analytics

Writing Analytics (WA) can be thought of as a sub-domain of learning analytics that uses analytic techniques to develop a deeper understanding of writing in educational contexts. Writing Analytics is defined as "*the measurement and analysis of written texts for the purpose of understanding writing processes and products, in their educational contexts*" (Buckingham Shum et al., 2016). It uses computational and Natural Language Processing (NLP) techniques to study the written products and processes involved in the creation of the product, and to develop tools providing automated feedback or grading on writing. It can thus be essentially thought of as comprising of two types of applications:

1. For automated analysis of writing products and processes
2. For supporting writing through the provision of automated feedback

Four workshops run on this topic in LAK have focused on critical perspectives and community building around writing analytics (Buckingham Shum et al., 2016), developing a writing analytics literacy and practitioner capacity (Knight, Allen, Gibson, McNamara, & Buckingham Shum, 2017), a hands-on training for developing this literacy by understanding technical affordances and aligning them to pedagogical feedback (Shibani, Abel, Gibson, & Knight, 2018), and mapping the state of the field and future directions (Shibani, Liu, Rapp, & Knight, 2019). The two main applications of Writing Analytics: automated analysis, and automated tools are discussed next.

2.4.1 Automated analysis of writing

This application of writing analytics to analyse writing involves studying text features and processes in writing using automated means, explained below.

2.4.1.1 *Studying Text Features in Writing*

Although Writing Analytics is a relatively new term, research studying writing by using computational techniques has a long history. For example, the Coh-Metrix tool analyzes texts automatically and provides 200+ measures to evaluate texts based on cohesion, language and readability (Graesser, McNamara, Louwerse, & Cai, 2004). They include text-based linguistic features related to lexical sophistication (word frequency, word concreteness, word familiarity, polysemy, hypernymy), syntactic complexity (incidence of infinitives, phrase length, number of words before the main verb), and cohesion (word overlap, semantic similarity, incidence of connectives). Another tool WAT, captures additional measures of global cohesion and lexical and syntactic complexity in texts.

Using such tools, a number of studies have identified linguistic indices of text that correlate to high- and low-quality text ratings graded by experts (S. Crossley, Allen, Snow, & McNamara, 2015; Vajjala, 2018). Linguistic features of texts have also been used to explore revision characteristics and patterns in student essays, which have implications on the type of automated feedback that best suit learners (Roscoe, Snow, Allen, & McNamara, 2015). Furthermore, the multi-dimensional nature of text with co-occuring linguistic features (S. A. Crossley, Allen, & McNamara, 2014), and flexibility in the use of narrative and cohesive linguistic features over time and how its relationship to literacy abilities were explored (Erica Linn Snow et al., 2016). These studies analyzing and assessing the characteristics of writing contribute to the theory in writing, but they need to be transferred to other forms for use in pedagogical applications. While the tools discussed above are used to study textual features by researchers and others with linguistic knowledge, a recently developed tool called 'Quantext' aims to support *teachers* in analyzing student texts using text analytics (McDonald & Moskal, 2017).

2.4.1.2 *Studying Processes in Writing*

Automated analysis of the process of writing predominantly involves developing visualizations to represent the process of drafting and revision. Recent tools that log keystrokes during document creation, and save interim drafts during the revision process now enable analysis of the revision process at a fine-grained level (Van Waes, Leijten, Wengelin, & Lindgren, 2012). To visualize modification patterns in an online document, Caporossi and Leblay (Caporossi & Leblay, 2011) developed a graph theory

approach to represent the movement of text through a document. In that work, log data of keystrokes and cursor movements from the document editing process was used to generate a graph. In other work that investigated collaborative writing processes, a revision map was created to represent the joint development of ideas by a group of authors (Southavilay, Yacef, Reimann, & Calvo, 2013). These maps helped researchers gain insights into the location and time of text edits made, and the collaboration of students to develop the document. Another work introduced the use of Sequence Homology Analysis (SHA) to study the evolution of public speech drafts by comparing the changes in characters (Wininger, 2014). SHA is a method from molecular biology to study differences in DNA strands. It was applied to identify the character differences between text drafts, and a draft network was proposed. This network maps sentences in students' drafts based on the strength of revisions made across the drafts.

Studying the revision process leads to a deeper understanding of writing by contributing to the theory and research on writing, which can then lead to its application in writing contexts. In addition to studying user behaviour and interaction through log data, this can inform design choices in writing tool development. While computational analysis techniques which are previously available can be used to study learning processes, new ways of looking at data help uncover new patterns of the current pedagogical contexts and learning scenarios; one such approach will be discussed in Section 4.4.6. However, to develop this process data to provide meaningful feedback that can improve writing, insightful revision patterns need to be found, which is an emerging area of research.

2.4.2 Automated grading, evaluation and feedback tools

Using technological advances, automated and semi-automated tools have been developed to overcome these issues of time, effort and reliability in assessing writing (Shermis, Raymat, & Barrera, 2003). They also provide more immediate feedback as compared to the other feedback types that require waiting for the grader to complete the evaluation. Several automated tools have been developed in order to assess student writing automatically. The scope of tools that evaluate essays varies from Automatic Essay Scoring (AES) systems that provide a score based on the assessment of standardized writing to Automated Writing Evaluation (AWE) systems that provide additional feedback to students on their writing (Warschauer & Grimes, 2008). Thus, the purpose of the tools varies from assessment to feedback for improvement. This

section will briefly talk about such automated tools, criticisms and advancements made in this field.

2.4.2.1 *Scoring systems*

Earlier tools were predominantly developed for the purpose of automated grading based on the textual features in essays. Automated Essay Scoring (AES) tools like Criterion and My access provided feedback on rhetorical and formal aspects of writing on the essays submitted to their scoring engines and also contained tools like dictionary, thesaurus etc. Criterion was developed by Educational Testing Service (ETS) as a web-based essay assessment tool to provide scores and feedback to students on grammar, usage, mechanics, style and essay discourse elements (Burstein, Chodorow, & Leacock, 2003). A study on the usage of Criterion for revising essays reported improvements in students' essay scores, error rates and introduction of discourse elements in subsequent versions of improved essays compared to the initial essay (Attali, 2004).

MyAccess, a commercial tool developed by Vantage Learning scored essays using IntelliMetric automated essay scoring system. The scores were generated based on five features: Focus and coherence, Organization, Development and elaboration, Sentence structure and Mechanics and conventions (Learning, 2003). It provided multi-lingual feedback on students' writing containing tools like dictionary, thesaurus, and translator and also provides dashboard and customization features for teachers. A study investigating the performance of the Intellimetric essay scoring system on Graduate Management Admission Test essays found good agreement between automatic and human scoring (Rudner, Garcia, & Welch, 2006). Such AES systems used linguistic features, machine learning models trained with a large number of graded texts to predict the scores of student essays in standardized writing tests, and/or used benchmarked essays for a topic which were then used to compare and grade student essays using Latent Semantic Analysis with high reliability (Rudner et al., 2006; Shermis et al., 2003). These tools focused mainly on the reliability in scoring compared to human grading and have been found to perform well for automated assessments.

Although the automated essay evaluation systems have been credited for their performance in some studies, they are also criticized in other studies for using shallow features, predetermined comments, and ignoring content meaning and argumentation (C.-F. E. Chen & Cheng, 2008; Ericsson & Haswell, 2006). Automated essay evaluations do not consider the social aspect of writing, which put forth arguments that they induce

training students to write for machines and not for humans. The efficiency of such systems was questioned since writing includes more meaningful engagement than merely formulaic features of text. Further, the errors flagged to students might direct students to place a lot of emphasis of errors, which may not be very serious threats to writing skills (Cheville, 2004). It could also be possible to game the system in order to get high marks by writing longer essays and plagiarising, since the systems cannot detect such features which can be easily detected by human graders (Kukich, 2000).

However, advocates of AES systems suggest that these concerns are not well founded and explain their right usage (Attali, 2013; Deane, 2013). They advise that the purpose of using automated tools should be carefully thought in order to put them to appropriate use. Since a machine cannot understand the meaning in writing like human beings, the scores of automated tools should not be interpreted similar to a human graded score. It should be noted that AES systems are not intended to replace human graders but rather complement them (Attali, 2013). The constructs measured by such systems should also be clarified in order to address the meaning, aggregation and reliability of features used for measurement with relation to human scoring. It has been reiterated by Deane that the deployment of AES should be appropriately chosen for some aspects of writing that can be assessed well by utilising the strengths of automated analysis (Deane, 2013). Instead of unquestionably rejecting the usage of such systems, it is emphasized to make best use of them for lessening cognitive load and increasing writing practice for students. For weaker features of such systems like argumentation, human beings could complement them. Thus, a combination of machine and human markers would be ideal, rather than using AES for sole assessments. The purpose of the tools could also be revised to provide feedback to students for improving the essays, rather than just scores for grading.

While automatic classification of writing is useful to develop large scale solutions such as automatic graders, the objective is to assign a grade that is sufficiently close to a human grader. Such approaches do not always provide data that can serve either as formative feedback to students, or assist human sense making in research contexts. Further work is required to deliver meaningful formative feedback to students that can aid improvement in their writing skills.

2.4.2.2 *Automated feedback tools*

As discussed earlier, writing involves multiple processes that can be supported by feedback for improvements. This is emphasized in the writing process approach, which defines writing as the iterative process, where seeking and receiving feedback while a text is being produced enables better quality texts (Curry & Hewings, 2003). Writing feedback helps students understand how they write and reflect on their own writing process, and also makes them aware of writing appropriately for the audience. While instructor and peer feedback is effective, they are also very time-consuming. They require considerable effort from the feedback provider and thus increase the workload of students and instructors. The feedback provided could also be subjective to the feedback provider and hence inconsistent across students.

To overcome this burden, Automated Writing Evaluation (AWE) systems include tools that use computational techniques to assess writing and provide immediate and consistent feedback for all students. In contrast to AES systems that only provide the scores for their essays, students also receive formative feedback to improve their writing. This helps students to practice writing more drafts and enhance their writing skills. Several such tools have been employed for university and school students to analyse text in the context of essays, problem solving, free form and collaborative writing. The tools use different forms of visualizations and reports for feedback as explained in this section.

An example tool is Glosser, which was developed as a web based tool to provide non-genre-specific feedback - called a 'gloss' – to students on their written essays. Glosser used text mining techniques to extract features related to issues in writing and highlight them to students as feedback and did not require pre-scored essays for training. The provided feedback was based on five rubric elements: Use of source material, structure and development of answer, control of writing style, grammatical correctness and quality of presentation (Villalón, Kearney, Calvo, & Reimann, 2008). An extended collaborative writing support environment called 'iWrite' was built using the Glosser feedback component and other cloud-computing API components (Calvo, O'Rourke, Jones, Yacef, & Reimann, 2011). This included modules for assignment management, peer assessments, collaborative writing and process mining of collaborative edits. The process mining component tracked the revisions made by students in a collaborative google document and visualized the collaborative process of

writing in three ways: Revision maps that show the development of the document using the edits made over a period of time, Topic evolution charts that visualize the development of topics in the document over time and Topic-based collaboration network that creates a network of students based on common topics written by students (Southavilay et al., 2013). In another study analysing collaboratively written texts, the evolving text in Wiki was visualized as a network of concepts in a compact way where the nodes represent concepts (noun phrases) and the edges represent relations between the concepts and colouring indicates authorship (Hecking & Hoppe, 2015). Network centrality measures were used to quantitatively characterize authors from the concept network and qualitative analysis based on the size of the network, density was used to observe the connectedness of concepts in the text.

To further provide adaptive and interactive support for learning, Intelligent Tutoring Systems (ITS) were developed to engage students in an open dialogue with an automated tutor by modelling user behaviour and competencies (Zukerman & Litman, 2001). An intelligent tutor called Writing-Pal (W-Pal) was designed to teach writing skills to high school students providing strategy instruction, modularity, extended practice, and formative feedback using game-based and essay-writing practice (Roscoe & McNamara, 2013). W-Pal helped school students to practice timed persuasive essays using SAT[4] prompts and automatically receive scores and feedback on their essays using linguistic text features (Erica L Snow, Allen, Jacovina, Perret, & McNamara, 2015). However, this feedback is based on the indices of the text and not the content itself. WRITEEVAL was another tool used to assess school students' textual responses to short answer questions in Science as correct, partially correct or incorrect using a text similarity technique (soft cardinality) and a semantic analysis technique (precedent feature collection) (Leeman-Munk, Wiebe, & Lester, 2014). It performed well and the assessments correlated with summative analyses of student performance. Nevertheless, it was considered that a detailed feedback would be more constructive than just the assessment of correctness to support the argumentation in Science. A range of computer-based writing instruction tools now exists to support students in their writing learning, see review in Allen, Jacovina, and McNamara (2015).

[4] SAT (Scholastic Aptitude Test) is a standardized test widely used for college admissions in the United States

More recently, a tool called “EssayCritic” has been developed to provide students with writing feedback in English as a foreign language (EFL) context. It performs semantic analysis by identifying the presence of specific topics in short texts (<500 words). This is done by training the system using a pre-defined knowledge base of themes and concepts related to a particular topic. Feedback provided by the tool is in the form of sub-themes identified and sub-themes suggested (currently missed) from the written essay. A study comparing the effect of automated feedback from the tool (target group) and peer feedback (comparison group) found that students receiving feedback from the tool wrote more sub-themes than the other group, although there was no significant difference in their final grades (Anders I. Mørch, 2017). The comparison group was found to be more focused on essay organization than the essay content compared to the target group. This kind of practice using pre-defined prompts or topics for providing feedback are however defined for specific contexts, which inhibit their applicability to domains other than the ones they’re tested for. For scaling up the usage of such tools to other learning contexts, they need to be recreated for the other domain leading to increased costs.

An approach which deployed some of the same underlying principles of cohesion as Writing Pal to provide feedback to students is OpenEssayist, a web application that made use of NLP techniques to automatically extract summaries from free-text essays, such as key words and key sentences and recognized essay structures for students to reflect on their own draft text. Several summarized text elements were provided to students to explore and reflect on their writing. In a later study using Open Essayist, Rainbow diagrams were used to pilot test visualizations of essays based on the key sentences and connectedness in an essay (Whitelock, Field, Pulman, Richardson, & Van Labeke, 2014). The diagram is a graph where each sentence of an essay is a coloured node and the edges connect the nodes containing the same word. The colour of the nodes varies in rainbow spectrum from violet (first sentence) to red (last sentence). This diagram does not use any domain-specific data, but builds from the input text to provide generic visualizations based on the occurrence of words.

Another tool called Writing Mentor is a free-to-use Google Docs add-on designed to provide feedback to struggling writers to help improve their writing in a self-regulated fashion (Madnani et al., 2018). Writing Mentor uses natural language processing (NLP) methods and resources to generate feedback in terms of features and

sub-constructs like use of sources, claims, and evidence; topic development; coherence; and knowledge of English conventions. This kind of generalized feedback point to students the key features and concepts in their writing for reflection but stay isolated from the pedagogical contexts. The feedback might lead to changes in their writing at surface level but might not engage students in a deeper learning for the subject.

Other writing tools that facilitate and support writing, but without automated feedback also exist. These tools categorised as Interactive Writing Platforms (IWP) provide prompts and scaffold the writing process for students, but do not process their input or give feedback (Strobl et al., 2019). An example is a tool called Thesis Writer which support scaling of academic writing instruction to guide students on the whole thesis writing process, allowing for interaction with instructors and peers (Rapp & Kauf, 2018).

Some tools use machine learning techniques to provide data-driven contextualization by making use of large text corpora with millions of examples. The Lightside application initially developed for summative assessments using machine learning (Shermis, 2015) was later identified as capable of providing formative feedback for revision. It was later commercialised and evaluated as Turnitin Revision Assistant that has a generalized set of features which are mapped to rubric elements of specific prompts to provide feedback on essays written for the prompt (Woods, Adamson, Miel, & Mayfield, 2017). But this approach of data-driven contextualization will not work in teaching and learning contexts where there is only limited data, which is the case in most classroom contexts. Further, it does not utilize the teacher or subject expert's pedagogic knowledge which is valuable to provide context in authentic learning settings.

Writing analytics tools can thus provide timely, formative feedback at scale for learners to engage with their drafts and develop their proficiency in writing. But as argued earlier in the introduction, these tools should be contextualized for and embedded in learning contexts to provide deeper reflection at the subject level.

2.4.3 Text feature – Rhetorical moves

The tools discussed earlier show that a variety of features can be identified from the given text to give different types of feedback. Among the many possible features that can be looked at from a piece of writing, the focus of the current research is on the rhetorical structures/ moves in text. A rhetorical move is "*a discoursal or rhetorical unit*

that performs a coherent communicative function in a written or spoken discourse" (Swales, 2004) (p. 228). This is a particularly important feature in scientific writing as it helps the reader to follow the flow of the text by 'signposting' their movements with the intended purpose. This key feature should be addressed in students' academic writing in order to help them write better. Different models have been developed to identify and categorize standard rhetorical structures from scientific text to understand writing genres.

Swales built a 'Create a Research Space' (C.A.R.S) model for the introductions of research articles by examining the sequence of rhetorical moves in them. It consisted of three main moves and sub-steps under each move. According to this model, the ideal moves in a research article follow a particular sequence. The first move is to 'Establish a territory' by providing a background to the work using previous research. This is followed by 'Establishing a niche' to position the current work in terms of the existing gap, followed by 'Occupying the niche' where the current contribution and research are presented. A sample extract from a research abstract annotated to exemplify Swales moves is shown in Figure 1.

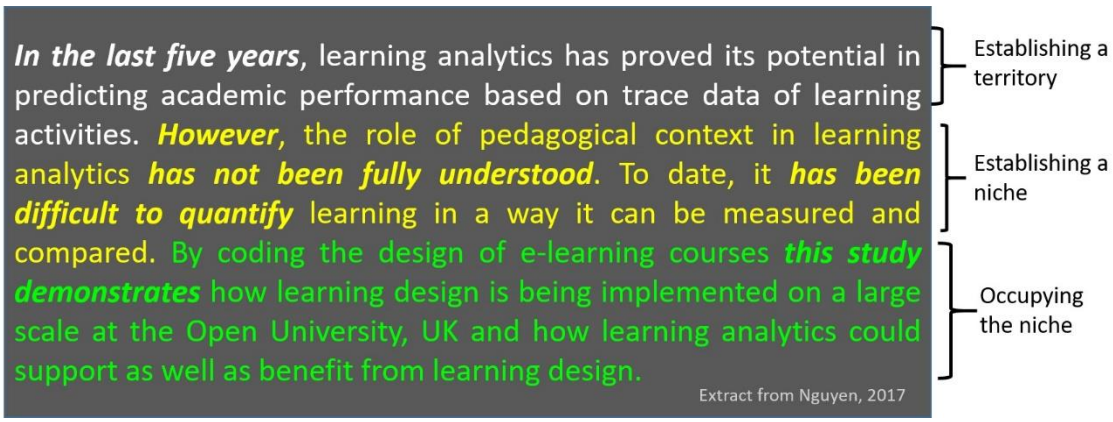


Figure 1: Swales CARS model shown using a sample research writing

In related work, Hyland developed a metadiscourse model consisting of interactive and interactional elements in academic texts which guide the reader with the flow of the text (K. Hyland, 2005). A sample interactive resource to manage the flow is a transition E.g. "in addition", which depicts the relation between the mentioned clauses. An interactional resource to involve a reader is a booster E.g. "it is clear that" which places importance on a proposition. These meta-discourse elements contain linguistic cues that provide orientation to the reader that support in delivering a coherent and persuasive writing.

Building on to these models further, an argumentative zoning approach was developed by Teufel to model the argumentative structure in research articles (Teufel, 2000). Scientific articles have a different structure compared to other texts in terms of linearity (do not follow chronological order) and bias (convincing readers of current work) amongst others. The Argumentative Zoning (AZ) approach was mainly developed to extract key information in research articles based on certain moves in the text. This helps researchers to identify the key information they are looking for in a search via document surrogates, that are not usually retrieved in document search, for example a surrogate that describes the Relation of current work to other work or conclusion. The document structures of scientific texts were analyzed to come up with seven argument zones that are commonly found in scientific research articles such as Aim, Background, Textual, Own, Contrast, Basis and Other (refer Figure 2).

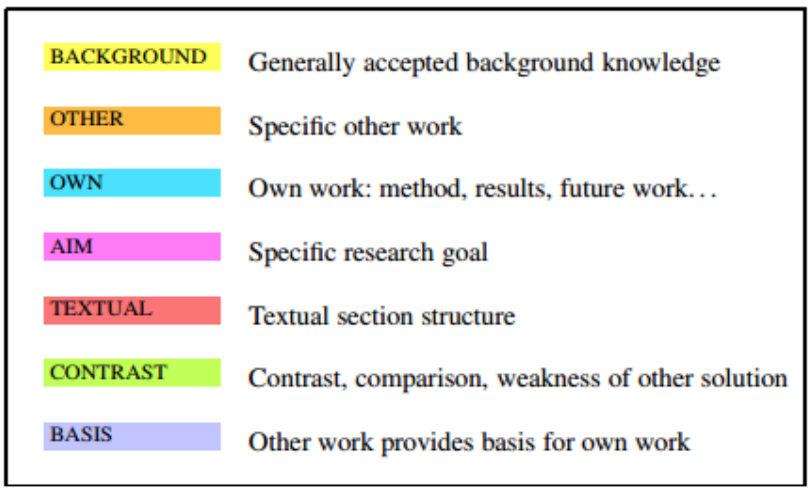


Figure 2. Argumentative Zones identified by Simone Teufel (Teufel, 2000)

These argumentative zones are sections of scientific text that intend to fulfil a specific purpose and contain linguistic cues that can be exploited to label them as specific rhetorical categories. Three labels that relate to the intellectual ownership of statements 'Background', 'Other' and 'Own' form the basic AZ scheme. 'Background' states the field's generally accepted methodologies and problems, and 'Other' refers to more specific descriptions of other work, both of which are not owned by the writer, but guide the reader to accept their own ideas. The author's own knowledge claim is categorized as 'Own'. Non-basic categories include Aim, Textual, Contrast and Basis. Aim refers to the specific research goal an article wants to achieve, Textual refers to explicit information on structure of a section, Contrast indicates critical problems identified from other solutions, and Basis identifies work that is continued from existing

work. The categories are mutually exclusive, so only one of them can be assigned to a particular sentence. These labels are defined to be domain-independent as they are based on rhetorical/ argumentative structures with no domain-specific knowledge.

2.4.3.1 *Tools for automated identification of rhetorical structures in text*

Based on the analysis of linguistic cues and rhetorical moves based on different models, tools that make use of Natural Language Processing (NLP) have been developed, which automate their detection. These automated tools have been developed in different contexts to help researchers and students identify the rhetorical structures in academic writing. The tools discussed in this section are broadly based on either the CARS model by John Swales or the Argumentative Zoning scheme by Simone Teufel. A comparison of the different coding schemes employed by the tools is shown in Table 1. Such automated tools can be employed to provide formative feedback on students' writing to create coherent writing with the use of rhetorical moves.

Mover and SAPIENTA are tools that help researchers identify the structure of research articles. Due to their intended use, they do not focus on providing feedback to students, but rather provide classified texts as output. 'Mover' classifies 3 moves and sub-steps in research articles (RA) from engineering and science field based on a modified CARS model. The tool is downloadable as a desktop application 'AntMover' and aims to provide immediate feedback on move structures in abstracts of scientific articles, thus reducing the time for manual annotations. SAPIENTA is a tool that automatically annotates scientific articles using a finer grained Argumentative Zoning scheme (CoreSC scheme). This tool was used to classify individual sentences in biochemistry and chemistry full papers. The first layer of the CoreSC scheme used by SAPIENTA for annotation consisted of 11 categories (refer Table 1).

The next three tools Research Writing Tutor (RWT), Academic Writing Analytics (AWA) and SciPo are specifically designed to help students in their writing by providing automated feedback on rhetorical structures in their texts. SciPo was developed to guide students to write abstracts and introductions in Portuguese using Argumentative Zoning (Feltrim, Teufel, Nunes, & Aluísio, 2006). It supports students writing using a set of guidelines for good academic writing. It is being excluded from Table 1 since the current work only focusses on tools that provide feedback on English writing.

'Research Writing Tutor' uses genre analysis to identify rhetorical moves and steps and provide feedback on students' academic writing, particularly research articles.

Natural language processing and machine learning techniques are used by this system to classify each sentence in the writing into moves or steps (Cotos & Pendar, 2016). Annotations are based on CARS model developed by John Swales from a variety of discipline-based journal articles. The tool provides feedback in terms of highlighted rhetorical moves from students' writing and the steps missing in those moves according to the CARS model. It also provides sample rhetorical structures as suggestions to help students improve their writing. The tool explored in the current study called Academic Writing Analytics (AWA)/ AcaWriter is explained in detail in the following section.

Table 1. Coding schemes of tools that automatically identify rhetorical structures

Tool	Mover	SAPIENTA	RWT	AcaWriter
Original Scheme	Modified CARS model -three main moves and further steps	finer grained AZ scheme -CoreSC with 11 categories in the first layer	Modified CARS model -3 moves, 17 steps	(Modified from AZ for sentence level parsing)
Coding Scheme Elements	1. Establish a territory *-Claim centrality* *-Generalize topics* *-Review previous research* 2. Establish a niche *-Counter claim* *-Indicate a gap* *-Raise questions* *-Continue a tradition* 3. Occupy the niche *-Outline purpose* *-Announce research* *-Announce findings* *-Evaluate research* *-Indicate RA structure*	-Background -Hypothesis -Motivation -Goal -Object -Method -Model -Experiment -Observation -Result -Conclusion	Move 1. Establishing a territory *1. Claiming centrality* *2. Making topic generalizations* *3. Reviewing previous research* Move 2. Identifying a niche *4. Indicating a gap* *5. Highlighting a problem* *6. Raising general questions* *7. Proposing general hypotheses* *8. Presenting a justification* Move 3. Addressing the niche *9. Introducing present research descriptively* *10. Introducing present research purposefully* *11. Presenting research questions* *12. Presenting research hypotheses* *13. Clarifying definitions* *14. Summarizing methods* *15. Announcing principal outcomes* *16. Stating the value of the present research* *17. Outlining the structure of the paper*	-Summarizing -Background knowledge -Contrasting ideas -Novelty -Significance -Surprise -Open question -Generalizing

2.4.4 Academic Writing Analytics (AWA)/ AcaWriter tool

Academic Writing Analytics (AWA)/ AcaWriter is the writing analytics tool that the current study uses to provide automated feedback on rhetorical moves in student writing. The first version of the tool used in the earlier parts of the study was AWA, and the upgraded version of AWA used in the later parts of the study was called AcaWriter. Since this tool forms the basis of the research in this doctoral study, it is explained in detail as follows.

2.4.4.1 *Academic Writing Analytics (AWA)*

The Academic Writing Analytics (AWA) tool was built as a research prototype that can automatically identify rhetorically salient features in writing, and highlight them as feedback for students to improve their writing. The tool was based on the Xerox Incremental Parser (XIP) that used computational linguistics and natural language processing to perform rhetorical parsing of academic texts. XIP was developed by the Xerox Research Centre, Europe as a robust analyser that tackles deep linguistic phenomena.

In linguistics, parsing refers to the process of analysing text into its constituent logical syntactic components. To ensure that the language analyser provides useful analyses of real-world input text, XIP employed a deep incremental approach with ordered rule layers in cascaded parsers (Aït-Mokhtar, Chanod, & Roux, 2002). The development of empirical evidence that underpinned XIP was incremental, meaning it was not built on a fixed development corpus. XIP used a concept-matching method to detect patterns of particular rhetorical formulations (rhetorical moves) in documents (Simsek, Buckingham Shum, Sandor, De Liddo, & Ferguson, 2013).

This concept matching discourse analysis framework (Sándor, 2007) uses syntactic relationships between words, broadly by extracting a list of constituent concepts from the text, assigning the constituent concepts to a list of words or phrases with a particular function, and mapping these to defined rhetorical functions. To identify the meta-discourse cues that signal a given rhetorical move by the author, XIP specifies a set of dependencies, co-occurrence rules, in which a specific set of concepts must co-occur, in any sequence. A sample rule containing concepts that constitute a 'contrasting idea' is shown in Figure 3.

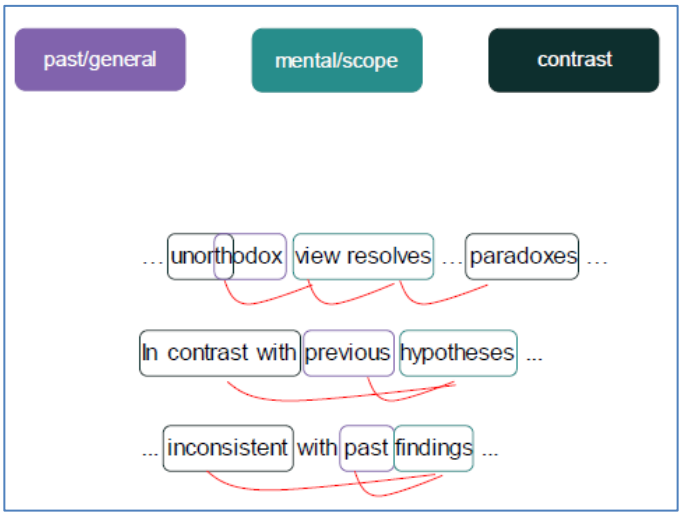


Figure 3: Sample concepts containing metadiscourse contributing to contrasting ideas (Image courtesy: XEROX)

The XIP parser was extended as an Academic Writing Analytics (AWA) tool which provides rapid formative feedback on academic writing of students to make their thinking visible using identified rhetorical moves. Summary, and Important were the main categories, and other rhetorical moves like Background, Contrast, Emphasis, Novelty, Question, Surprise, and Trend contributed to the Important category (see Figure 4 for examples). Such moves are a key component in good academic writing and are seen to be correlated to essay quality (Simsek et al., 2015), with detailed evaluations discussed in an earlier thesis (Bektik, 2017). Feedback on the presence of these moves should help students reflect on their writing and the rhetorical structure of it. Note that XIP also consisted of a different parser for reflective writing to identify the depth of reflection in student writing from linguistic features, which is not covered in this section due to the focus of this thesis on academic writing (More details can be found in Gibson et al. (2017))

Summary: Summarises the author's goals, contribution or conclusions, often referring to another part of the article
Important: Indicates a significant aspect of the problem

Function	Explanation	Example
BACKGROUND	consensus or background knowledge	*They note that administrative units, such as admissions and fund raising, remain the most common users of analytics in higher education today.*
CONTRAST	disagreement, tension, options, inconsistency	*There is also a shift away from an institutional perspective towards a focus on the concerns of learners and teachers.*
EMPHASIS	emphasis on importance, significance	*We propose that the design and implementation of SLA present significant challenges and opportunities...*
NOVELTY	novelty, improvement	*Other new models for learning are emerging from a variety of digital sources.*
QUESTION	a question or missing knowledge	*Current data is insufficient to conclude if...*
SURPRISE	an unexpected outcome	*We have identified in recent years an unusual number of reports that...*
TREND	a trend, growth, pattern or tendency	*Other new models for learning are emerging from a variety of digital sources.*

Figure 4: Rhetorical moves identified by XIP for AWA tool

The identified rhetorical moves were highlighted as feedback for students to reflect on their writing (Sample screenshot in Figure 5).

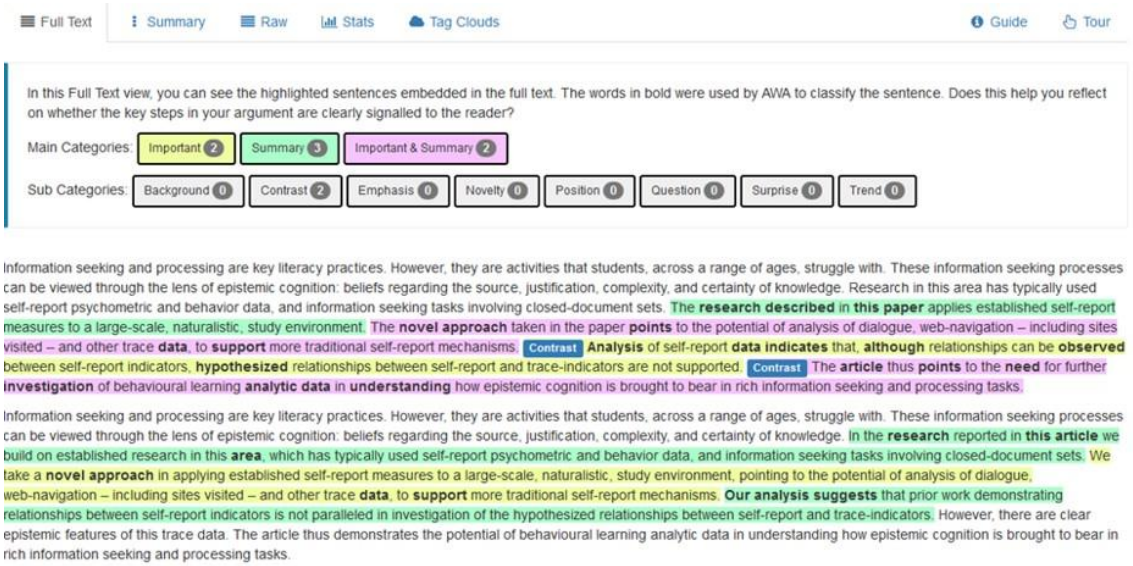


Figure 5: Screenshot of AWA highlighting rhetorical moves on a sample text

In the pedagogical context, AWA has been implemented in undergraduate Law to provide feedback on the academic writing of students. In that context, the identification of rhetorical moves in the form of highlighting were evaluated with feedback from students, and changes were made to the parser as required (Knight, Buckingham Shum, et al., 2018). However, the AWA tool only highlighted the rhetorical moves for student reflection, and did not provide additional feedback with suggestions to students for

improving the text further. The updated version of AWA discussed next contains additional tool features.

2.4.4.2 *AcaWriter*

AcaWriter, the upgraded version of AWA discussed earlier, has improved technical infrastructure and tool features. It is a web-based tool that provides formative feedback on rhetorical moves in writing, released under an open license[5]. AcaWriter uses an updated parser called 'Athanor' to detect rhetorical moves similar to XIP, through a robust text analytics framework called 'Text Analytics Pipeline' (TAP) – more details in Knight, Shibani, et al. (In submission). The key rhetorical moves that AcaWriter can detect in academic texts are shown in Table 2, while Figure 6 shows highlighted moves identified in a sample text in the user interface (UI) of AcaWriter.

Table 2: Rhetorical Moves identified by AcaWriter with Examples

Move	Description	Example Sentence
Summary (S)	**S**ummarizing the text's aim, goals and conclusions	This paper will examine the question of how we develop scalable learning analytics applications.
Background (B)	Referring to **B**ackground or prior work done in the area	While data was previously studied in educational research, LA now enables more… Recent studies indicate that the effects of the drug could be permanent.
Contrast (C)	Pointing to **C**ontrasting ideas, issues and disagreements	However, a recognized challenge in the field of learning analytics is the uncertainty around LA's pedagogical relevance
Emphasis (E)	**E**mphasizing and ringing attention to important ideas in the text	The key elements for contextualizing LA include… It is important to note that the policy applies to all universities.
Perspective (P)	**P**erspective or Stance from a piece of work	Research has suggested a link between brand perception and customer loyalty.

[5] https://cic.uts.edu.au/open-source-writing-analytics

Novelty (N)	Mentioning **N**ovel and innovative ideas	The new model suggests a view of learning that is an embodied and relational process.
Question (Q)	Highlighting an open **Q**uestion or insufficient knowledge	Little research exists on how automated feedback impacts student writing.
Surprise (S)	Pointing to **S**urprising facts or unexpected findings	Surprisingly, the results indicate a weak link between customer satisfaction and brand value.
Trend (T)	Recognizing **T**rends in research and drifts over time	With the growing quantity of data generated, there is increasing interest in analytics.

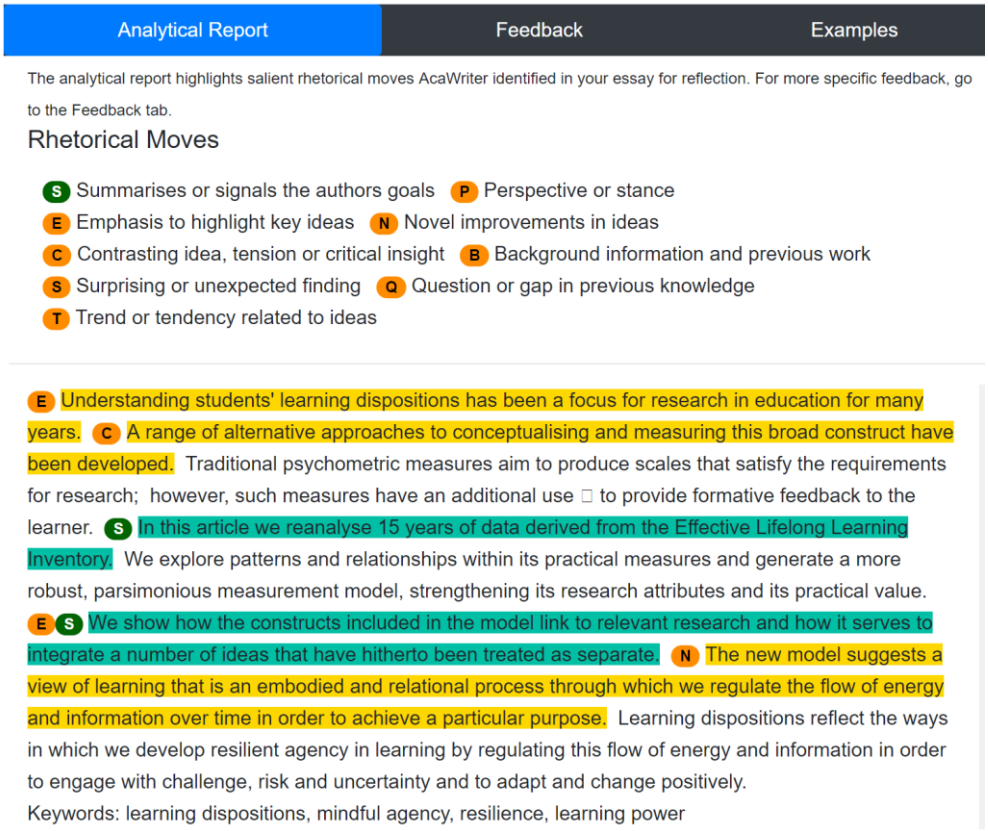


Figure 6: AcaWriter UI highlighting rhetorical moves in the text

The typology of rhetorical moves developed can be used to provide generalized feedback to all analytical writing, which include concepts and arguments supported by pieces of information. However, there are certain rhetorical moves that are more

significant than others for some contexts, including pedagogic contexts that have particular learning foci, as will be illustrated in my empirical studies later. By identifying these contexts using an expert's subject knowledge, AcaWriter provides contextualized feedback by integrating the tool with the learning design using assignment codes tailored for specific genres of writing across subjects, which is the key value added by AcaWriter in comparison to other writing analytics tools. AcaWriter contextualization is hence implemented at the subject level and can be drilled further to assignment levels to provide contextualized automated feedback on writing. This is done by LA designers working as a team with instructors and subject experts for the context to align analytics from the open source tool with the learning design. Sample feedback messages from AcaWriter for Law essays are shown in Figure 7.

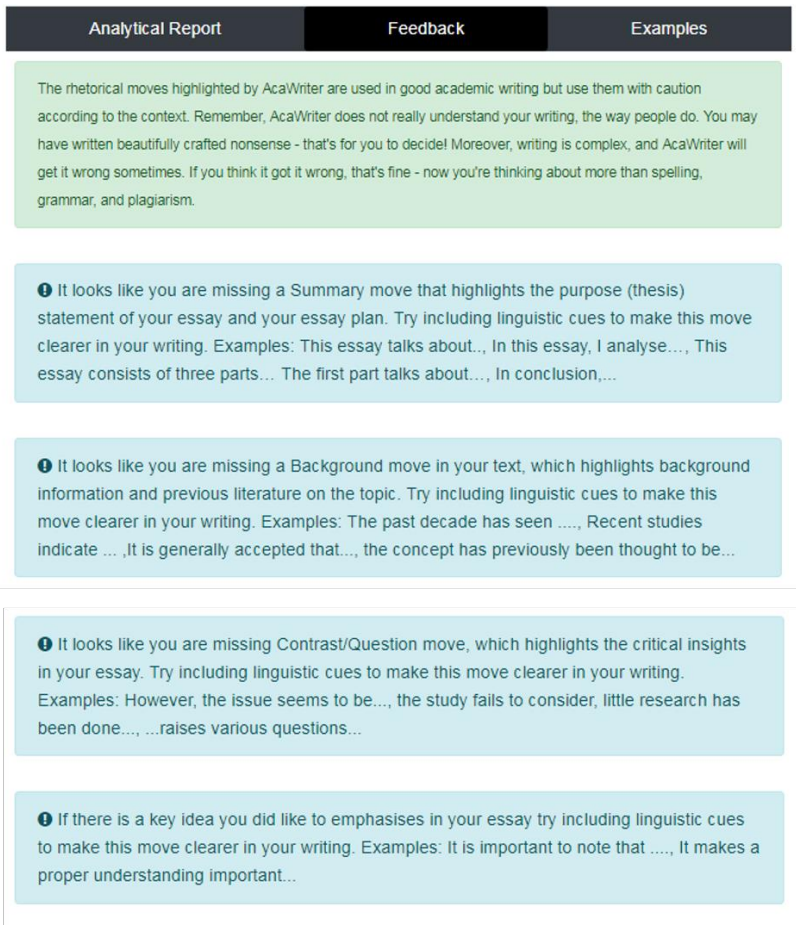


Figure 7: AcaWriter UI showing feedback messages

The full set of applications related to AcaWriter is available open source (Buckingham Shum, 2018), with Jupyter notebooks available to guide people through the tool design by connecting low-level textual features to high-level writing feedback (Shibani, Abel, et al., 2018). Along with the technical components, AcaWriter also builds social infrastructure by creating educator resources illustrating how AcaWriter can be integrated into writing improvement tasks in classrooms: http://heta.io/resources/wawa-improve-business-report-writing-accounting/. The functionality of AcaWriter with its technical background and empirical evaluation is detailed in a recent article (Knight, Shibani, et al., In submission).

2.5 Research and Evaluation of Writing Support

Earlier sections provided a review of related work done to support and improve academic writing in students using human and automated means. However, in order to know if they work in the intended and impactful way, they need to be evaluated in authentic settings. For such evaluations, different methods have been used. One approach is the deployment of surveys and other feedback mechanisms that measure user perceptions on effectiveness/ usefulness. Using quantitative and qualitative analysis methods, these perceptions on learning are explored – for example, studying the ratings and feedback to uncover useful and not-so-useful aspects of the writing support provided (Ball, 2009; Roscoe, Wilson, Johnson, & Mayra, 2017). Other approaches include the analysis of written products that students produce as a result of the writing support provided, to study the actual impact of the support in terms of the changes it induced. This includes studying textual features and the final *products* of writing using grades, counts of sentences, words, errors etc. (Bitchener et al., 2005; Suzuki, 2012). Revisions made in writing have also been studied using manual coding to categorize the types of revisions like unit (e.g., word, phrase, sentence), type (e.g., addition, substitution, spelling), quality (Crawford, Lloyd, & Knoth, 2008), and error correction types (Chandler, 2003). There have been recent efforts to automate classification of revisions based on the content of the changes made (Zhang, Hwa, Litman, & Hashemi, 2016).

Another approach is the study of the *process* of revision that students are involved in during writing (automated methods using analytics to study this process were discussed earlier in Section 2.4.1.2). Text revision is considered an important process in writing to support the reworking of writer's thoughts and ideas, playing a major role in

the outcome of the writing (Fitzgerald, 1987). The cognitive process theory of writing defines revision as a recursive process that can be called any time during writing (Flower & Hayes, 1981). As per this model, writers engage in processes like task definition, evaluation, goal setting and strategy selection to make revisions, thus leading to improvements in a text. By studying this key process of revision, we can study how the writing support we provide helps students improve their writing. Analysis of this revision process has typically relied on resource-intensive manual observation and coding of writing behaviour. It has been studied using personal testimonies of participants regarding their cognitive process in revising, or by process tracing and participant-observer methods that observe the behaviours involved in revision (Fitzgerald, 1987). Such resource intensive manual observation and coding are improved with advanced data collection and analysis techniques in more recent work of my own, which I will elaborate in Sections 4.2.5.4 and 4.4.5.4. Furthermore, student engagement with automated writing feedback has not been studied in past literature by tracking their process of revision. To contribute to this body of knowledge, my study combines the different approaches discussed for the evaluation of writing support in order to study the impact of automated feedback in terms of the writing products and explore how students interact with the feedback in the writing process.

2.6 Integrating technology in pedagogical practice

While the technological advances in data and tools discussed earlier provide new opportunities to support student writing, they should be well integrated in pedagogical contexts to meaningfully impact practice. It has been noted that when harnessing the potential of technology in education, it is important to optimise its characteristics to best impact learning practice. Conole argues that “the complexity of digital environments require us to develop 'schema' or approaches to thinking about how we can best harness the benefits these new technologies confer” (Conole, 2008). She emphasizes that new technologies should be driven by sound pedagogies with alignment to learning theories. Similarly, in Learning Analytics, a critical question on how theory could, or should shape research in this new paradigm has been raised (Knight & Shum, 2017).

Learning analytics promises much more than scaling up using big data, because it is about meaningful data for learning (Merceron et al., 2016). While much learning analytics work has been conducted with large quantities of data in institutional

contexts, for senior management, curriculum designers, and researchers, there is an increasing emphasis on using LA to directly support learners. There have been calls for developing student-facing LA solutions to encourage students towards more sophisticated metacognition about their own learning processes (Kitto, Lupton, Davis, & Waters, 2017). Such applications in learning analytics need attention to theory-informed pedagogic decisions. This is particularly important for the challenge of implementing automated writing feedback, since this must be relevant for the students' learning contexts.

The integration of learning analytics tools in pedagogic design should also be aligned to subject curricula in order to find new ways of solving existing pedagogical issues using learning analytics. This can be supported by the pedagogical frameworks developed in education for the use of technology in teaching and learning contexts and the design of assessments. They support the integration of information and communication technology (ICT) and design elements in pedagogy. Existing approaches that provide a conceptual understanding for such design of learning analytics include Evidence-centred design and Orchestration, which are discussed next, followed by the application of Learning Design (LD) in Learning Analytics using a Co-design methodology.

2.7 Approaches for technology integrated practice

2.7.1 Evidence centred design

Evidence-Centred Design (ECD) is an approach which encourages the design of assessments with clear evidence on the data to be collected and inferences to be made before an assessment is designed (Mislevy, Almond, & Lukas, 2003). It provides a conceptual assessment framework consisting of models that answer specific questions to design an assessment as below:

- The Student model that defines what the assessment wants to measure
- The Evidence model that defines the instructions on how to measure it
- The Task model defining the situations where the assessment can be designed
- The Assembly model defining the targets of how much should be measured
- The presentation model that describes how the assessment tasks should be presented

- The Delivery system model that defines the collection of the above models to work together

ECD aligns with a construct-centred approach emphasizing definition of the construct to be measured and asking critical questions on the collection of evidence from an assessment using an explicit model. It further organised the processes within the design the implementation phases of an assessment with the use of layers like domain analysis, domain modeling, conceptual assessment framework, assessment implementation, and assessment delivery (Mislevy & Riconscente, 2006). It supports the argument against the usage of off-the-shelf assessments that are standardised to all contexts.

In Learning Analytics, ECD helps to guide researchers frame the questions to be asked from students' learning (Writing, in this case). It also helps researchers and instructors best support learners by designing effective tasks for their learning. Components of ECD can be utilised for pedagogic design of learning analytics applications to impact writing in authentic classroom settings. It is aligned with the 'Orchestration' approach discussed next, which aims to manage the processes in authentic computer-supported learning scenarios.

2.7.2 Orchestration

Orchestration in Technology Enhanced Learning (TEL) involves co-ordinating and managing the different pedagogical activities in a formal learning context. The main 5 aspects in orchestrating TEL (Prieto, Holenko Dlab, Gutiérrez, Abdulwahed, & Balid, 2011) are:

1. Design/Planning of learning activities and tools to enact them.
2. Regulation/Management of learning activities manually or by automated means in order to modulate the characteristics of the subject in terms of time, workflow, group management etc.
3. Adaptation/flexibility/intervention of activities to suit the local context and the emerging occurrences during its enactment in a classroom.
4. Awareness/assessment of what's happening in the classroom in order to make well-informed responses.

5. Roles of the teacher and other actors (learners in learner-driven orchestration) who drive the orchestration as guiders to facilitate learning.

Three further aspects define the challenges and how orchestration should be carried out as below:

- Pragmatism/practice of orchestration to change educational practice in authentic settings
- Alignment/synergy to co-ordinate different elements like learning activities, tools and scaffoldings to achieve the learning goals,
- Models/theories to backup empirical data with theoretical models to guide orchestration

The 5+3 conceptual framework described above was further modified to develop a more complete framework which distinguished the role of different elements in orchestration and clarified the aspects (refer Figure 8). Based on the revised framework, orchestration in learning technology research has been defined as *'the process of designing and managing in real-time (including awareness and adaptation mechanisms) the learning processes in an authentic computer-supported learning scenario'* (Prieto, Dimitriadis, Asensio-Pérez, & Looi, 2015) (pg 12).

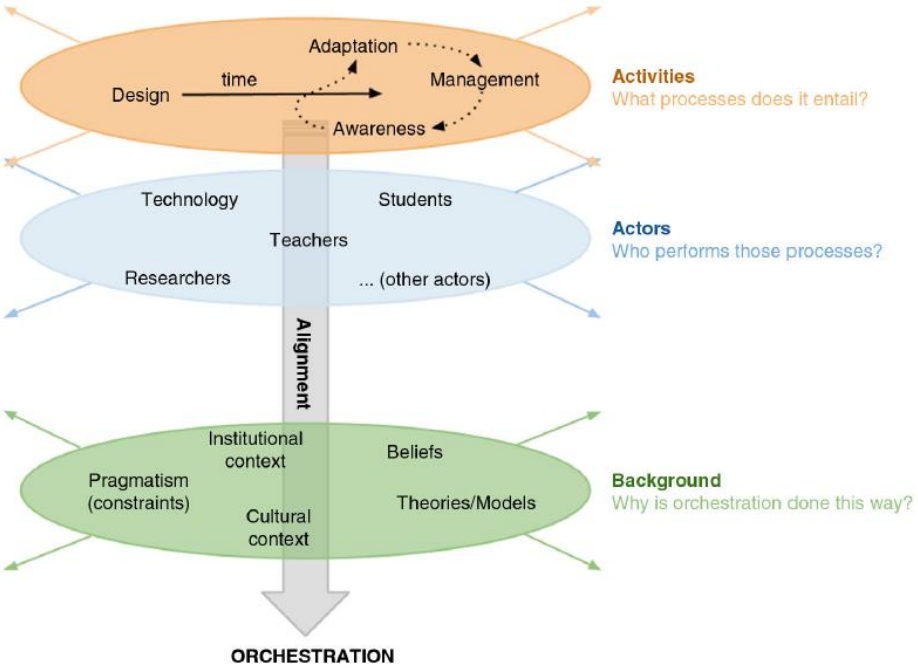


Figure 8: Revised conceptual framework for orchestration used by Prieto et al. (2015)

This framework separates the activities, actors, and background in orchestration, and provides a coherent framework with the different elements of learning. It helps to design pedagogical activities and research instruments involving technological tools for authentic classroom settings, which can be employed by Learning Analytics. In recent work, orchestrating learning analytics aims to promote inter-stakeholder communication by focussing on classroom-level constraints faced in everyday educational activities, in order to favour adoption at the practitioner level (Prieto, Rodríguez Triana, Martínez Maldonado, Dimitriadis, & Gašević, 2018). Similar to the ECD approach discussed earlier, by giving importance to such contextual factors through orchestration, Learning Analytics can impact learning, and improve adoption at institutional levels.

2.7.3 Learning Design for Learning Analytics

It is known that for LA to work effectively, it must be coupled with pedagogical approaches. The contextualization of LA, then, arises from its integration in pedagogical contexts to augment the learning design and provide analytics that are aligned with the intended learning outcomes (Knight, Shibani, & Buckingham Shum, 2018). While 'learning design' is the widely used term, its preferred usage, 'designing for learning', distinguishes between what can be designed and what emerges subsequently: As Goodyear and Carvalho explain, "We can design tasks or tools or other things that help someone learn, but we cannot do the learning for them" (Goodyear & Carvalho, 2014). The usage of 'Learning Design' is hence a shorthand to referring to 'Designing for Learning' in this thesis.

A pedagogically driven approach prompts the alignment of learning analytics to learning design towards contextual frameworks that are defined for the pedagogic intent of analytics applications, such that learning analytics can provide the necessary data, methodologies and tools to test the assumptions of the learning design (Lockyer et al., 2013). An understanding of the pedagogical contexts enhance our interpretation of analytics into actionable insights and meaningful interpretations (Gašević, Dawson, Rogers, & Gasevic, 2016). This learning design approach addresses a known concern that even high-quality technologies may not be used by students if they are not embedded in the curriculum. A clearly defined learning design helps students use the tools to add value to their learning by closing the gap between the potential and actual use of technologies (Shibani, Knight, Buckingham Shum, & Ryan, 2017). Existing work

on aligning learning analytics and learning design are discussed next. For LA to be impactful, it is argued that there is a necessity to make a transformative shift from exploratory studies to evaluative research that can study the impact of LA at an institutional level (Dawson, Joksimovic, Poquet, & Siemens, 2019). This will facilitate the move of LA applications from laboratory based environments to authentic classroom settings for use by practitioners. Learning analytics pedagogic intervention designs have been proposed for integrating LA technologies as part of a larger educational context (Wise, 2014). This way, Learning Analytics can focus on 'augmenting', rather than revolutionizing existing high quality pedagogy by enhancing classroom practices (Knight, Shibani, et al., 2018). By finding new methods to solve existing pedagogical issues, LA can contribute to existing good practices by improving them with insights and technical affordances, rather than analytics remaining separated from current classroom teaching and learning practices.

To aid this LA-LD alignment, Bakharia et al, proposed a conceptual framework linking learning analytics to learning design. In this framework, the teacher plays a central role by bringing context to the analysis, making decisions on the feedback to be provided to students and in the adaptation of learning design (Bakharia et al., 2016). Similarly, Alhadad and Thompson (Alhadad & Thompson, 2017) propose the mediation of effective teacher inquiry processes, enhancing opportunities for genuinely evidence-informed practice by avoiding the oversimplification of learning enhancement to a data-driven process. Despite the importance of aligning LA and LD, few empirical studies demonstrate how this alignment happens in practice across pedagogical settings. A detailed systematic review of 43 studies in the current landscape connecting LA and LD can be found in (Mangaroska & Giannakos, 2018). Based on this review, Mangaroska and Giannakos suggest that a framework should be developed to capture and systematize learning design grounded in learning analytics, and study how learning design choices influence learning and performance over time. They also emphasize the need for educators to design for learning by making use of LA affordances as opposed to being providers of knowledge only. The next section will discuss how it can be implemented with involvement from stakeholders in the design of LA.

2.7.4 Co-design: Stakeholder involvement

It has been emphasized that learning analytics is about learning (Gašević et al., 2015), and has to be better integrated into existing educational settings for learning analytics

research and practice. However, large-scale adoption of learning analytics at an institutional level comes with challenges that call for new adaptive forms of leadership, collaboration, policy development and strategic planning (Macfadyen, Dawson, Pardo, & Gaševic, 2014; Tsai & Gasevic, 2017). While enabling such adoption to impact learners at scale, there are challenges in bringing relevant LA to practitioners at the implementation level as well.

The bulk of empirical LA studies focus on student outcomes and performance measures using tools with little focus on the barriers to adoption of these tools like practitioner involvement (Klein, Lester, Rangwala, & Johri, 2019). Among the key groups of stakeholders, including Learners, Educators, Researchers, and Administrators (Romero & Ventura, 2013), there is limited research on the role of educators in integrating learning analytics in authentic practice. Among the notable studies that explore teacher interaction, two approaches have been taken. A set of studies have investigated teacher design and inquiry processes, for example in a professional development design workshop on designing for learning using learning analytics (Alhadad & Thompson, 2017), while the 'completing the loop' project investigated how learning analytics from Learning Management Systems (LMS) could be delivered to support teachers (Corrin et al., 2016). This was done by interviewing university teachers to understand what analytics they would find useful, and implemented a tool to deliver meaningful analytics.

Other studies have investigated educator perspectives and sense-making on learning analytics, for example regarding information from LA focused on online collaborative learning (van Leeuwen, van Wermeskerken, Erkens, & Rummel, 2017). Similarly in a study by Holstein et al, teacher expectations were explored to see how Intelligent Tutoring Systems (ITS) with real-time dashboards could be designed for blended classrooms (Holstein, McLaren, & Aleven, 2017). This study used design interviews including generative card sorting exercises, semi-structured interviews, directed storytelling, and Speed Dating sessions to understand teacher notions before designing an ITS for them. Other studies have examined teacher views to qualitatively evaluate the usefulness of LA applications after they are implemented (Echeverria et al., 2018; Koh, Shibani, Tan, & Hong, 2016). More generally, the sub-field of teaching Analytics has focused on capturing and analysing teacher actions to help teachers improve educational designs prior to their delivery (Sergis & Sampson, 2017).

While these studies give insight into how educators reflect on specific LA applications, they generally do not provide detail about long term usage, or the ways that educators adopt and adapt learning analytics to their specific context. Thus, across this body of work the perspectives of teachers regarding learning analytics and their use in existing practice is limited. To enable greater opportunities for cross-fertilization between research and practice in learning analytics, there should be added emphasis on the perspectives of educators, who are one of the main stakeholders to effectively implement learning analytics.

The inclusion of stakeholders like teachers and students in the design process of LA has been suggested to achieve responsible LA innovations (Knight & Anderson, 2016; Prieto-Alvarez, Martinez-Maldonado, & Anderson, 2018). This participatory approach emphasizes the need to take into account practical considerations of different stakeholders to bring LA in authentic practice for making it effective for learners (Greller & Drachsler, 2012). Co-design is one 'approach where learners, educators, institutions, researchers, developers, and designers are all included across different stages of the design process, from exploration to actual implementation' (Prieto-Alvarez et al., 2018). Such participatory approaches enable openness of data and platforms across stakeholders such that the needs and values of respective stakeholders are met in accordance to Value-Sensitive Design (Knight & Anderson, 2016). Value Sensitive Design is an endeavour that proactively considers human values throughout the process of technology design (Davis & Nathan, 2015). It offers a systemic approach with specific strategies and methods to help researchers and designers explicitly incorporate the consideration of human values into design, and illuminate tensions among stakeholders. A recent study illustrated the use of Value Sensitive Design for LA system design, and discussed implications on the holistic integrity of LA systems (B. Chen & Zhu, 2019). To take such a Value Sensitive Design approach to design LA for learners, a sensible first step is to be involve educators, who are the main stakeholders in the design of LA for students. This has been reiterated in Orchestration of LA, which points us to the fact that considering classroom-level constraints by enabling inter-stakeholder communication, and orchestrating LA to suit them, can better aid the adoption of LA technologies (Prieto et al., 2018). Given that the limitations of designing LA without regard to specific learning contexts are explained in depth earlier, and aligning LA with LD has been highlighted as a potential solution, educators can bring in this alignment with LD for their learning contexts. This method of contextualizing LA for authentic

classroom scenarios by involving educators in a co-design process will be explained in the implementation of the current study in Chapter 4: Design Iterations in Learning Context 1.

2.8 Summary and Implications

From existing literature, it is seen why writing is considered a crucial skill for students to develop in their higher educational practice. Due to the significant role played by effective writing in academic and professional roles, its development is given paramount importance. To support students in developing this key skill, various forms of instruction and feedback have been studied in past research. A common approach is the provision of feedback on writing by instructors who assess student writing. However, due to the time-consuming nature of assessing writing, instructors often don't have the time and expertise to provide formative feedback that helps students improve their writing before summative assessments.

Also, for students to move from feedback to self-monitoring for sustainable learning, they ought to learn evaluative judgement of their own work. This is enabled by designing self-assessment activities with writing, and is often coupled with feedback practices. Peer feedback is one such method which can enhance student learning and evaluative judgement by providing and receiving feedback on writing. But it also depends on the feedback provider to give useful feedback, and on the feedback receiver to interpret and apply the feedback effectively. Another method that encourages students to evaluate the work of others to improve judgement, and learn from concrete examples is through the provision of exemplars, and these are seen as good aids to help students understand assessment criteria for specific types of writing. The kinds of writing support discussed from related work have implications in the current study to recognise and design effective forms of writing support to students.

New technologies, forms of data, and artificial intelligence have contributed to the emergence of the field of Learning Analytics, which applies those affordances to optimise educational practice. In its sub-field focussed on writing called Writing Analytics, new techniques enable novel forms of feedback that never existed before, and more fine-grained, time-efficient analysis. This has implications on the types of analysis we can perform on student data to study writing and revision, and the types of automated feedback that can support students on their writing. A detailed review on

writing tools shows the scoring and feedback mechanisms currently available. These tools have a range of functionalities from automated scoring to automated feedback, and use textual features like cohesion, word usage, grammar, sentence organization, rhetorical structures etc. to provide feedback on student writing. A particular problem with existing tools is the lack of balance between generalizable and contextualizable support for writing. Few tools offer generalized support based on textual features with no consideration on the topic of content, while few others offer specific feedback for prompts on a particular topic that cannot be easily transferred to other writing. In the current study, a writing analytics tool called AcaWriter is used (previously called AWA), which can provide contextualized feedback on student writing based on assessment criteria of specific writing using generalized rhetorical structures in writing.

Existing research has also explored evaluation methods to study the application of writing support in various forms, but lack empirical evidence of impact in writing practice. This includes manual and automated methods to study student perceptions of writing instruction, revisions made to the writing as a result of the provided support, and tracing the process of revision students are engaged in using the support in experimental and authentic writing scenarios. By building on these methods, the current study evaluates the impact of AcaWriter feedback on products and processes in writing. This is done by implementing writing support using AcaWriter in pedagogical interventions across authentic classrooms in higher education, and evaluating its impact using empirical studies, detailed in Chapter 4: Design Iterations in Learning Context 1 and Chapter 5: Conceptual Model and Design Transfer to Learning Context 2.

Furthermore, the literature review discussed the need for pedagogic theory behind technology in education and the design of learning analytics to support learners in practice. It highlighted opportunities for the alignment of learning theory with learning analytics using existing theoretical frameworks to design for learning in authentic higher education classrooms. Evidence Centred Design is one such framework that encourages the design of assessments with clear evidence on the data to be collected and inferences to be made before an assessment is designed and makes the case against the use of off-the shelf assessments for learning analytics. Orchestrating learning analytics can also promote inter-stakeholder communication by focussing on classroom-level constraints faced in everyday educational activities, in order to favour adoption at the practitioner level.

To couple LA with pedagogical approaches, there also needs to be an alignment of learning analytics and learning design. A clearly defined learning design helps students use learning analytics to add value to their learning by closing the gap between the potential and actual use of technologies. Existing frameworks have emphasized this link between learning analytics and learning design by giving teachers the central role of bringing context to the analytics. However, there is limited knowledge in the implementation of learning analytics in authentic practice by involving educators in the design process. Such active involvement of stakeholders has been reiterated by participatory approaches that involve educators in the co-design process for value sensitive design. This approach will be elaborated in the following chapter, detailing the methodology used in the thesis.

Chapter 3: Methodology

This chapter discusses the methodology employed in my overall research design. Section 3.1 first introduces the research questions driving the study. Section 3.2 then discusses the design-based research approach and the rationale behind this overarching educational research methodology used in the thesis. The learning contexts where the study is implemented are discussed next in Section 3.3, followed by data collected for the study and its mapping to the research questions in Section 3.4. Section 3.5 discusses the methods used for data analysis, mapping them back to the research questions they address, and an explanation of the procedures and timeline in data collection are presented in Section 0. Finally, Section 3.7 discusses ethical considerations and how they were tackled in the study.

3.1 Research Questions

As detailed in Chapter 2: Literature Review, there is a significant gap in the existing literature with regards to the impact of writing analytics tools. The gap was also discussed in terms of the lack of theory and implementations of pedagogically integrated learning analytics in authentic practice. Hence, the overall aim of my research is:

> *"To investigate the implementation, and impact on student writing, of an automated feedback tool in higher education teaching practice."*

Based on the overall aim, three main research questions (RQs) described in detail in this section guide the study.

The first research question aims to study the impact of automated feedback on student writing in terms of the writing product produced. The actual impact of the writing analytics tools in students' writing skills was found to be a less studied construct in literature when compared to the usability and accuracy testing of such tools. In addition, the evaluation of tools were either based on user perceptions of the tool, or on text features of the writing produced using the tool, but rarely a combination of both. Hence the focus of RQ1 is on studying the impact of automated feedback on students' writing products, by exploring their perceptions of automated feedback, and studying

the impact of feedback by measuring changes made to the writing product, i.e, revisions made. RQ1 and its sub questions are as follows:

RQ1 - Writing Products: What is the impact of automated feedback on rhetorical moves in student writing?

RQ1a. What are students' perceptions of the writing task with/ without automated feedback?

RQ1b. What is the impact of automated feedback on student revisions?

The second research question aims to go deeper into studying student engagement with automated feedback by exploring the writing processes involved. Existing literature reveals limited knowledge about the patterns of student engagement with automated feedback by tracking their revision to study what goes on in this process. Learning analytics can address this limitation through the capability to track students' drafting and revision process. Hence, RQ2 will use such processes to examine in detail how students interact with automated feedback and their trajectories of feedback use. It will also investigate whether scaffolding the students with additional support, like peer feedback and prompts, can impact their engagement with automated feedback. RQ2 is therefore described as follows:

RQ2 - Writing Process: How do students engage with automated writing feedback?

RQ2a. How can we study students' interactions with automated feedback?

RQ2b. How does scaffolding (using peer feedback and additional instruction) impact student engagement with automated feedback?

The third research question aims to investigate stakeholder perspectives in practical implementations of AcaWriter in authentic classroom settings. In many learning analytics research studies, this element of practical application and the factors influencing educators in adoption are not discussed. The multitude of steps behind implementations in practice are also barely reported, which inhibit the ability to uncover successful traits of design-based research in learning analytics. Hence, RQ3 will explore *educator* perspectives by examining their motivations, challenges, and support required for adoption. It will also examine further in detail the implementation of writing analytics in authentic classrooms, and its outcomes. RQ3 is as follows:

<u>RQ3 - Educator Perspectives:</u> What are practitioner perspectives on automated writing feedback in authentic practice?

RQ3a. What factors influence adoption in authentic classrooms?

RQ3b. How do the practitioners engage in implementation across disciplines, and what are the outcomes?

3.2 Design-Based Research

The overarching approach of the study in this thesis is Design-Based Research (DBR), which is "a systematic, but flexible methodology aimed to improve educational practices through iterative analysis, design, development, and implementation, based on collaboration among researchers and practitioners in real world settings, and leading to contextually-sensitive design principles and theories'' (pp. 6–7) (Wang & Hannafin, 2005). DBR is a methodological paradigm that strives to solve real-world problems through multiple iterations of development, based on lessons from practice and collaborative partnership between researchers and practitioners (Anderson & Shattuck, 2012).

First introduced by Brown (1992) and Collins (1992) as 'design experiments', a core part of DBR involved situating the work in naturalistic contexts, because of the belief that many of the questions important to them could not be adequately addressed by laboratory-based examinations. The focus of DBR is on developing both theory and practice through addressing practical design problems in applied contexts. In addition, Reeves (2006) notes that DBR is appropriate for educational research because it emphasizes content and pedagogy rather than technology, by giving special attention to supporting human interactions and nurturing learning communities. The pedagogical outcome defines the goal, and the learning environments can be modified until the goal is reached, with the underlying assumption that it is difficult to control the variables in real classroom contexts. This also helps reflect on the process to reveal design principles that can be shared to inform other instructors and researchers, and future development projects, thus maximising the possibilities for dissemination. Furthermore, DBR contributes to the creation of new theoretical principles that are intertwined with practice rather than basing the implementation on an existing theory from the start (Hein, 2017).

Barab and Squire (2004) highlight some key characteristics of DBR, which make it distinctive compared to traditional psychological experimentation methods adapted from Collins (1992), as follows:

- **Location of Research:** In contrast to experimentation conducted in laboratory settings, DBR occurs in the 'buzzing, blooming confusion of real-life settings where most learning actually occurs' (Barab & Squire, 2004).
- **Complexity of variables:** DBR often involves multiple dependent variables, including climate variables (e.g., collaboration among learners, available resources), outcome variables (e.g., learning of content, transfer), and system variables (e.g., dissemination, sustainability). This is different to experimentation methods which usually involve only one or two dependent variables.
- **Focus of research:** While the focus of experimentation is on identifying a few variables and holding them constant, DBR focuses on characterizing the situation in all its complexity, much of which is not derived by reasoning from self-evident propositions.
- **Unfolding of procedures:** Experimentation methods use fixed procedures, but DBR uses flexible design revisions, where the initial set of procedures are revised depending upon their success in practice.
- **Amount of social interaction:** DBR encompasses complex social interactions with participants sharing ideas, distracting each other, and so on – reflective of real classroom settings, in contrast to controlled experiments where learners are isolated to control interaction.
- **Characterizing the findings:** Experimentation methods generally focus on testing a hypothesis. On the other hand, DBR involves looking at multiple aspects of the design and developing a profile that characterizes the design in practice.
- **Role of participants:** While experiments treat participants as subjects, DBR involves different participants in the design to bring their expertise to produce, analyse, and improve the design.

The above factors align with the aim of my research thesis, which is to apply learning analytics in *authentic practice* for students to improve their writing. While making great use of qualitative data, DBR does not preclude the statistically meaningful comparison of different cohorts and conditions in order to assess the impact of an intervention. Examples can be seen in other work (Feng, 2015; Koh, Tee, et al., 2016; Nelson, Ketelhut, Clarke-Midura, Bowman, & Dede, 2005) and in this thesis, see the results reported in Sections 4.2, 4.3, 4.4, and 5.2. However, the complexity of the real classrooms, teachers and cohorts at the centre of the empirical work precludes investigations into the impact of detailed variables. More controlled conditions with higher definition data would be required to investigate, for instance, subtle variations in training procedures, or the cognitive significance of typing patterns from keystrokes (Conijn, Loo, & Zaanen, 2018).

While the authenticity might compromise the ability to prove hypotheses in a controlled manner, DBR makes my research focus on the pedagogical outcome, i.e, improving student writing in authentic classrooms more prominent. DBR supports effective integration of technology in practice, which was identified as a challenge for learning analytics in Section 2.6. It helps understand the complexities of such research in practice at a holistic level by considering multiple variables. Thus, DBR is an ideal choice for my study due to its applicability to real-life classroom settings. It accounts for the social interactions that are a naturally occurring part of authentic educational practices when exploring my research questions. It also emphasizes the involvement of stakeholders to shape the design, which forms a key aspect of my research involving educators and students in higher education. In using this research methodology, the empirical work of my study is implemented in two higher education learning contexts, described next. This ability to scale up to settings other than the ones originally developed for also contributes to system level outcomes and strategies for implementation at a policy level, which define the characteristics of Design Based Implementation Research (DBIR) (Penuel, Fishman, Haugan Cheng, & Sabelli, 2011) - discussed more in Section 7.4.

3.3 Learning Contexts

The DBR iterations of the current study involved implementations of writing interventions in authentic higher educational learning contexts within an Australian university. These contexts involved academic writing as one of the key skills student

develop as part of their learning as identified by prior work (Graham et al., 2013; Lea, 2004), and involved specific types of writing. The two contexts where the study took place, and their writing contexts are described below. It is to be noted that while the current study includes two learning contexts to demonstrate its effectiveness, the reported intervention designs are also widely applicable to a number of other learning contexts that develop academic writing as a core competency in students, for example, Research Abstract writing, as reported by Abel, Kitto, Knight, and Buckingham Shum (2018). These contexts were chosen because the instructors identified the need for additional writing support for their students, they were interested in working with the researchers to make use of AcaWriter for this purpose, and the researchers confirmed that in principle, AcaWriter's capabilities appeared to match the forms of writing being taught.

3.3.1 Undergraduate Law subject

The first learning context was an undergraduate Law subject teaching 'Civil Procedure' consisting of about 280 to 400 Law students in each cohort. This was the main context where most iterations (from iteration 1 to iteration 3) were designed and implemented. Civil Procedure was a compulsory subject which ran for one semester in the faculty of Law in the university, and generally for all Law students in Australia. It was located in the middle of their degree as a mid-degree subject in the overall course structure, and hence students usually had prior experience in academic writing and were expected to produce high quality writing.

Writing is a key disciplinary skill for Law students with emphasis on clear and engaging writing with the use of appropriate arguments (Knight, Buckingham Shum, et al., 2018), and was identified as an area to target student learning. The specific type of writing that students were asked to produce in this subject were ***argumentative law essays*** of about 2000 words. These academic essays should discuss a topic of their choice from a list of assigned topics which are quite provocative, clearly outlining the legal arguments to argue for or against the particular proposition. The topics are set out so that both sides are arguable. In their key written assignment, the instructor had developed a marking rubric consisting of the following elements: Statement of argument, Statement of essay plan, Identification of issues, Analysis, Sustained thesis & Original insight, and Engagement with literature/cases. Students were required to submit the argumentative essay individually half way through the subject for an

assignment, which was assessed using the above-mentioned elements of the assessment criteria.

3.3.2 Undergraduate Accounting subject

The second learning context was an undergraduate Accounting subject which taught 'Management Decisions and Control' for about 400 to 600 business students in each cohort. This was the learning context for iteration 4, in which the task design from the previous Law context was transferred to a new discipline. In addition to the development of discipline-related skills, this subject also contributed to the faculty graduate attributes by developing students' ability to 'convey information clearly and fluently in high quality written form appropriate for their audience', which is a Course Intended Learning Outcome (CILO). The cohort had a wide mix of local and international students, who had varied levels of English ability from high to low.

As part of their Accounting degree, students had to practice and develop their written communication skills in a professional business manner. This subject helped students develop their writing skill with an individual writing assignment. The assessment piece or assignment was about 1500 words, and oscillated in terms of it being an essay or a report. In the current study context, students were asked to write a ***business report*** by defining an organization's performance by selecting one of the organizations in the given list. It was assessed against four elements of assessment criteria: 1. Organisational analysis, 2. Defining performance, 3. Justification of their definition of performance, and 4. Written communication.

3.4 Data Collection

Data for the study was collected primarily from *students* who participated in writing interventions in subjects they undertook as part of their usual learning. Data was also collected from *instructors* who co-designed the writing interventions with the researchers, using interviews to elicit their perspectives on implementing learning analytics in authentic practice. Data came from multiple design iterations of the writing intervention, and included the following to address different research questions:

- Students' responses to feedback questionnaire
- Students' assessment of given texts and their improved texts
- Revisions made by students in the given texts

- Logs of students' interactions with the online platform (AWA-Tutor)
- Logs of students' interactions with the automated writing feedback tool (AWA/ AcaWriter)
- Intermittent drafts of students using the automated writing feedback tool
- Audio recordings of students' peer discussion conversation while engaging with automated writing feedback
- Audio and video recordings from instructor interviews

3.5 Analysis

The study uses mixed methods to design and analyse data for answering the research questions. Quantitative methods use experiments and statistical measures to study relationships between variables, and are predominantly driven by numbers validating a hypothesis (Khandker, Koolwal, & Samad, 2009). On the other hand, qualitative approaches provide insights that are deeper in nature with interpretive rich descriptions even from small sample sizes (Erickson, 1985). By bridging across the traditions and integrating the strengths of both methods, a mixed-methods approach employs both quantitative and qualitative methods as appropriate. By making use of multiple methods this way, triangulation can be achieved to study the same phenomenon through different lenses, leading to convergent results (Jick, 1979).

In learning analytics, there is a concern that the favouring of quantitative approaches might lead to a focus on finding generalizable structural relations rather than understanding nuanced processes in learning (Wise & Cui, 2018). To find the balance, my study adopts the mixed methods approach to examine the data from multiple sources that are in varied forms. While some data yield numerical results, others need to be fully understood in depth using human interpretations. For this reason, I followed the mixed methods approach for my thesis to analyse the different types of data collected from surveys, logs, and interviews using different types of data analysis like statistics, thematic coding, and text analysis for the respective research questions. A mapping of the research questions to the corresponding data and analysis is shown in

Table 3.

Table 3: Mapping of RQs, data sources, and analysis methods

Research Question (RQ)	Data source	Analysis
RQ1. Writing Products: What is the impact of automated feedback on rhetorical moves in student writing?		
RQ1a. What are students' perceptions of the writing task with/ without automated feedback?	Students' self-rating of usefulness, Students' open-ended responses	Quantitative analysis, Qualitative analysis, Comparison plots
RQ1b. What is the impact of automated feedback on student revisions?	Students' revised texts	Quantitative analysis, Text analysis, Comparison plots, Graph analysis
RQ2. Writing Process: How do students engage with automated writing feedback?		
RQ2a. How can we study students' interactions with automated feedback?	Student drafts and requests for feedback from logs	Graph analysis, Text Analysis
RQ2b. How does scaffolding (using peer feedback and additional instruction) impact student engagement with automated feedback?	Students' Usefulness rating, Students' Open-ended responses to survey questions and self-evaluation prompts, Peer Discussion recordings	Quantitative analysis, Qualitative analysis, Discourse analysis
RQ3. Educator Perspectives: What are practitioner perspectives on automated writing feedback in authentic practice?		
RQ3a. What factors influence adoption in authentic classrooms?	Instructor Interviews	Thematic analysis
RQ3b. How do the practitioners engage in implementation across disciplines, and what are the outcomes?	Instructor Interviews, Tutor Surveys, Co-design session notes	Qualitative analysis, Design representations

3.6 Procedure and Timeline

Following the Design-Based Research methodology, the study implemented several design iterations to collect data from students and instructors over a span of three years. Data used to answer the research questions came from different design iterations, and are described in detail in Chapter 4: Design Iterations in Learning Context 1 and Chapter

5: Conceptual Model and Design Transfer to Learning Context 2 under each iteration. The timeline of the design iterations and the key activities involved are shown in Figure 9. The first three iterations were designed for the Law subject described in Section 3.3.1 and implemented over four semesters. Then, the transfer of design to an Accounting subject (described in Section 3.3.2) took place in the next design iteration. Finally, interviewing instructors and surveying tutors were conducted.

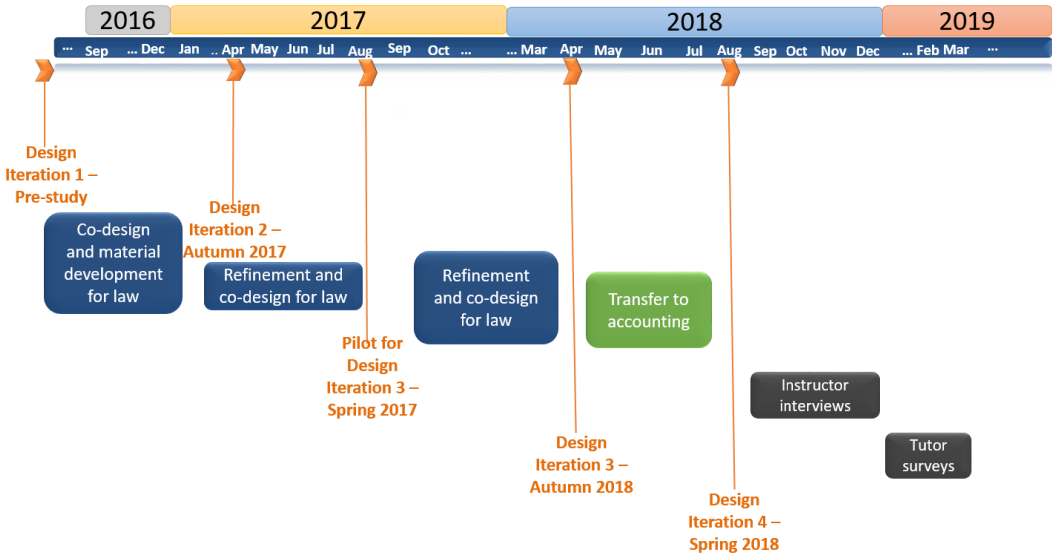


Figure 9: Timeline of Design Iterations in the study

3.7 Ethics

Trust and ethical issues surrounding the use of data and technology in educational settings is a known concern for learning analytics (Slade & Prinsloo, 2013). A recent report states that many institutions are still formulating their approach to ensure that LA is used appropriately for ethical practice in the field (Corrin et al., 2019). The concerns in the current study to be considered for the ethical use of LA are as follows.

3.7.1 Use of control conditions

To investigate the impact of automated writing feedback in authentic practice, the study embedded group comparisons in normal instructional design using control conditions. Students were assigned to one of the control conditions in the Law design iterations so that they either receive automated feedback from AcaWriter (automated feedback group), no feedback (No feedback group) or instructor feedback (Instructor feedback group). This difference in feedback only affected one of the tasks that students do – a revision activity, maintaining the remaining writing support the same for all students. Even without the provision of automated feedback, the writing interventions were

designed in a way that they were pedagogically sound and able to support improvements in student writing based on other effective mechanisms reviewed in Section 2.2. The additional benefit of the group receiving automated feedback was also uncertain, since no prior evidence of the impact of AcaWriter on student writing was measured in a systematic way. Once a positive effect was observed in the automated feedback group in a stabilised task design, the next iteration onwards were designed to provide automated feedback for all students to ensure equal learning opportunities.

3.7.2 Use of student and instructor data

Technological tools that collect *data* often have underlying privacy issues as users are not always fully aware of the data that is being collected about them, the ways in which it is being used, and who has access to it. Since the current work involved collection of data from student activities and educator engagement from authentic classrooms, it ensured that ethical practices are maintained in this research. A participant information sheet allowed students to know the intended purpose of the research, providing them an option to opt out of the use of their data for research. 'The opt-out approach is a method used in the recruitment of participants into research where information is provided to the potential participant regarding the research and their involvement and where their participation is presumed unless they take action to decline to participate' (National Health Medical Research Council, 2007). Since the activities students undertook were part of their normal educational practice in the classroom, their data was presumed to be used for research unless they opted out of its use by emailing the researcher (with no need to provide a reason). Fewer than 5 students opted out of the use of their data in this whole study. Data was managed securely, made accessible only to researchers, and reported anonymously so that the identities of participants were protected. Furthermore, the study did not collect or analyse any sensitive data (E.g. Demographics) that was not pivotal to the aim of the study.

3.7.3 Use of analytics

The second concern is the use of *analytics* to provide feedback on student writing. The nature of such automated feedback is different from what humans are capable of providing owing to the complexity of language, making it inherently imperfect. For instance, the automated feedback might wrongly highlight a rhetorical move as another move or miss to identify a rhetorical move. To teach students the critical usage of such technologies, the current research included educators as partners to design tasks that

could lead to better understanding of how students can effectively use such feedback in spite of its imperfections. The tasks provided opportunities for students to critique and disregard the feedback as required. Students were instructed to consider other factors in their writing not highlighted by the tool as well, so they were made aware that the automated feedback might not be comprehensive enough to consider all aspects of writing, and that they would still need to exercise their own judgement. In this way, the learning design acted as a safety net for the learning analytics support, with instructors acting as gatekeepers for the ethical use of AcaWriter. The design of learning tasks also ensured that the analytics were only used for additional feedback for students, and neither replaced existing support, nor graded students using automated means. The automated detection of rhetorical moves in AcaWriter is powered by human-coded NLP rules, which are transparent, inspect able, and based on published research articles. This ensures transparency and provides algorithmic accountability, in contrast to a machine learning AI system black box which cannot explain why it classified a given text in a certain way (See discussion on algorithmic accountability in Knight, Buckingham Shum, et al., 2018).

Chapter 4: Design Iterations in Learning Context 1

The design-based research approach of this thesis led to the development of several design iterations in authentic learning contexts. These design iterations involved writing interventions implemented for students to improve their subject-related writing skills in real classrooms with the use of automated feedback. The results and findings from the previous iterations were used to improve the intervention in the forthcoming iteration designs. Figure 10 presents an overview of the different design iterations involved in the study. This chapter discusses the design iterations in the first learning context (undergraduate Law subject described in Section 3.3.1), along with the data and analysis for research questions corresponding to each iteration[6].

The study adopts a perspective of 'augmentation', by taking this design approach to analysing teaching and learning contexts, to investigate where existing good practice of writing instruction might be augmented by learning analytics, further strengthening that practice. This approach is likely to increase adoption both of the analytics, and of the underlying practices, thus driving forward implementation of such learning designs, and the potential to research them. A similar call has been made in learning analytics applications for researchers to capture their pedagogical intents by aligning learning analytics to learning design. In this way, Learning Analytics can help to test the assumptions of learning designs by providing the necessary data, methodologies and tools to support the learning design in lieu of self-reported measures (Lockyer, Heathcote, & Dawson, 2013). In turn, knowledge of the pedagogical context that gives rise to the data is critical to its interpretation. The use of design abstractions can support this alignment as exemplified in the design iterations. In taking a design approach, we specifically focus on a particular set of goal oriented design representations to support learning (Goodyear, 2005). The task designs of the writing intervention are described

[6] Parts of this chapter draw from the following published articles: Knight, Shibani, et al., In submission; Knight, Shibani, et al., 2018; Shibani, 2017b, 2018; Shibani, Knight, et al., 2018a; Shibani et al., 2017, Shibani, Knight, et al., 2018a, Knight, Shibani, et al., In submission, Shibani, In Submission and Shibani et al., In Submission.

using design representations like design patterns, conjecture maps, and workflows, and the rationale of the designs are explained in detail in each iteration.

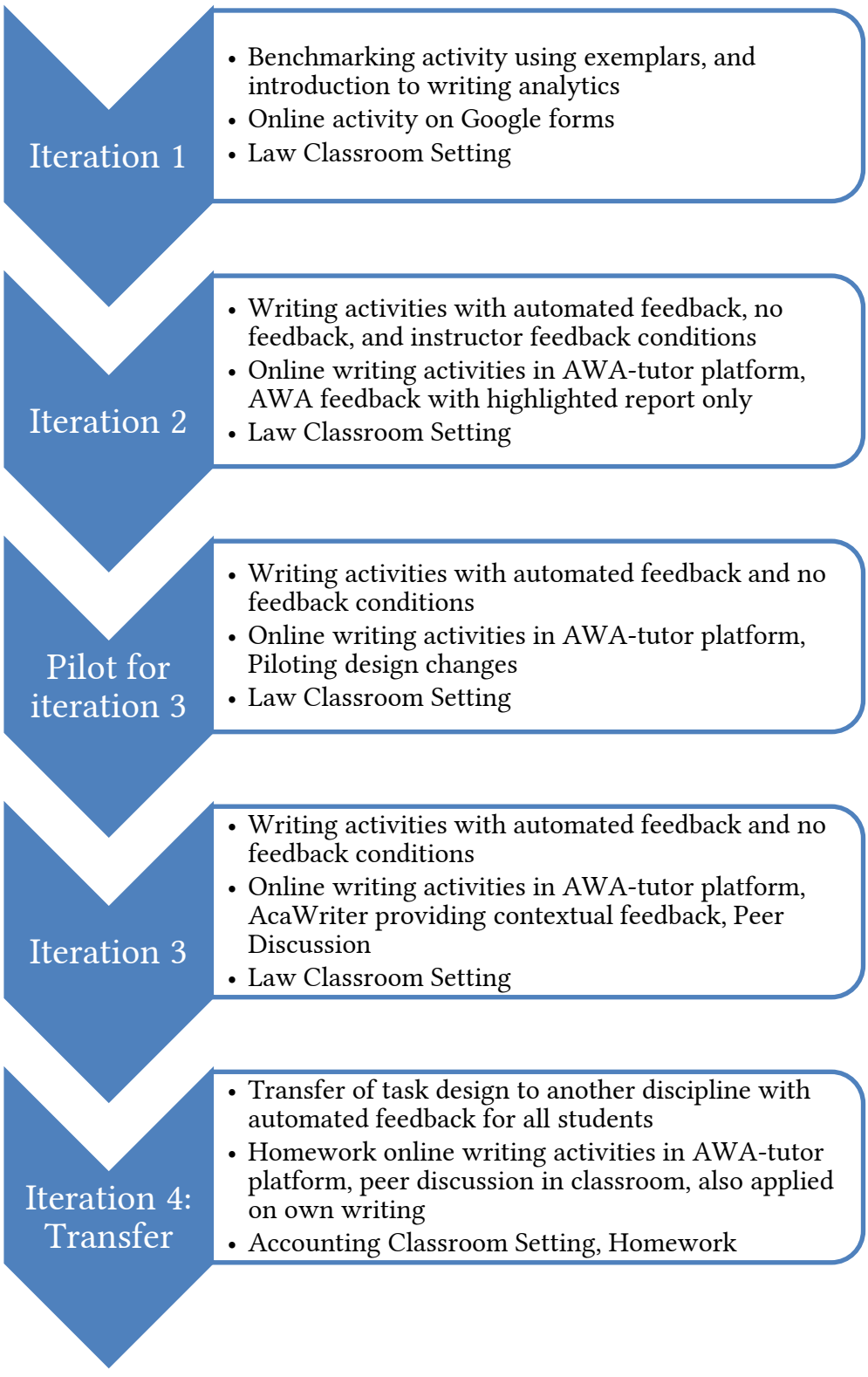


Figure 10: Overview of design iterations in the study

4.1 Design Iteration 1

Iteration 1 was the first implementation that integrated automated writing feedback into a writing benchmarking task (see Section 2.2.4 for definition), trialled with undergraduate Law students. This iteration was implemented prior to the main design iterations for this research, and thus serves as a baseline for subsequent versions of the task design.

4.1.1 Task Design and rationale

The iteration engaged students in a benchmarking activity, in which students assess exemplars, in order to develop their skills of self-assessment and evaluative judgement (Hendry et al., 2012). In this task, all students engaged with exemplars of varying quality and assessed them in order to develop their understanding of the assessment criteria, and learn to apply them on a given piece of writing. Another goal of the task was to introduce writing analytics tools as a source of automated feedback to improve student writing, thus *augmenting* the task design using analytics. Therefore, the exemplar texts that students were given were annotated with feedback, either from Grammarly[7] or AWA[8], or using instructor feedback forming the control group conditions. The three feedback conditions: AWA feedback, Grammarly feedback and Instructor feedback were used to study the impact of feedback (if any) on students' ability to assess the given texts. For instance, the control condition was used to observe whether students who received exemplar texts with AWA feedback were able to assess text quality better than the ones who received exemplars with Grammarly feedback. Students' ability to assess the exemplar texts annotated with feedback was measured based on the difference between the actual score of the text as marked by the instructor versus the students' self-assessed score. The task was facilitated using Google forms, which collected minimal data of students' assessment (students' score for the exemplar texts based on their judgement using the assessment criteria). The task design is shown in Design Representation 1.

[7] Grammarly: https://www.grammarly.com
[8] The tool providing automated feedback on rhetorical moves at this time of the iteration was AWA, which was used prior to the development of its upgraded version AcaWriter

DESIGN 1: Benchmarking and Automated Writing Analytics

Problem: The challenge was to design a task to engage students with exemplars and their assessment, in order that they have an activity that (1) prompts them to critically apply the assessment criteria, (2) prompts them to engage actively with exemplars, (3) provides us as researchers with information regarding their ability to appropriately assess texts.

Task: This initial base task design consisted of a task in which all students were provided with three exemplars of varying quality, and asked to assess those exemplars using the assessment criteria. The application of the assessment criteria involves a mediating process of evaluative judgement in the application of assessment criteria, which in turn should produce the outcome of improved self-assessment ability. The task design was modeled on an existing common practice at the institution.

Tools/materials and participant structures: This task was designed for individual completion, making use of the instructor's rubric, and both high and low quality exemplars. It was facilitated using Google forms in which students logged in with their student emails.

Iterations and Augmentation: The task design was modeled on an existing common practice at the institution. To augment this with writing feedback, in the initial iteration of the task, the exemplar texts provided had been marked up using writing feedback (from either a tool for feedback on rhetorical structures in writing (AWA), or one focusing on spelling and grammar, or from the instructor). This was done with the intent of foregrounding salient features of the texts through the provision of NLP-derived feedback in the form of highlights.

Design Representation 1: Design pattern of iteration 1 task design

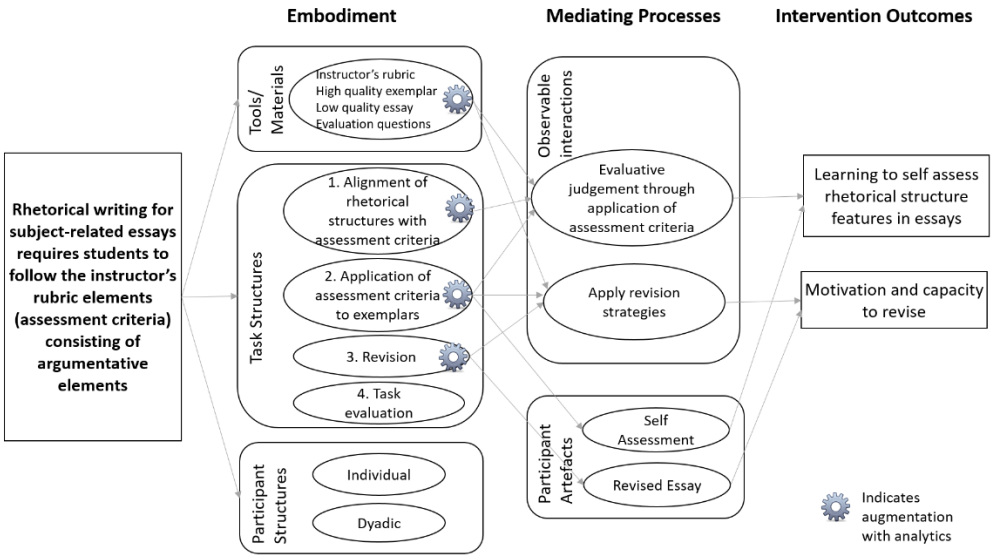
4.1.2 Discussion

Previous analysis of this activity indicated that students appreciated access to the exemplars, and criteria. However, there were no clear differences between the ability of students to appropriately apply the criteria in the group with, or without, the automated feedback. In addition, the mode of interaction with both the exemplars and the automated feedback was rather shallow. Thus, we sought to develop the task to provide opportunity for deeper interaction with feedback and improvements in the self-assessment skills of students. Finally, while the use of Google forms provides for a low-cost and familiar user experience, it limits data capture and presents some technical difficulties for authentication. Therefore, we wanted to improve the platform so as to facilitate other types of writing activities efficiently and collect student data as required without using third-party apps.

4.2 Design Iteration 2

Design iteration 2 was the first main iteration of the study where the tasks in the design, and the online platform to facilitate the activities were established in a standard format. The task design, materials and the learning analytics platform for this iteration were systematically developed by the researchers by *co-designing* them with the instructor of the Law subject in face-to-face discussions. The design was developed to provide both a learning experience for the students, and to facilitate the development and evaluation of learning analytics interventions by the researchers.

The key components of the design are represented in a conjecture map (Sandoval, 2014) shown in Design Representation 2A. The specific sections of the task design which were augmented by analytics are displayed with a 'gear' icon. The box on the left conveys the theoretically principled high-level conjecture, specifically, that in order to induct students into disciplinary practice, they must learn to write using the rhetorical structures that make up argumentative forms. This conjecture is then applied through its *embodiment* within a designed learning environment that uses three main elements: *Tools/Materials, Task Structures and Participant Structures*. The writing instruction is driven by the materials defined by the instructor (in tools/materials) and tasks described in the task structures, listed in the order that students complete them.



Design Representation 2A: Conjecture map of the task design in iteration 2 (Knight, Shibani, et al., 2018)

A key feature of the conjecture map is that it separates the logic of the design from the specific instantiations, and provides a clear perspective on pedagogic sites at

which learning analytics can augment the design towards the conjecture. In this case, a tool called AWA-Tutor (the Academic Writing Analytics Tutor) was developed that guided students through the tasks, and collected data for research, by integrating Writing Analytics in pedagogic contexts (Shibani, 2018; Shibani et al., 2017). This tool provides an example case for the use of abstraction to develop research and implementation understanding of a learning context. A detailed system description of AWA-tutor is provided in section 4.2.2.

Automated writing feedback from AWA embedded in an AWA-Tutor tool allows students to submit their drafts and receive immediate feedback to make further revision in their texts. This enables students to assess their revisions based on the feedback and encourages further revisions upon assessment. This immediate feedback, made possible by analytics, can aid reflection and encourage improvement in student revisions on their drafts, the augmentation it enables is shown in the conjecture map as gears. While the current design is based on an automated tool which provides feedback on rhetorical structures in the text, the design can also be extended for tools that provide feedback on other text features. Because the tasks were developed within an online tool, a separate evaluation question was also built into the structures, as indicated in the conjecture map. The artefacts highlighted in this conjecture map include specifically those artefacts about which learning conjectures were made. Since the artefacts from observable student interactions in the tasks could be useful proxies for students' learning, they are of interest to researchers and practitioners. In the context of our design iterations, the interest in analysis of artefacts is further developed not only in describing how conjectures may be made regarding the trace obtained through use of an online tool designed with pedagogic principles in mind, but also how learning analytics may augment these task designs. The next section provides an explanation of the task elements in more detail.

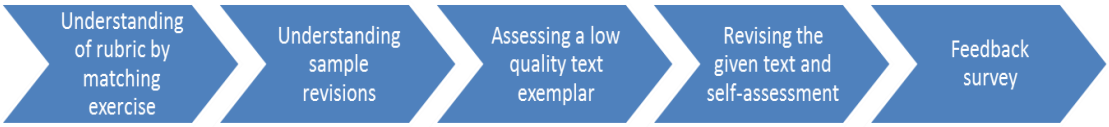
4.2.1 Task Design and rationale

This pedagogic intervention was designed for students to improve their ability to evaluate the quality of writing and revisions and improve a draft based on feedback/ self-assessment. Study conditions were designed to address a key concern of the research questions: understanding the efficacy of different feedback types for student revision, in order to identify helpful feedback for students. Students were randomly pre-

assigned to one of the three groups by the instructor based on the feedback they would receive in the revision task as follows:

1) AWA Feedback Group – These students received feedback on request from the AWA tool. They could request feedback on their revised text as many times as required. They watched a short video on how to use the tool at the start of the activity.
2) Instructor Feedback Group – Students from this group saw an instructor-highlighted PDF file with static feedback on the parts that need improvement in the given essay.
3) No Feedback Group – This set of students received no feedback on the essay to make improvements. They worked on the text based only on their assessment of it.

The sequence in Design Representation 2B shows a simplified workflow of tasks that all students carried out in the intervention (see below for explanation of the tasks and the differences among between control groups).



Design Representation 2B: Workflow of tasks in design iteration 2

The design pattern for this task design explaining the theoretical background behind the design changes and augmentation by analytics can be found in Design Representation 2C.

DESIGN 2: Benchmarking, Text-Revision, and Automated Writing Analytics

Problem: We wanted students to critically consider how specific features in the text instantiate responses to the assessment criteria, and to develop the student's interaction with the application of the criteria for building their understanding of how to – practically – improve a text.

Task: The initial task (task 2) was amended, and an additional task was added (task 1). In task 1, students were asked to match excerpts from a text to the criteria that they addressed (for example, a sentence providing background information aligns with the criterion "Identification of relevant issues", while a sentence providing evaluation or analysis of a claim or piece of evidence aligns with the criterion "Critical analysis, evaluation, original insight". The revised task 2 involved students assessing a single exemplar text using the assessment criteria, and being specifically asked how they would suggest improving the text. In task 3, then, the students were asked to edit the text they were provided with, and (task 4) to evaluate the improvements that they had made (i.e., to provide a new assessment of the quality of the text). Following task 4 the students were provided with their own text revisions, and those of an instructor on the same text, providing a 'good' exemplar to demonstrate the improvements made. While the original task (above) was intended to produce a mediating process of evaluative judgement, the revision task was – in addition – designed to produce a mediating process of revision strategy application, to produce the outcome of increased capacity and motivation to revise, and improved self-assessment ability. The first task was specifically designed to develop evaluative judgement through understanding of the assessment criteria, and thus to improve self-assessment through understanding of rhetorical structures.

Tools/materials and participant structures: As in design 1, this task was designed for individual completion, making use of the instructor's rubric, and in task 2 a lower quality exemplar, with task 4 providing the higher quality comparator. The instructor's rubric and the lower-quality exemplar drive the first and second-to-fourth tasks from the task structures list respectively. The online activities were facilitated using the AWA-Tutor platform.

Iterations and Augmentation: This task design developed from that described in design 1. As in that case, a between-subjects design was used to provide some students with instructor-based (static) feedback, others with dynamic feedback from AWA, and others with no feedback. Prior work has been conducted to establish conceptual relations between the instructor's criteria, rhetorical structures, and their specific instantiation in AWA (Knight, Buckingham Shum, Ryan, Sándor, & Wang, 2017). These relationships were foregrounded to the AWA group through static highlights flagging the AWA moves on the sentences to be aligned with the criteria. Then, the revision task was also augmented by AWA, with feedback provided on-request (via a button) to students as they revised the

Design Representation 2C: Design pattern of the task design in iteration 2

The pedagogical design was supported by a learning analytics platform called AWA-Tutor (explained in Section 4.2.2) that facilitated the online activity for students in all groups by stepping them through a series of several subtasks. Students worked individually on the activity using their own laptops. They entered the system by accessing the platform from the web. The URL for the activity was supplied from the LMS as part of their weekly lesson. Student details were pre-stored in the database and each student was directed to a specific group's tasks upon login. The individual tasks in the design are explained along with their pedagogical reasoning below:

4.2.1.1 *Rubric Understanding*

In our learning context, although students are already aware of the instructor's marking rubric, they may not know how to apply the rubric, and how particular rubric facets are related to linguistic features that automated tools might help them identify. Therefore, the first task was a matching exercise where students were asked to identify sample sentences from an essay that would match elements of the instructor's marking rubric in an interactive drag and drop interface. This engagement with exemplars is an effective method for students to understand the assessment criteria (Hendry et al., 2012; Rust et al., 2003). The task thus supports understanding the different rhetorical markers from sentences that would be useful to signal to the reader the important components of their essay with respect to the rubric. A screenshot from task 1, the matching exercise is shown in Figure 11 where green indicates correctly matched elements and red indicates wrongly matched elements[9]. Students were required to match all instances correctly before moving to the next task. In the AWA feedback group, students also saw the corresponding AWA tags for the exemplar sentences to support understanding how the tags were related to the rubric (refer Figure 12).

4.2.1.2 *Learning Revision Strategies*

In the second task, students viewed a sample essay which was revised by the instructor to give them an idea of the kind of features to focus on and how revisions to an existing text might be used to improve that text towards the rubric. This helped them learn strategies for revision, and understand the task requirement. A screenshot showing sample comments from the instructor teaching revision strategies is shown in Figure

[9] A screencast video of how students interact with the sub-tasks in AWA-Tutor is demoed in https://www.youtube.com/watch?v=K212XabCL5w, and the platform detailed in Section 4.2.2

13. The process of revision could otherwise include anything from making surface level changes like spelling and grammar to modifying the content of the topic. For this particular revision task, students were encouraged to focus on rhetorical structures in the text that could be improved.

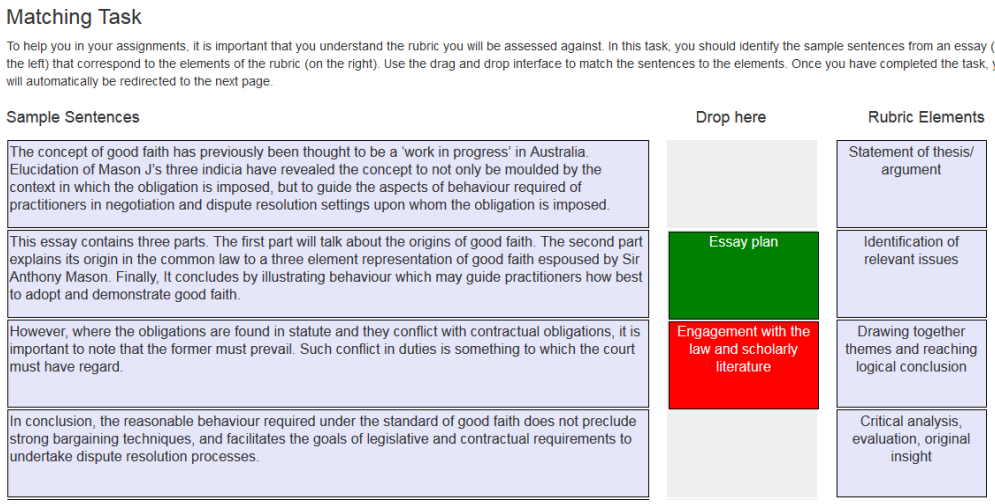


Figure 11: Screenshot of the matching exercise task for feedback groups not receiving AWA feedback

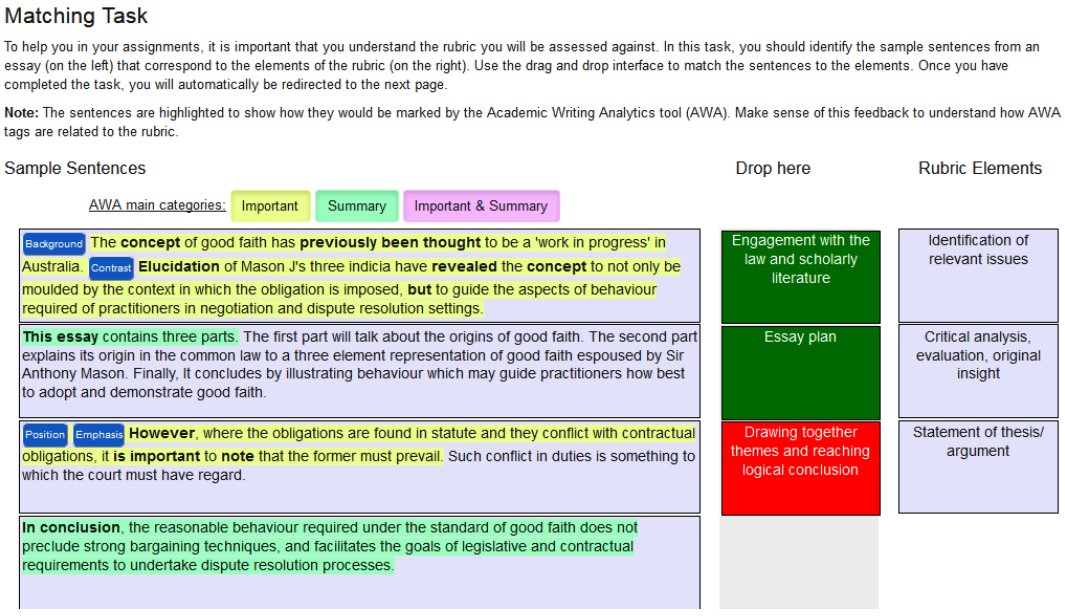


Figure 12: Screenshot of the matching exercise task for AWA feedback group

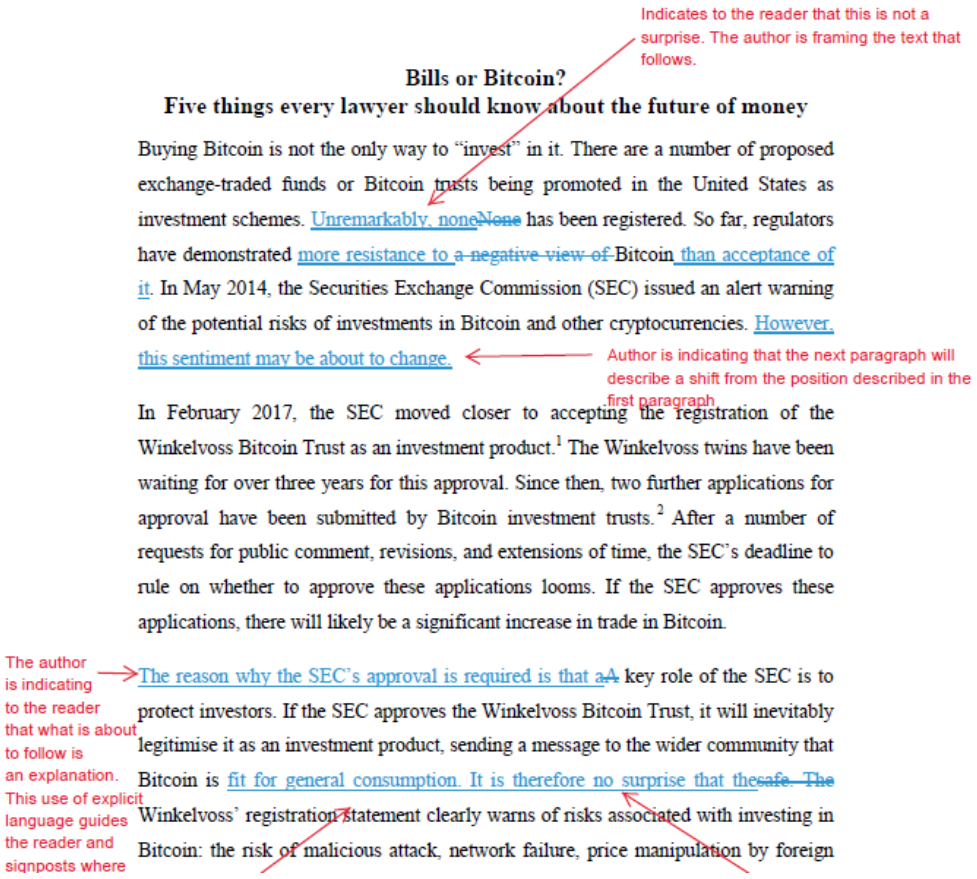


Figure 13: Screenshot showing sample comments made by the instructor to teach the revision strategy

4.2.1.3 *Essay Assessment*

The third task consisted of an essay assessment in which a low quality essay exemplar pdf was provided for reading, followed by questions to evaluate its quality (see Figure 14). Students provided an assessment of the essay's quality, which, in the fourth stage, they then worked on to revise. This task was designed for students to acquire evaluative expertise by transitioning from feedback to self-monitoring (Sadler, 1989), with the self-assessment intended to enhance students' capacity to make judgements and self-regulate their work for sustainable learning (Boud et al., 2015).

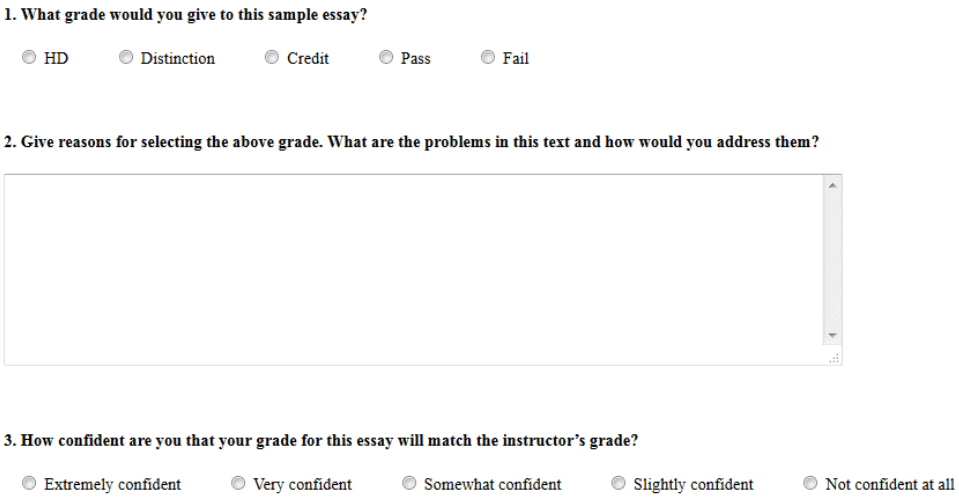


Figure 14: Screenshot showing sample questions asked for essay assessment

4.2.1.4 *Essay Revision*

Students fourthly, worked to act on the issues that they identified in the text, with encouragement to engage in the kinds of revisions cycles that support learning to write (MacArthur, 2007). In this task, students received different types of feedback on the essay to make revisions based on their group (AWA Feedback Group, Instructor Feedback Group and No Feedback Group). To facilitate the use of feedback, it was provided in a frame to the right of the editor frame in which they revised their text. The revision task interface for Instructor Feedback Group is shown in Figure 15. For the AWA feedback group, the frame on the right contained feedback on the editor text from the AWA tool as shown in Figure 16. The No feedback group was provided with the text editor only to make revisions. After revising the essay, students completed a self-assessment on the revisions made by answering questions, to reflect on the improvements they made in the essay.

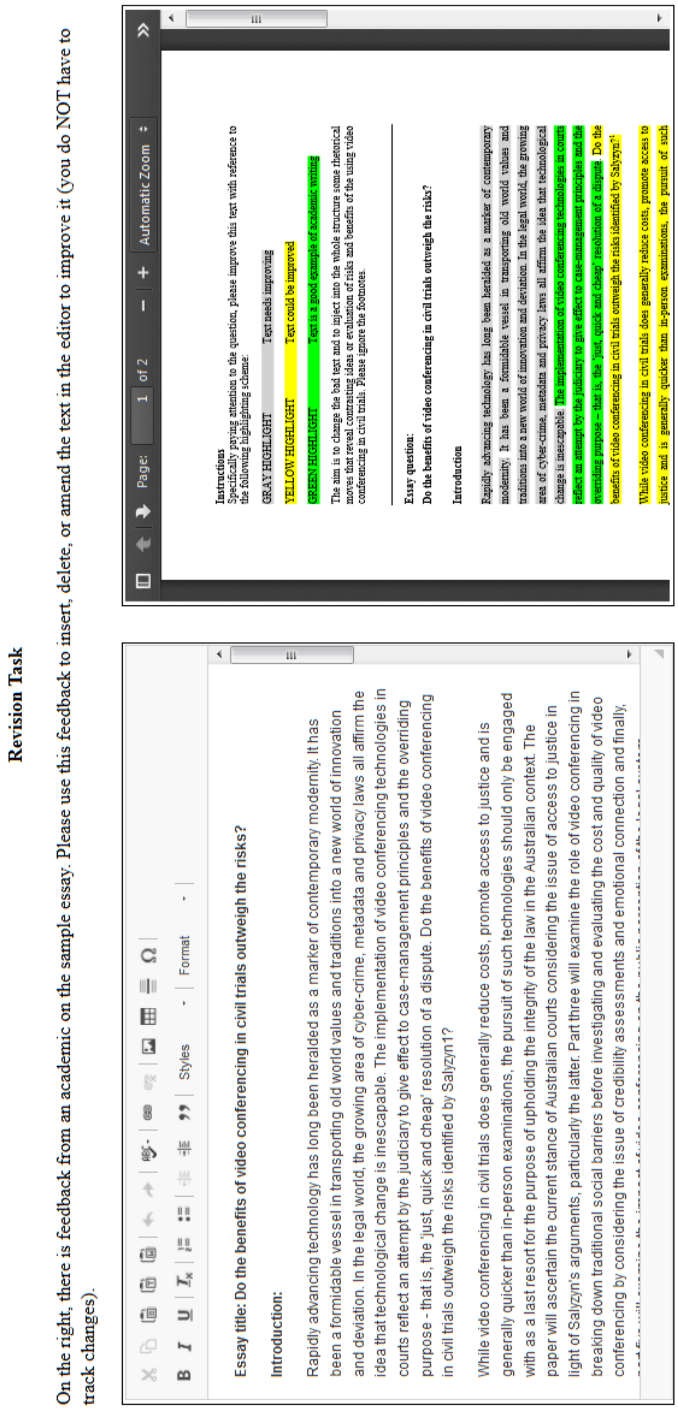


Figure 15: Sample screenshot from the revision task for Instructor feedback group

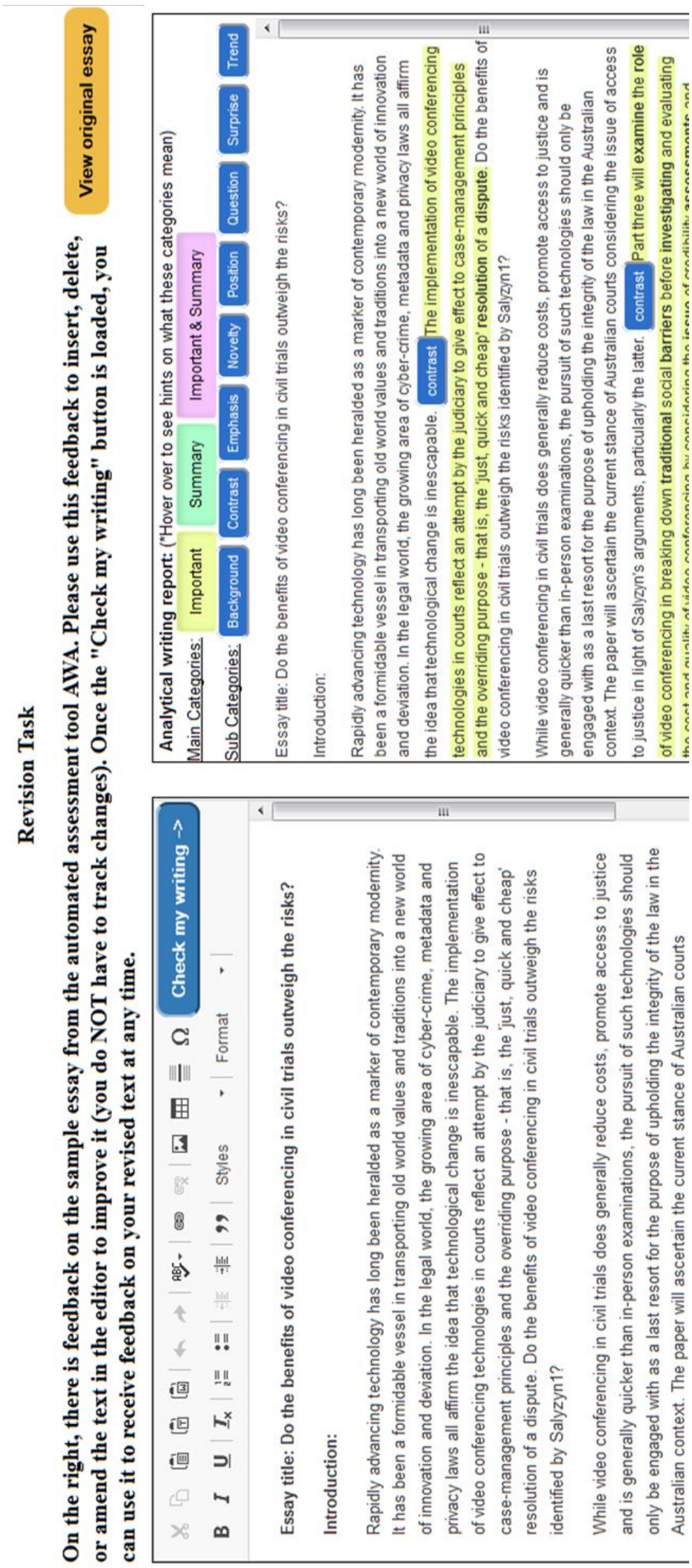


Figure 16: Sample screenshot from the revision task for AWA feedback group

For this revision task, students can be provided with different types of feedback as required by the pedagogical design. In the current setting, there was a static feedback

by the instructor in the form of a pdf which gave comments on the given essay by highlighting the key parts that need improvement, and a dynamic feedback on rhetorical moves from the automated analysis tool AWA that can be accessed to get feedback at any particular version of the essay as required.

4.2.1.5 *Feedback Survey*

In the final part, students provided feedback on the task and the feedback obtained for revision by answering a few questions (screenshot in Figure 17). This was to help researchers evaluate the effectiveness of the intervention and to study the impact of automated feedback from the perspective of the students. At the end of the activity, students were given an option to download a version of their revised essay and a sample revised essay by their instructor for future reference. This was to help them reflect on the improvements they made in the essay by comparing with an exemplar improved essay from the instructor.

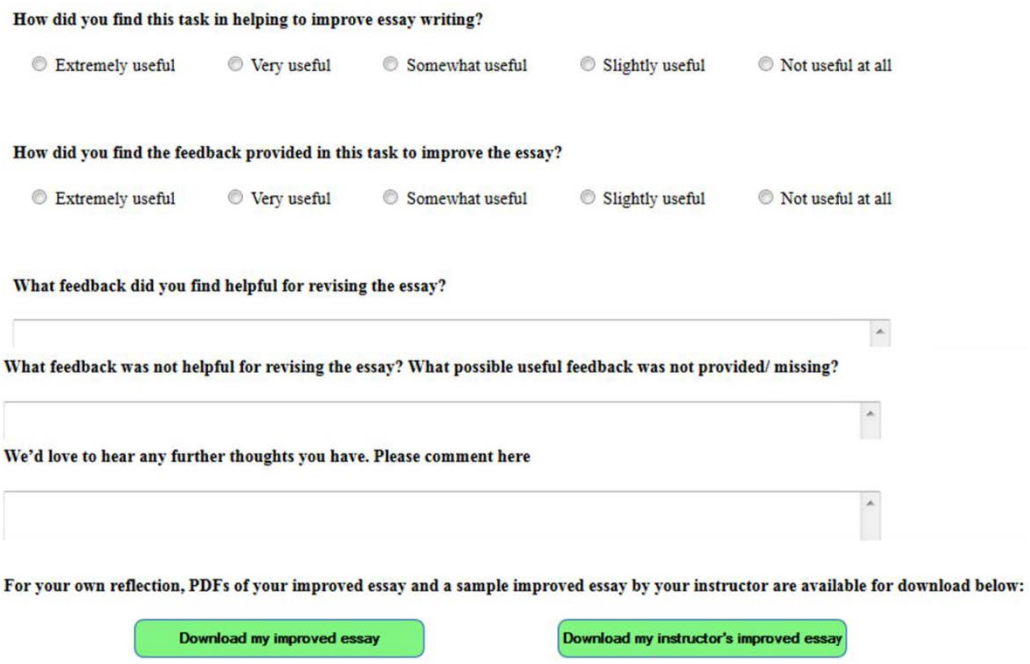


Figure 17: Feedback Questions for the task

4.2.2 Technical Infrastructure – AWA-Tutor platform

The technical infrastructure for this pedagogic implementation was provided by a writing analytics tool called AWA-Tutor, which acted as a platform to ground automated writing feedback in robust learning design. Increasingly, the importance of aligning learning analytics with learning design is being understood as a way to uphold

its core aim of improving educational practices, while also collecting meaningful data about learner's activities that can be interpreted in context (Lockyer et al., 2013). In light of this, the web-based writing analytics tool "AWA-Tutor" (Shibani, 2018) was developed to facilitate this writing intervention by integrating analytics with pedagogy[10].

AWA-Tutor extended the Academic Writing Analytics tool AWA, by scaffolding the entire writing intervention described earlier. Students were guided through a series of tasks through the platform, such as understanding the instructor's rubric, improving a sample text, reviewing exemplar improvements, self-assessing their work, and reflecting on the quality of the automated feedback. The tool was designed in a modular fashion to support the learning design of an instructor, who can select the task components to be included, and personalize the feedback experience for different students. It was implemented using PHP, Javascript and a MySQL database. The platform architecture is provided in Figure 18.

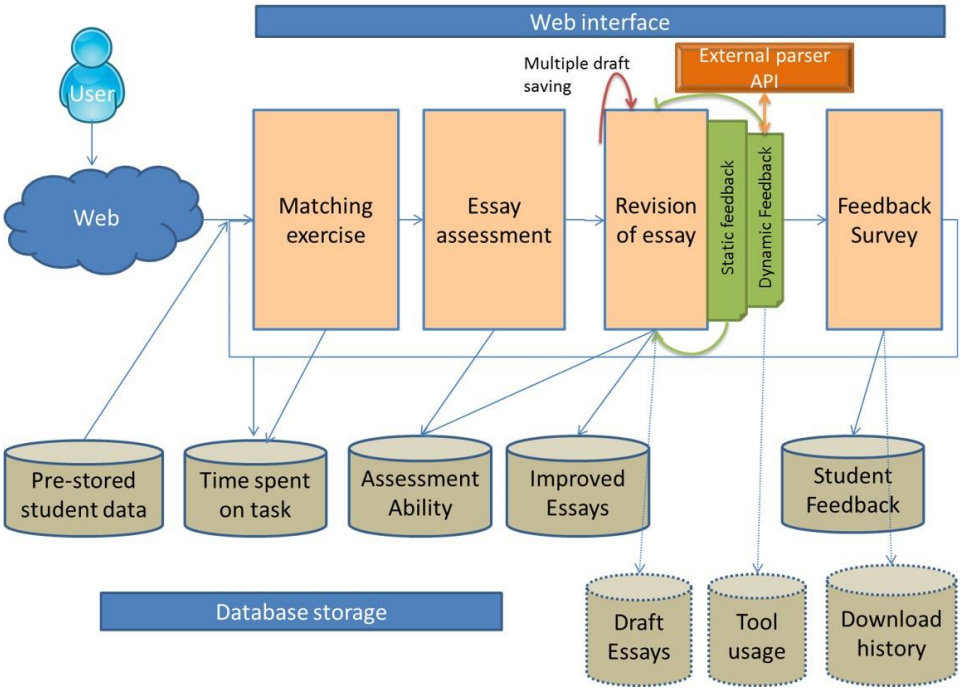


Figure 18: AWA-Tutor platform architecture

[10] A short demonstration video is available: https://www.youtube.com/watch?v=K212XabCL5w

AWA-Tutor also captured detailed activity traces: the time taken by students to complete certain tasks of the activity, snapshots of drafts at customizable time intervals, students' requests for automated feedback and the feedback received, and feedback survey responses. These traces of student activity were stored in database tables by different components of the web interface as students use the platform. The time spent for the whole activity to complete all tasks from the start to the end was recorded for all students. The different components of the web interface, and the types of student activity data stored for research purposes are explained in the following section.

The matching exercise in AWA-Tutor was implemented using a customized DHTML drag and drop quiz script[11] . From this exercise, the duration taken by students to complete the task by matching all elements correctly was stored. The essay assessment component stored students' assessment data on the given essay in the form of grades, confidence level that it would match an instructor's grading, and qualitative comments on the problems identified in the essay and recommended improvements. A basic document editor from CKEditor was used for the revision task[12] that helped preserve formatting of text which would be lost in a normal text box. A tutorial explaining the use of CKEeditor in PHP has been published as a blog post (Shibani, 2017a). From the provided text editor, data was stored in specific intervals (every one minute) to capture students' drafting process. This process of drafting and revising, which were previously hard to study in authentic writing scenarios, are now reconstructable by tracking those analytics.

As defined in the control conditions described earlier, students in the AWA group could get feedback from the tool on any of their drafts as needed. Text cleaning and formatting were performed in PHP to provide live feedback on the text, or post-task processing. The analytical parser engine that provides feedback on the text was accessed from a version of Xerox Incremental Parser[13] by connecting to an external API (Note: As of 2018, the Xerox parser is depreciated, and has been replaced by an open sourced TAP API, hosted by CIC). Data from the editor was cleaned to send data and receive AWA feedback from the parser on rhetorical moves. The final improved essay was stored in the database for all students. The dotted database tables in Figure 18 show that

[11] http://www.dhtmlgoodies.com/index.html?whichScript=drag-drop-quiz
[12] http://ckeditor.com/
[13] https://open.xerox.com/Services/XIPParser

data that would be stored differently for different students based on their usage behaviour. The number of draft essays stored depends on the time they spent working on the task – higher time spent stores more drafts. AWA use also varies from student to student – ranging from students who made few requests for feedback to students who requested feedback many times.

The feedback survey stored responses from students on the questions asked about the usefulness of task and feedback. The option to download a version of their own revised essay in the last activity was made possible by dynamically generating a pdf of their revision from the previous page using FPDF (http://www.fpdf.org/). Details of students who downloaded their own improved essay and instructor's sample revised essay from the last page for reflection were recorded in the database as download history by tracking the clicks from the respective links.

4.2.3 Data Analysis

Development of the learning analytics platform and the pedagogical design described above facilitated capturing student trace data (explained in Section 4.2.2) that was analysed to provide insights into their learning, and the impact of feedback on that learning. Data from this iteration was analysed to partially answer research questions RQ1a and RQ1b on students' perception of the writing intervention, and the impact of automated feedback on revisions using mixed methods. New automated and manual methods to study student revisions were trialled using the analytics collected in this iteration. Student activity data stored in the database tables were downloaded as csv files, which were then imported into RStudio for data analysis in R.

4.2.4 RQ1a) Student perceptions of the writing intervention

Students rated the perceived usefulness of this activity in order to improve their essay writing in a scale of 1-5 (where 1 = not at all useful, 2 = slightly useful, 3 = somewhat useful, 4 = very useful and 5 = extremely useful). They further provided qualitative comments on what feedback was found to be useful, what feedback was not useful, and any other additional comments about the whole activity. Even though the activity was carried out in a tutorial session in class, not all enrolled students completed the activity since it was not a mandatory requirement. For the purpose of this analysis, only the complete dataset of 201 students who finished all parts of the activity is considered for analysis. This consisted data of 91 students from AWA group (awa), 71 from Instructor group (ins) and 39 from No feedback (none) group. This data was analysed to study

students' feedback on the perceived usefulness of the activity and on the provided feedback for answering RQ1a. The perceived usefulness of the activity across the three groups is shown in Figure 19 using a box plot[14].

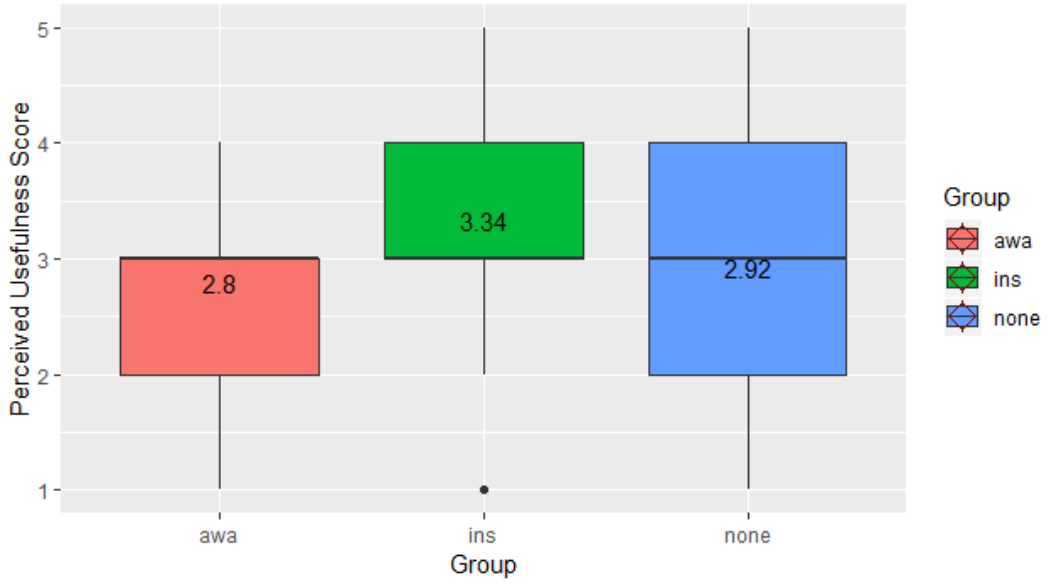


Figure 19: Perceived usefulness of the activity across comparison groups in iteration 2

The Instructor feedback group found the activity to be most useful (M = 3.34, SD = 0.84), followed by No feedback group (M = 2.92, SD = 1.04) and AWA feedback group (M = 2.80, SD = 0.75). A one way analysis of variance showed that the effect of groups on the usefulness score was significant, $F(2,198) = 8.32$, $p < .001$. The assumption of homogeneity of variance was satisfied with Levene's test (p =.12). Post hoc analyses using Tukey's HSD indicated that the usefulness rating of Instructor feedback group students was higher than the usefulness rating of AWA group students ($p < .001$), and no significant difference was noted amongst the other groups.

The qualitative comments of students were explored to further understand student perceptions of this activity. The No feedback group provides a baseline group,

[14] 'Box-and-whisker-plot' commonly called 'box plot' provides a visual representation of the five-number-summary: the minimum value, the lower quartile, the median, the upper quartile, and the maximum value (Watson, 2012), and is used widely in this thesis. The numbers displayed in the boxes denote the mean values, and are not to be confused with median values, which are denoted by bold lines. Possible outliers are represented by dots extending from the whiskers. Note that the 'awa' group in the figures represents the group of students receiving automated feedback from the tool AWA or AcaWriter.

as these students evaluated the usefulness of the broad pedagogic activity with no additional feedback component.

4.2.4.1 *Useful task elements*

Across the groups, students found several sections of the task useful in improving their essay writing. A number of students explicitly mentioned that it was useful to have the initial sample text, on which the instructor had modelled the kinds of revisions that could be made to improve a draft, saying things like:

"The annotated sample with comments was helpful in revising the essay, as it gave examples of what was done well and done poorly." Respondent 123, Instructor feedback group

"The exemplar answers were very helpful in highlighting the areas of the essay which needed improving, which students may initially overlook. The highlighting of different sentences is also useful in indicating what components of the writing were critical and what sections were maybe unnecessary identification or description. The sentences crossed out and rewritten were especially useful for proposing alternative ways of writing a sentence in an improved manner" Respondent 82, AWA feedback group

"It was useful to see how a simple change (like swapping one paragraph for another) can make an essay a lot clearer and relevant to the topic at hand. I will make sure I apply this kind of task to my own essays - asking myself if my essay would be clear to a pair of fresh eyes." Respondent 180, No feedback group

Students also appreciated having access to both their own text to download, and a sample revision of the same text they had edited which the instructor had marked up with improvements, saying for example:

"A very good exercise! Glad we can download both our improved version and the instructor's improved version. Hopefully this will be a way we can get feedback on the feedback we provided in our edit." Respondent 145, Instructor feedback group

"The provision of the instructors improved essay provides a useful benchmarking tool to compare my changes against the changes made by the instructor" Respondent 109, Instructor feedback group

4.2.4.2 *Evaluating design decisions*

Some students felt that it would have been useful to have some readings beforehand on the topic to have a better idea on the essay they work with.

"I really love the idea behind this exercise. I think it would be more beneficial to complete if we had to do some prior reading - for example of the Salyzyn essay that was referred to in the paper so that we could have some context. I find it difficult to write or revise something without having a background in the area." Respondent 179, No feedback group

"It didn't make sense to be asked to revise an essay on a topic we haven't really studied ... because in terms of content, I'm not sure how to improve it" Respondent 28, AWA feedback group

This could be incorporated if a revision task is designed in the future for modifying the content of the essay as well. The focus of the current task was on the rhetorical structure improvements and hence there was no emphasis on the content.

Some students were unsure of the usefulness of the revision task which required them to work on essays written by others. On the other hand, others found it useful to apply their critical lens to an essay written by someone else, as it would eventually help them look critically at their own essays:

"Everyone has their own unique styles that should translate on to a page of work that is of their own design. Not sure if people learn from someone else's mistakes at the very end of a Uni day." Respondent 167, No feedback group

"I believe having to personally assess an essay forces you to critically engage to a greater extent than one may have to. From looking at an essay from a marker's perspective one can take a step back and understand the little details that a marker is looking for. I also feel that by assessing someone else's work it provides you with better skills to assess your own work from a more neutral perspective. Self-reflection and editing are a key aspect of writing a poignant and quality academic essay that accurately engages with the criteria." Respondent 32, No feedback group198

The allocation of time was not incorporated in the platform, but was provided as a run sheet with an approximate time division for the sub-tasks. This led to some incomplete submissions as some students stayed in the first few tasks longer than expected, not allowing them enough time for the rest of the activity. The activity could

be re-designed in the future to incorporate time allocation for sub-tasks in the platform for smooth task completion, which was indeed incorporated in later iterations (see next design iteration in Section 4.3.1). It could also be built as an out-of-class activity where students can engage with the tasks in their own time out of class, independent of the pace of other students.

"The activity was engaging but it would be more enjoyable if it was clear of the time allocation for the tasks and the number of tasks involved" Respondent 204, No feedback

"I think it would have been useful to know how much time I had...I feel as though I rushed myself with the editing and therefore didn't do my best work" Respondent 109, Instructor feedback group

"A bit more time in amending the essay would have been greatly helpful." Respondent 135, Instructor feedback group

4.2.4.3 *Evaluating the provision of meaningful feedback to students*

In terms of the feedback provided to improve their essay, students from the Instructor feedback group felt that more explanations were necessary to understand the changes that they needed to make in their essays. They wanted to learn how to resolve the identified problem by receiving suggestions for improvement. Such direct recommendations would help students solve the current problems in hand and aid them in resolving similar problems in the future. Similar comments on the given feedback were also seen in the AWA group. They wished to receive more direct feedback in the form of corrective advice on what to improve, rather than highlighting the key sentences.

"Rather than just highlighting the text, I think it would have been worthwhile to have colour coded explanations of why each section was highlighted... Sometimes I was unsure about why a piece of text was highlighted so I wasn't sure how to make a change." Respondent 109, Instructor feedback group

"Didn't give many alternatives as to how the phrasing could be improved" Respondent 143, Instructor feedback group

"The highlighting only alerted to me what was good. However, there should be highlight to alert me to problems in the essay as well. the highlighting only showed me what was a 'summary' etc. There should be more categories and types of feedback such as grammar issues, sentence structure." Respondent 11, AWA feedback group

"It could offer an alternative or some tips regarding essay writing so a student who has seen where they go wrong can understand how to amend the essay they have written." Respondent 95, AWA feedback group

"I found the feedback unhelpful as I couldn't distinguish which parts needed fixing, even though it was stated as "important". Couldn't understand what 'important' meant in this context - important to fix or important as in it was a good part of the essay that didn't need to be fixed?" Respondent 47, AWA feedback group

Student feedback from this task could be used to find ways in which the tool can be tweaked to provide better feedback. However, few of the suggestions students made with regard to the feedback received provide direct actionable alterations for that feedback. It is crucial for students to recognise the intended usage of the tool and how to use it best to help in their writing context. The scope of the tool must be explained clearly in terms of what it can and cannot do. For example, some students noted that they would have liked grammar and spelling feedback, but this is not a feature targeted by the task or the tool deployed in this research. Students noting that the focus of the tool is on rhetorical markers of academic writing should set expectations of the tool to students in order to effectively use the tool. Further guidance regarding the use of the AWA tool, in the form of examples of use, and a user-guide, would also support this effective use. Directing students to lessons where they can read more about effective writing practices would also be beneficial.

4.2.4.4 *Studying additional downloads as a proxy for student motivation*

To analyse the motivation of students who were engaged in the activity, a proxy measure was studied based on additional files that students downloaded for further reflection. This was made possible by AWA-Tutor which tracked students' clicks on the buttons which let students download files for further reading and reflection in the feedback page. Students were provided an option to download the two files below as part of the activity, and their click logs were stored:

1. Ins-File: This was a sample revised essay which was improved by the instructor based on the given essay. It was provided as an additional exemplar for students to refer to, so they can learn how to revise better.
2. Self-file: This was their own revised essay, which was dynamically generated as a pdf from the text editor in the Revision task (previous page).

It was provided for students to reflect on their own revisions made, and compare it with that of the instructor's changes in the exemplar for further understanding.

Table 4 shows the count of students in each group (n) along with the count of students who downloaded the two additional files. Out of the 91 students in awa group, about 58% of students downloaded the instructor file, and 32% downloaded their own file. In the instructor group containing 71 students, a higher number (66%) of students downloaded the instructor file, but only 28% of the students downloaded their own files. In the No feedback group of 39 students, 46% downloaded the instructor's file and 23% their own file. Fisher's exact test for count data did not find any association between the group and the download categories (p =.78), providing no evidence for differences between the groups.

Table 4: Count of students who downloaded additional files across groups

	n	Ins File	Self File
AWA Feedback Group	91	53	29
Instructor Feedback group	71	47	20
No Feedback Group	39	18	9

These findings indicate that the students did not fully engage with all parts of the task, especially as it was an optional activity. This might be because of their lack of motivation to go beyond the mandatory tasks to improve their writing skills, or failure to read the instructions carefully. While there was no significant difference across the three comparison groups, there seems to be a trend across all the groups in how students perceive their own file in comparison with the instructor's file (blue and green bars in the figure). Almost half of the students who downloaded the instructor's file did not download their own file, in spite of both being presented next to each other. This suggests an interesting notion that students appreciate expert suggestions to further reflect on improvements, but do not want to have a copy of their own writing for self-assessment. A reason might be that students only made minor revisions which they still remembered on the top of their heads to compare with instructor-suggested changes, or they did not make good changes to the essay to keep a reference of.

4.2.4.5 *Comfort in receiving automated feedback*

In addition to the measures studied above, students' self-rating on how comfortable they were in receiving automated feedback on their writing were explored. Some students were not comfortable in receiving automated feedback and felt that a tool cannot provide context-sensitive feedback like a human. This is a known problem with the incorporation of such tools. Students should be made aware that automated tools are not a replacement for instructors/ tutors, but rather a support mechanism they can use to get additional feedback when required. As discussed earlier, making students understand the scope of the tool would help them put the tool to appropriate use it was designed for. This will help them know the context of using a machine versus a human for the desired feedback.

"I don't feel as comfortable with an online tool. I think I would feel more comfortable with a human providing feedback" Respondent 21, AWA feedback group

"An automated program would not be able to tell me what points I am missing information-wise like a lecturer or tutor would be able to" Respondent 25, AWA feedback group

Students who received automated feedback from AWA also rated their level of comfort (1-5, where 1 = not at all comfortable and 5 = extremely comfortable) in receiving feedback from a tool. Hence, a subset of the dataset containing only the AWA group students' rating was analysed to assess the relationship between the comfort level of using automated feedback and their perceived usefulness score. Students were generally not very comfortable in receiving feedback from a tool (M = 2.77, SD = 1.07). Students' comfort level in receiving automated feedback was also found to be positively correlated to their perceived usefulness of the activity, Pearson's $r(91) = .44$, $p <. 001$. This could provide a possible explanation for the low usefulness score of the AWA feedback group. The low usefulness score of AWA group students in this activity is in contrast to the findings from the previous study which reported that students found AWA to provoke useful reflection regarding their essay writing. Students seemed to have judged the usefulness of the current activity in terms of their tool usage rather than the wider pedagogic design of the tasks targeting improving their essay writing. Thus, the expectations for different types of feedback by the tool and their comfort level in receiving automated feedback could have contributed to the usefulness score for this group.

4.2.5 RQ1b) Impact of automated feedback on student revisions

In addition to studying student perceptions of the intervention, a suite of techniques were trialled to study the actual impact on student writing in the form of revisions made. These methods can further provide a triangulated view for evaluating the impact of automated feedback on student writing as discussed in Section 2.5. The product and processes involved in the revision process can be studied with text analysis and visualization techniques; these methods are exemplified using data from this iteration.

4.2.5.1 *Data context*

The data used in this section comes from the main revision task in which students engaged. In this task, students worked on revising a short essay extract provided to them (original essay), to produce an improved version, with a focus on improving the rhetorical structure of the text, in study conditions with and without automated writing feedback. This original essay serves as the baseline to study individual improvements made by students. The advantage of this method over using students' own writing is that it eliminates the need for additional marking of students initial drafts (since the provided original essay is a standard base text of known quality). The original essay given to students for revision was on the legal topic of using video conferencing in civil trials, consisting of 550 words with four small paragraphs and a reference list. Drafts from students' revisions were captured every minute using the AWA-Tutor tool which scaffolds the tasks in the intervention, as detailed above. Students' usage of automated feedback, and its output every time the student requested feedback, were captured to study the impact of automated feedback as well. This data will be used in the second part of analysis that studies the process of revising drafts.

4.2.5.2 *Scores analysis*

The most commonly used method by instructors to analyse the quality of writing is grades or scores for the written text. Score is regarded a valid measure to assess a text's quality because it involves an expert marking the text by making judgements based on a standard rubric. While there might be variation between markers that affects the quality of the grading done, the idiosyncrasies of the marker did not affect the scores in this case since all essays were scored by one expert tutor. Hence, this score acted as the gold standard human assessment to validate other measures that were auto generated using textual features.

A subset of the revised essays containing 123 essays (46 from AWA feedback group, 40 from instructor group, and 37 from No feedback group) were sent for marking to a subject tutor. The marker was blinded to which group the essays were from. The revised essays were marked by the tutor on a scale of 0-3 (0 - degraded, 1 - no change, 2 - minor improvement, 3 - major improvements) based on the essay assessment's seven rubric elements: Statement of Argument, Statement of Essay Plan, Identification of issues, Analysis, Sustained thesis, Original insight, Engages with literature/cases. A combined overall score was calculated by summing the scores of all the assessment elements' scores as an indicator of quality. If no changes were made, the minimum score a student could possibly get for a revised essay was 7, and lower if they degraded some aspects of the essay. The maximum possible score was 21 if they made major improvements, scoring 3 in all parts of the criteria. The overall scores of essays revised by students across the three groups are shown in Figure 20.

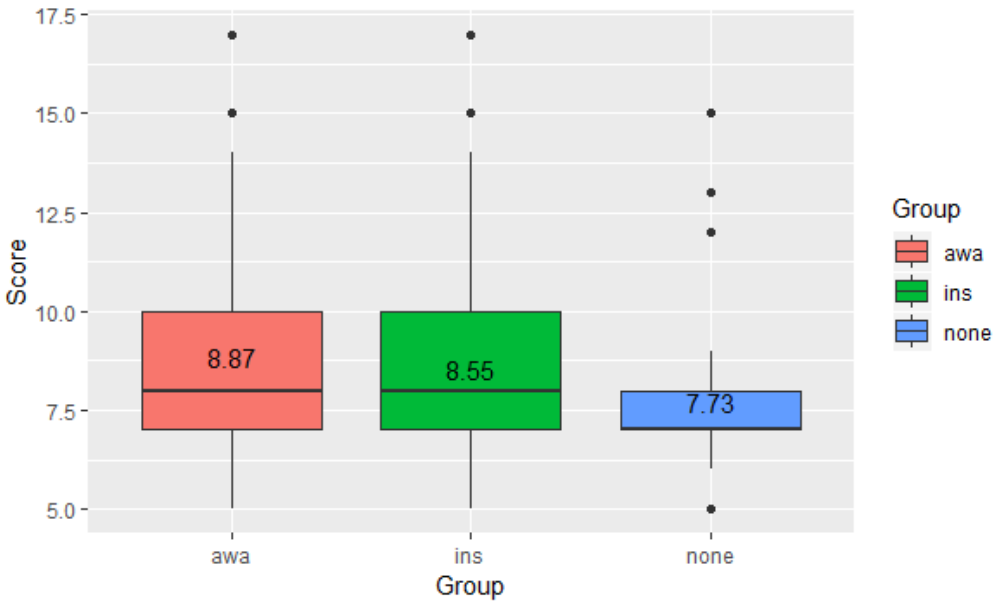


Figure 20: Scores of students' improved essays across comparison groups in iteration 2

The AWA feedback group scored the highest for improvements in the revised essay on average (M = 8.87, SD = 2.52), followed by the Instructor feedback group (M = 8.55, SD = 2.92) and the No feedback group (M = 7.73, SD = 1.95). However, ANOVA showed that the effect of groups on the score was not significant, $F(2,120) = 2.19$, p = .12. The high standard deviation numbers indicate that that scores were spread out from the mean more for the Instructor feedback group and the AWA feedback group, than for the No feedback group. This suggests that even students within the same

control group (Instructor feedback group for example) varied widely in terms of the quality of revisions made to the given essay.

From the scores generated above, bands were created to further characterise students based on their essay revision quality/ performance as *Degraded, Neutral, or Improved* based on individual assessment criteria elements. If the students made improvements in one or more of the assessment criteria, they were categorised as belonging to the 'Improved' band. If they did not fall under the above band, and reduced the quality by making irrelevant changes which further degraded the quality of the given essay in any criteria, they were categorised as 'Degraded'. Students who did not fall into either of these categories, i.e, those who did not make any changes to the essay were classified as 'Neutral'.

The proportion of students in each group categorised by this revision quality indicator is shown in Figure 21. A finding from the figure is that students from the instructor and No feedback groups degraded their essays more than the students from AWA feedback group. This might be because the automated feedback directed students to make relevant changes to the essay by highlighting the key areas to focus on for reflection, the students made those changes using rhetorical moves. The other groups which did not receive this might not have been able to focus on key improvements, leading them to make changes at a lower surface level and degradations. However, there was no statistically significant association between groups and revision quality in a chi-squared test: $\chi 2(4) = 7.7$, $p = .10$.

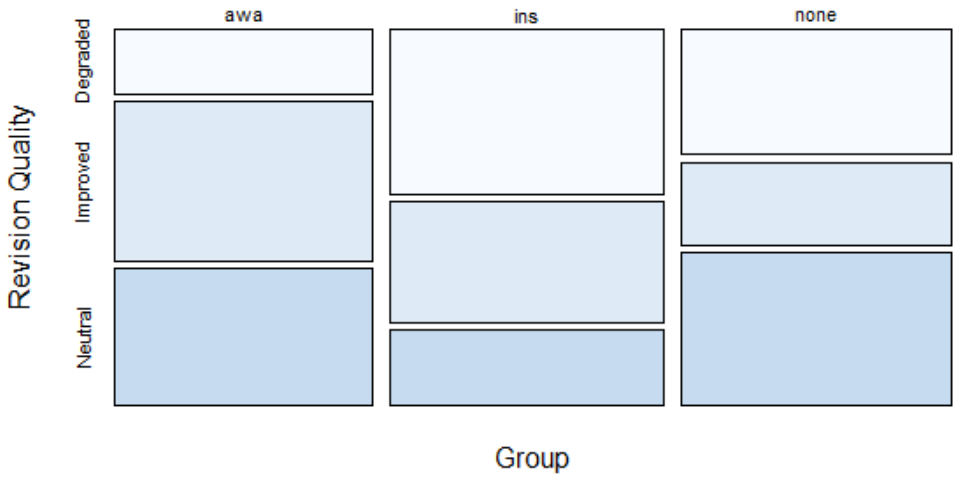


Figure 21: Proportion of students across groups categorised by revision quality

A problem with this method of manual marking for scores is that it is time-consuming, since it takes about 10 to 15 minutes roughly for each essay of this length to be marked by a tutor. It is also expensive and not scalable for large number of students. There are features of students' revised texts that can be exploited to identify useful indicators of revision quality. Hence, several methods to come up with revision metrics and visualizations to analyse writing revision were piloted by studying them in relation to the standard scores, discussed next.

4.2.5.3 *Methods to study revisions - Metrics*

To study the features of revision in students' improved texts and identify possible metrics, scores were used as the standard measure. Sample students' revised essays which were used to characterize good and bad essays with these techniques included the most degraded (16) and the most improved (15) essays to distinguish the extremes from the graded essays above. The first two metrics are simple quantitative measures which can be used to quantify the revision in terms of textual features. These measures assist in providing an overview of the revisions made and are usually good first-hand indicators of revision quality (detailed in the technical report, Shibani, Knight, and Buckingham Shum (2018b)). Other linguistic features that are used to quantify texts like cohesion (McNamara, Graesser, McCarthy, & Cai, 2014) are not included since they do not detect the nuanced differences made in the small texts in this context (i.e. original and improved/degraded essays as a whole do not give significantly different cohesion scores).

Word counts

A common measure to characterize text differences is a simple word count. Although it is a simple measure, word counts can be an effective indicator to get a general sense of the quantity of revisions made by a student at the word level (Fitzgerald, 1987). To further characterize revision, the relative change in word count in comparison to the original essay can be calculated as the difference between the word counts in the given and the revised essays. Even though this measure of the number of words changed cannot identify individual additions and deletions, it can still be an indicator for good or bad performance, depending on the context. Figure 22 shows the relationship between the performance category and the word count change measure.

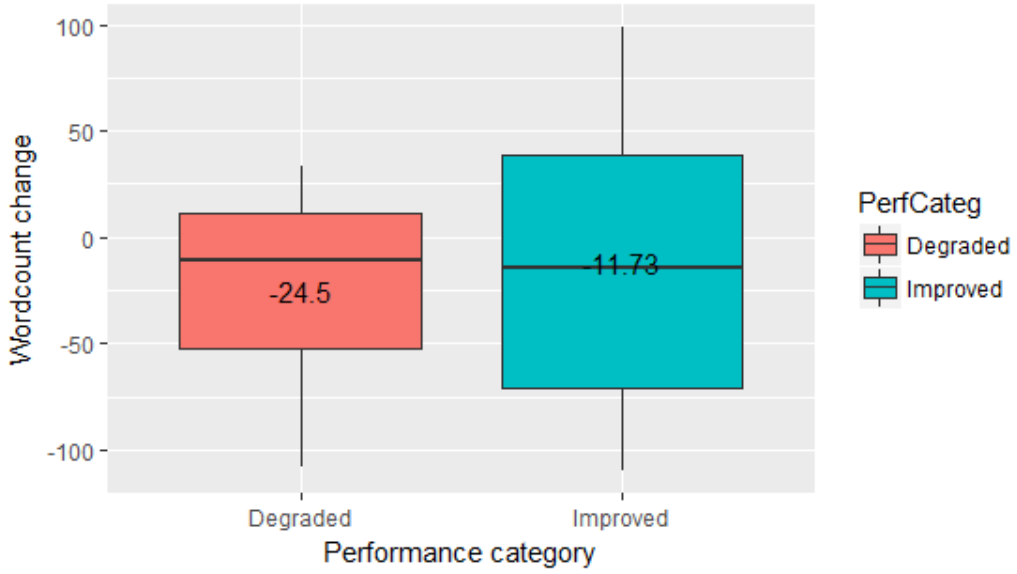


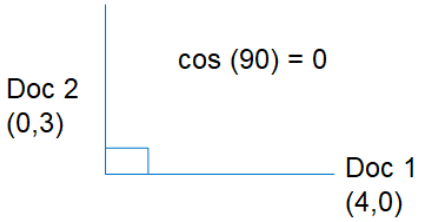
Figure 22: Change in word count measure against revision performance category (Means of each category displayed on the box plots)

In the examples used for our analysis, the mean word count change was -24.5 for degraded essays (*SD* = 48) and +11.73 for improved essays (*SD* = 70). The actual values of word count changed ranged from -103 to 33 for degraded essays and -110 to 99 for improved essays. For improving the essay, students generally added more words increasing the word count, and the degraded essays consisted of fewer words indicated by the negative value due to deletions. But there were also improved essays which had lower word counts. This shows that improvement need not always involve addition of words; students could make improvements by deleting words or rephrasing them to be more succinct. Pearson's correlation test did not indicate a relationship between the number of words changed and the total numerical score as such: $r(29) = .03$, $p = .98$. A similar result was observed showing no significant difference across performance category groups (degraded/ improved) in the number of words changed in a Welch's two sample t-test, $t(24) = -.58$, $p = .56$, Cohen's $d = -.21$. However, this simple measure of counting words does not capture other information including text-reorganizations, which may have significant impact on text quality. Simple variants on this method could capture sentence or paragraph counts, which represent some structural information. In addition, distinct words (i.e. words in the new text but not the original) could be represented to give some measure of textual change.

Text similarity and distance measures

Text similarity measures are commonly used in information retrieval and document clustering to compute the similarity or dissimilarity between two documents (Huang, 2008). In the current context, we are interested in studying the dissimilarity of revised essays in comparison to the original essay to quantify the changes made. This can be measured using Cosine Distance, which is a measure that quantifies the distance between two documents in a vector space model. This measure is used to calculate the distance measure between the given essay and the improved essay for all students to quantify the changes made by students. The measure of cosine distance quantifies the dissimilarity between two documents by computing $1 - x.y/(||x||\,||y||)$, where x and y are two document vectors. To calculate this measure, text documents to be compared are represented as vectors and the cosine of the angle between them measured in an n-dimensional vector space, where n is the number of words. Figure 23 demonstrates how the cosine distance measure is calculated using two examples, both assumed to contain two words in a two-dimensional vector space as shown in the corresponding tables.

	Doc1	Doc2
Word 1	4	0
Word 2	0	3



	Doc1	Doc2	Doc 3
Word 1	1	2	1
Word 2	4	3	6

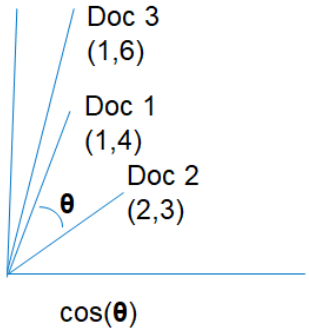


Figure 23: Sample vector space representations for calculating the cosine distance between two documents

In the first example, the two documents (doc1 and doc2) align with the word1 and word2 vectors respectively since they do not have words in common. The angle between their vectors is 90^0 and hence the cosine distance is 0 showing that the documents are dissimilar. Alternatively, if the two documents contain exactly the same words, then their cosine distance would be 1 (since *cos* (0) =1), showing that they are similar. In the second example, three documents are shown in the vector space to

demonstrate how Doc3 is closer to Doc 1, when compared to Doc 2 due to their word distributions in the vector space. This will lead to a higher cosine distance between Doc 1 and Doc 2, when compared to Doc 1 and Doc 3 showing that the latter is more dissimilar. An important point to note in this measure is that only the direction of the vector contributes to the angle and not its magnitude (frequency), so it does not matter how many times the word occurs in the document. For a more concrete example, consider a reference document containing the terms "essay" and "text" once each, that is compared with a comparison document (1) containing "essay" 5 times and "text" 0 times, and comparison document (2) containing "essay" and "text" once. Here, the reference document is more similar to document (2) than document (1).

To calculate the cosine distance between the given essay and the revised essay of students in our analysis, the 'stringdist' package in R was used. A higher distance measure signifying higher difference between the given and the improved essay implies more revisions made by the students. The mean cosine distance (scaled) was 5.2 for degraded essays (SD = 4.8) and 9.77 for improved essays (SD = 7.6). The cosine distance measure positively correlated to the numerical revision score: $r(29) = 0.41$, $p < .05$. So in general, the higher the distance between the original and the improved essay, the higher the performance. Welch's two sample t-test also found a difference at $p = .06$ level, across performance category (degraded/ improved) in cosine distance, $t(23) = -1.2$, $p = .06$, with a large effect size (Cohen's $d = -.73$) indicating that degraded essays may have a lower cosine distance. This would mean that students who made more changes to the essay had better performance than students who did not make many changes to the text. From our examples, the cosine distance serves as a good proxy to study the performance of students in revision. However, this measure is based only on the angle between the document vectors, and does not take into account the frequency of words, potentially losing some information about the revisions made. In cases where a semantic similarity measure would be helpful to analyse the content of revisions, more advanced word2vec word embedding can be used to represent the texts (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013).

4.2.5.4 *Methods to study revisions – Visualizations*

The next two methods use visualization to provide insights into the quality of actual revisions made. A good visualization should help the user make sense of complex information in order to support particular tasks, in this case, revealing the feature order

(e.g. word/ sentence order) in a text. To illustrate the use of these methods, we will introduce individual essay revisions of varying quality from two students. These two examples are from opposite ends of the improvement/degradation spectrum, to study their unique characteristics in detail. The methods described here are particularly useful to analyse essays which are very similar (revised from the same base text).

N-gram graph

Words and their occurrences in the revised essays can be analysed to study the revisions made to the given text. This can be visualized using an n-gram graph that captures all important words in the essay along with the position of co-occurring words. In this graph, nodes represent words and edges represent their occurrence next to each other in terms of the 'n' number of words selected. An example trigram graph constructed on our example revised essays using the 'igraph' package in R is shown in Figure 24.

The graph shows the density of occurrences and connectivity among the words used in the essays. The improved essays contain additional words like connectives (e.g. however) that connect the given words, producing a denser graph than the degraded and the original essay. This is visible from the change in clusters created in the graphs. This could be a useful way to represent essays visually to study their revisions, going beyond word-count or broad-similarity measures. Graph metrics such as distance, centrality and connectivity metrics could be utilized to further study these graphs quantitatively. Graphs could also be used to develop word clouds, or concept maps, to visualize the key concepts and their connectedness in a document. This visualization may be particularly useful in studying revisions made by adding many new concepts and ideas to the given base text which can be observed by changing clusters of words.

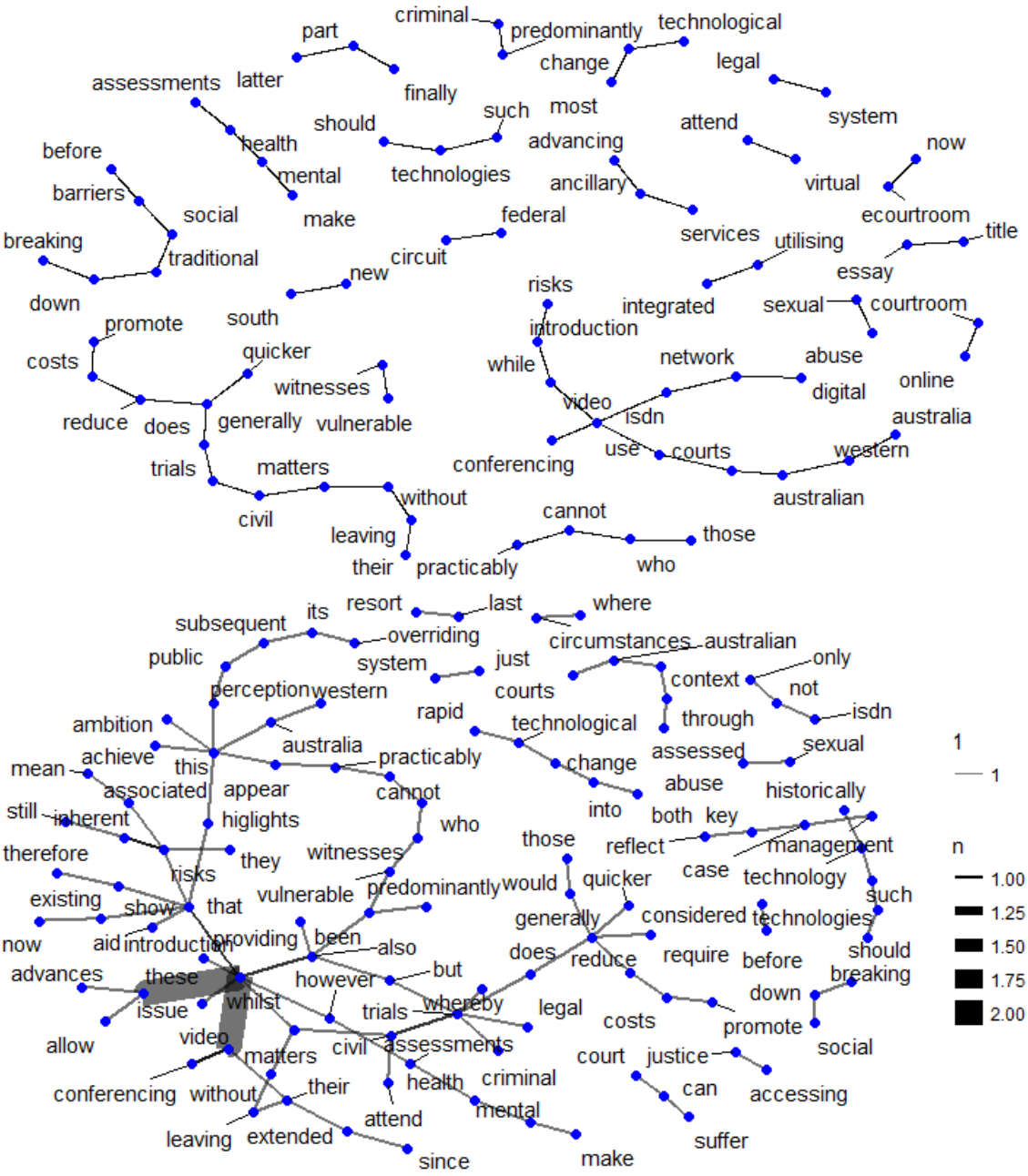


Figure 24: Sample trigram graphs from a degraded (top) and improved essay (bottom)

Rhetorical moves graph

The next analysis on the revised texts is based on rhetorical moves, since the core aim of the task given to students was to understand the use of rhetorical moves and discourse markers to improve an essay. To illustrate this analysis, the improved and degraded essays are mapped as a rhetorical move graph to study their sequences. Writing analytics tools like AcaWriter can automatically detect rhetorical moves in an essay using natural language processing techniques, simplifying analysis of this kind (Sándor, 2007). The rhetorical move graph is based on the method of visualizing

rhetorical moves as proposed by the sequence mining approach to study the patterns of rhetorical moves in student texts (Knight, Martinez-Maldonado, Gibson, & Buckingham Shum, 2017). The differences in the moves graph between sample degraded and improved essays can be seen in Figure 25, which shows the presence of rhetorical moves across paragraphs for the given text (left), a sample degraded text (middle) and sample improved text (right) which consolidated sentences to make two paragraphs.

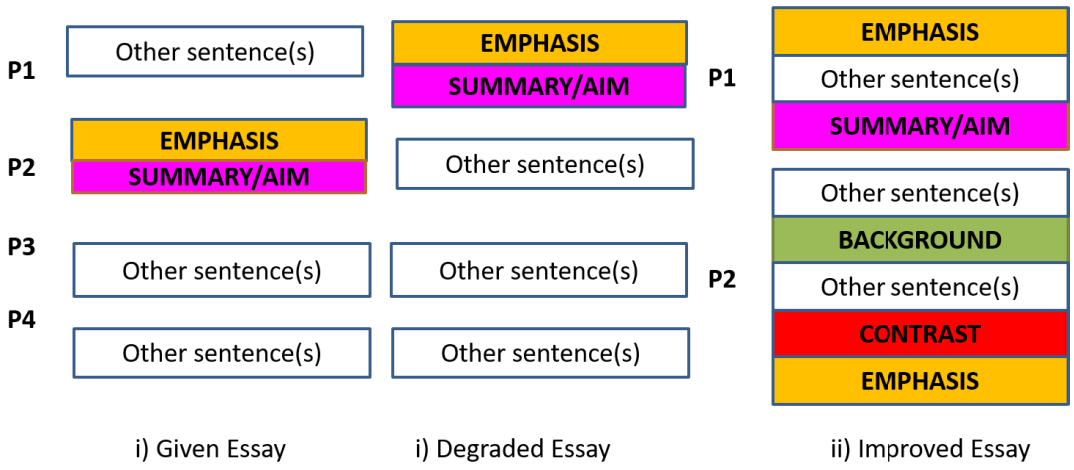


Figure 25: Sequence of rhetorical moves from a degraded and an improved essay

Emphasis is a move that usually highlights the key statement the essay's argument and *Summary* signals to the reader the author's intent and textual organisation. These were provided in the original essay along with some supporting statements. In our example shown in Figure 25 , the degraded essay did not add new rhetorical moves and only reorganized the given sentences. The other paragraphs contained no rhetorical moves to explicitly guide the reader with the flow of the text. On the other hand, the improved essay included discourse markers that indicate rhetorical moves such as explaining *Background* work, *Contrasting* critical statements, *Emphasizing* in the combined second paragraph, and removing inappropriate paragraphs. Although the examples described above show only a type of revision made using rhetorical moves, they illustrate the possibility of using such patterns to characterize other revisions as good and poor.

4.2.5.5 *Method to study the process of revision – Revision Graphs*

To study the *process* of revision in students' writing, a novel approach to revision analysis called 'Revision Graph' was developed. Because the analysis of revisions is generally conducted on multiple drafts of a produced text, it is challenging to interpret

differences between texts. Analysis of revisions on a provided 'base text' may provide an innovative approach to addressing this gap. This builds on previous work in revision analysis by studying the process of revision over multiple drafts created through the text revision exercise. This analysis over multiple drafts using Revision Graphs can aid in uncovering the previously unknown processes involved in the editing of the final revised essay. This new analysis focused on the ordering of sentences and revision actions in a visual representation at sentence level. Performed manually in this iteration as a prototype, Revision Graph generation was later automated in design iteration 3 (discussed in Section 4.4.5.4).

In the following revision graph, the *nodes* represent sentences from the drafts and the *edges* (arrows) represent changes in the organization of sentences across multiple drafts. The sentences are represented in the sequence of occurrence across the paragraphs. The colours of the nodes indicate the type of revision action made at the sentence level: i) minor revisions are when students predominantly use the given text, but add or substitute few words, ii) major revisions are when students add a substantial number of words and explanations to the given text with the inclusion of their own writing, iii) no changes made and iv) no change in the current stage, but deleted in the next stage. Red inverted triangles represent the request of automated feedback during the revision process. Dotted edges are used to represent the repetition of similar concepts across multiple sentences inside a draft. This could be a good indicator of word repetition/ overlap leading to high cohesion in the document.

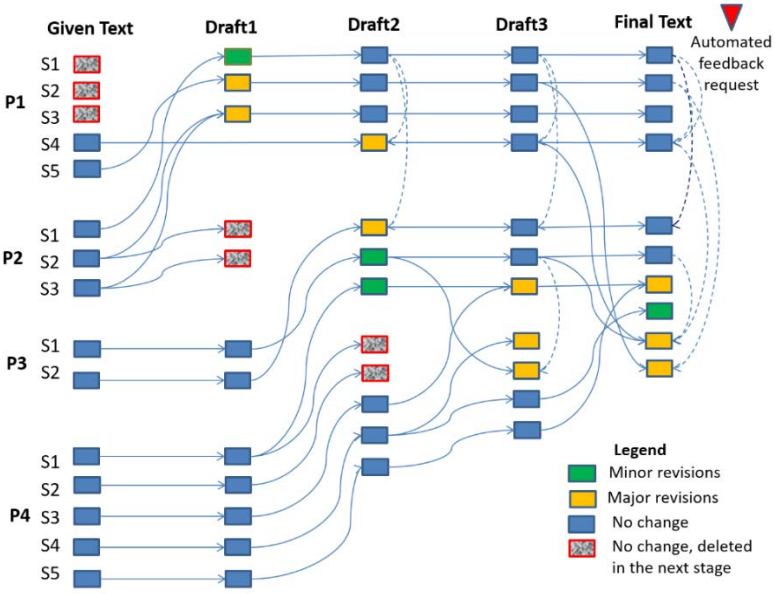


Figure 26: Manually constructed revision graph of a sample improved essay

Figure 26 shows the revision graph constructed manually from the sample improved essay's drafts to show the evolution of a high-scoring revised essay. The drafts were selected at set intervals (every 6 minutes in this analysis) using the time spent on revision. The graph shows the stages in the revision of the *given text* containing four paragraphs P1-4, and 15 sentences, to the *final text* containing two paragraphs and 10 sentences. In the first draft stage, the student has deleted some broad introductory sentences from the original essay. In *draft 1* the first paragraph of the draft has been shaped up by making minor and major revisions to the given sentences and reordering them, while the other paragraphs remain untouched. In *draft 2*, the student has deleted the previous second paragraph and mainly worked on the revision of this paragraph from the other paragraph sentences. Here the text has been reduced to three paragraphs.

In the third draft, the first paragraph remains stable, and the student turns their attention to paragraphs 2 and 3, making extensive changes, revising and consolidating them to produce a final text consisting of only two paragraphs. The number of references (i.e. repetitions) to previous words increases in each stage of the draft as shown by the dotted edges. The final text has many such cross references made to the previous sentences, which has improved the cohesion of the text. This student requested automated feedback (red triangle) after completing the final text and made no more changes after that. This information is made visible by matching the timestamp of feedback request with those of the drafts. It informs that the changes made to the text by the student were not an effect of the feedback received. In cases where we do not have such process information to study writing, it is feasible that the revision effect is attributed to feedback, but they are in fact not related. This revision graph is thus serving its purpose of making visible, at an appropriate granularity, the nature of the revisions, and whether the automated feedback component impacted subsequent revisions.

In the revision graph of a sample degraded essay shown in Figure 27, there are no revisions made by the student to the given sentences. The introductory sentences have been removed in the first draft, and sentences have been reorganized in the second draft. No further changes have been made from the second draft to create the final revised essay, leading to a degraded version of the given text. The last three drafts have

remained stable, meaning the student has stopped working in the last few minutes of the revision task.

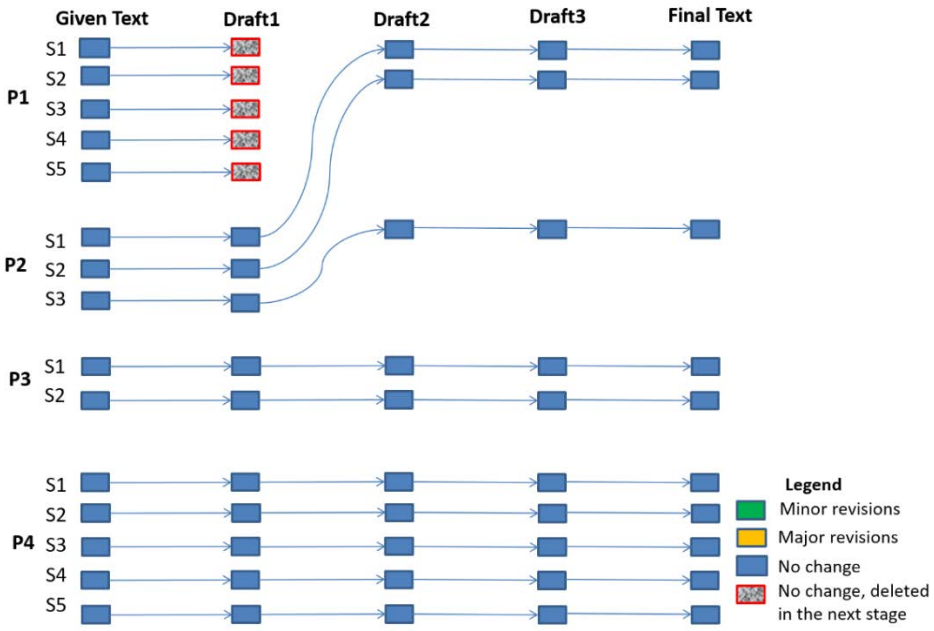


Figure 27: Manually constructed revision graph of a sample degraded essay

Sample pseudo code is shown in Figure 28, which would operate on writing process data collected automatically (students' drafts at specific intervals) by AWA-Tutor. The above manually constructed revision graphs were later automated for a large scale analysis of revision process in the next iteration, described in Section 4.4.5.4. The prototypes thus provided a starting point for making use of such representations and metrics to study the quality of improvements made by students in the writing task, some of which were explored in more detail in later iterations.

```
Create the corpus of n ordered drafts based on timestamp
For each draft document:
    Parse document into sentences
    Create a column of nodes in the graph for each sentence ordered by
    paragraph
    If n > 1: (Start from the second draft to add edges)
        For each sentence:
            Calculate similarity score with each sentence from the previous
            draft
            If similarity score > similaritythreshold:
                Add edge in graph between the current sentence and
                previous draft sentence
                Use similarity score to colour nodes for minor, major
                revisions and no change
            Calculate word overlap with all other sentences in the same draft
                If word overlap > overlapthreshold:
                    Add dotted edge in graph between the
                sentences
If no edge originates from a sentence in draft (excluding last draft):
    Mark sentence node as red deleted node
```

Figure 28: Sample pseudo code for automatically constructing a revision graph

In the current context, the revision graph was used to study the revisions made across drafts in terms of the strength of revision actions like insertions, deletions, substitutions etc. It can also be applied to other specific changes we would like to study like the types of revisions (e.g. concepts, rhetorical moves, surface errors, etc.). Another application of this revision graph, as mentioned previously in the revision process analysis of a good revised essay, is to study the effect of automated writing feedback using actual revisions made by students. This way of evaluating the effectiveness of Learning Analytics applications (automated writing feedback in this case) is thus made possible using Learning Analytics itself (tracking the revision process in student drafts for detailed study).

Further research is required to investigate the potential of these techniques to differentiate texts by quality beyond examples of extreme cases of performance; thus, having demonstrated the utility of the revision graph in principle, to test its performance on text corpora at scale requires software implementation. An automated approach to do this analysis reduces human effort and has evolved from a later iteration (see Section 4.4.5.4). To extend their usage in educational contexts, further work could be done to characterize essays based on the discussed features to provide meaningful feedback to educators and students. The feedback might be based on writing patterns

that emerge or revision types, e.g. to draw attention to the fact that there have been no substantive changes in graphs after 2 drafts or within a defined time interval, or changes that only involve surface level error corrections. The revision patterns observed from revision graphs could be the first step towards studying the contexts in which automated feedback can work better, and other contexts in which other forms of feedback like human feedback are well suited. Further cognitive processes can be studied using think aloud techniques to capture the mental models while adopting/ rejecting the feedback.

4.2.6 Discussion

In this iteration, we designed a pedagogic intervention to support student learning, and both implemented and evaluated the potential of the writing analytics tool AWA and its automated feedback on rhetorical moves. The pedagogic design and development of the learning analytics platform AWA-Tutor (to support the activity flow of the intervention) exemplified a learning-design oriented approach to learning analytics, which was built upon further in future iterations. Traces of student data made available by the platform enabled many types of analysis with the use of quantitative and qualitative techniques. Methods to study revision products and processes were piloted for future use. The results from the study provide insights to make changes in the next iteration for improvements to both students and educators' experience. Based on our preliminary findings, students generally found this activity useful in developing their writing. Qualitative analysis of students' comments on the activity shed light on the usefulness of different subtasks in the pedagogic design that contributed to their writing skills. It was observed that students found the exemplars, self- assessment and revision skills applicable for their own writing in the future. Such interventions are seen to improve students' understanding of the instructor's rubric and their writing skills by deliberate practice of these skills.

Analysis of students' comments regarding the feedback they received provided information to instructors and researchers on the types of feedback that students find useful, and their expectations of feedback from instructors and automated tools. Students particularly highlighted a desire to receive explanations on why certain sections are highlighted and how to improve the text further, which was taken into account when improving automated feedback from the tool in later iterations. Giving such actionable feedback for students to close the gap between the expected and current

performance is a principle of good feedback practice, which has to be followed for any kind of formative feedback (Nicol & Macfarlane-Dick, 2006). When tools are used, students also require proper guidance to interpret the results and use the tool for its intended purpose. It was also found to be crucial to design effective pedagogical practices keeping in mind the inherent limitations of automated tools. This led to changes in task design in the next iteration where students were provided with scaffolding in the form of peer feedback and discussion, as additional support while using automated feedback. Student motivation was studied using a proxy of downloads, and their comfort with automated feedback was studied from their rating leading to the insight that students required more explanation and induction for participating in the activity effectively. Taking this into account, the next iteration made changes in how the task was introduced to students, by explaining its rationale with more concrete examples. For details on the major changes in the next iteration design based on lessons from the current iteration, see Section 4.3.1.

Analysis of scores obtained by students on the revised essay in this task added insights in addition to students' self-reported data on the actual revisions made. Students from the automated feedback group had learned to identify parts of the writing that needed improvements more than the other groups with very few degraded essays produced in comparison to the other groups. Using these scores as the standard measure, other methods to study revision were developed. These methods illustrated the use of text analysis and visualization techniques, exemplifying its potential for learning analytics applications by using students' essays from the revision task. The combination of product and process centric methods to study revision helps to uncover deeper aspects of writing. The methods demonstrated different granularities of analysis of revisions starting from a set of essays, extending to individual essays and internal processes in Section 4.2.5. Word count change and a text similarity measure – cosine distance – were used to quantify the changes made in the final product of revision in a set of degraded and improved essays. The cosine distance was a good measure of the performance since students who make more changes to the given text generally score higher in the task. Visualization techniques like the n-grams graph based on word occurrences in text and the rhetorical moves graph constructed from the sequence of rhetorical moves, were useful to study further the revisions made in individual texts. Such visualizations of texts are an effective way to observe the characteristics of these revised texts.

The construction of a revision graph to visualize the revision process of essays introduced a novel approach to study the evolution of writing in terms of the actions that led to the final product. This visualization revealed a pattern of activities like addition, deletion and re-organization of sentences in the generation of the improved essay showing the importance of understanding textual restructuring and the revision process in writing. It demonstrated the opportunity to study the different ways that good or poor writing may evolve in its revision stages and served as a prototype for the automated revision graph developed in a later iteration discussed in Section 4.4.5.4. It became a particularly useful tool to study the effects of writing feedback on students' revisions at multiple stages in the later iterations.

4.3 Pilot for Design Iteration 3

A pilot was run in preparation for the next main iteration, design iteration 3. It was run in the immediate next semester which did not allow time for significant changes to be completed in the tool's automated feedback, but some minor changes were made, explained next.

4.3.1 Task Design and rationale

From the results of the previous study, further improvements were identified to improve student uptake. New ways of helping students by providing better feedback and improving social sense making in writing were piloted, by improving the explanations of the purpose of the activity, and introducing a dyadic setting.

1. To improve students' understanding of the purpose of the activity, a better explanation was required to teach students the use of rhetorical moves and discourse markers. This was done by including an initial reading as a new component at the beginning of the activity which explained the use of these elements in their essay writing context. AWA feedback in the revision page of the AWA-Tutor interface also included descriptions of rhetorical moves in the tool's user interface using a help button. This enabled a better understanding of these moves by students, in contrast to the earlier label only user interface in AWA.
2. A dyadic setting was introduced to pilot peer discussion in combination with automated feedback to improve sense making and learning.

The initial reading added in the task design was implemented using a supporting material (in Appendix A) written by the instructor. This was read aloud during the tutorial and given as a handout for students to download from their learning management system. Based on feedback from students on the previous iteration regarding allocation of time, AWA-Tutor interface included an ideal time for completing each sub-task in all its pages, and showed students a progress of the activity using a progress bar as shown in Figure 29.

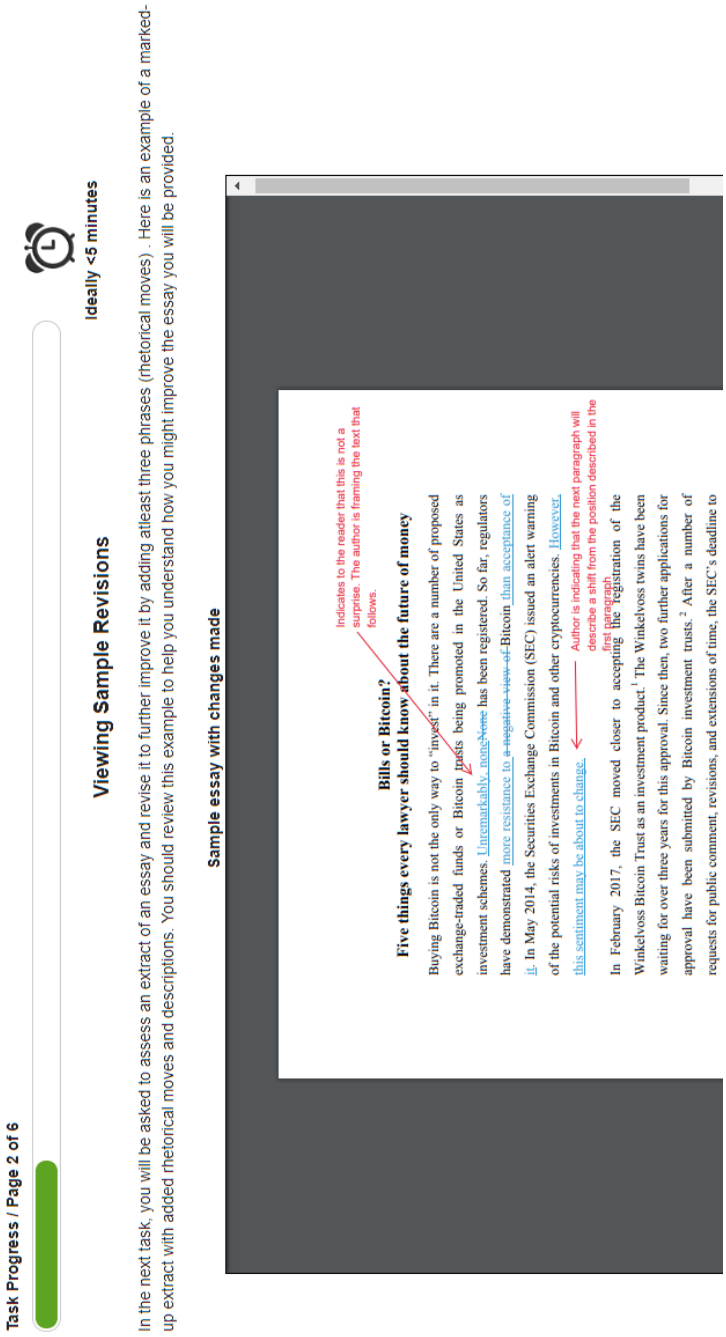


Figure 29: Screenshot showing changes in AWA-Tutor UI, which included a progress bar and time allocation at each sub-task page

The other tasks remained almost the same as in the previous iteration, but instead of the tasks being completed individually, a critical new feature was that some students were asked to complete it collaboratively with a peer. This peer discussion component was added to improve student sense making and engagement with the automated feedback. While working with learning analytics tools and dashboards, the human context is often emphasized as central in interpreting and making sense of the analytics (Siemens, 2012). Sense-making and interpretations are hence important in writing analytics tools, in order for students to understand and implement the automated feedback that is provided on their writing. One way of providing sense-making support is through peer feedback and discussion, where students can interpret, discuss and critique automated feedback on writing with their peers.

It was hypothesised that combining peer discussion and automated feedback would bring two benefits. First, peer discussion overcomes limitations in automated feedback by complementing it with contextual feedback by peers which might identify features of writing missed by the tool. This brings in a human context which is lacking in automated feedback and enhances the social and cognitive processes involved in writing. As discussed further in Section 2.2.3, students also learn from each other while providing feedback on each other's writing by making judgements about their performance (Allal et al., 2005). Second, automated feedback may address a concern in peer feedback regarding student's abilities to provide meaningful feedback discussed in the literature review (Section 2.2.3), by scaffolding this feedback and provoking discussion around the identified features. Hence, this design combined known effective practices like peer feedback and discussion with automated feedback, in order to complement each other and to aid in improving students' writing skills (Shibani, 2017b).

In order to study the impact of automated feedback in detail, this iteration had just two feedback conditions: *Automated feedback* group and *No feedback* group. To study the effect of collaboration, *individual* vs *dyad* conditions were set up – some students completed the activity individually, while others completed it together with a peer. Students in the automated feedback group accessed AWA feedback on rhetorical moves during their revision task. The feedback consisted of the analytical report highlighting the rhetorical moves similar to the last iteration. A design pattern representing the key changes made in this iteration is provided in Design Representation 3A.

PILOT DESIGN 3: Benchmarking, Text-Revision, Peer-Discussion, and Automated Writing Analytics

Problem: Building on the previous designs, we additionally wanted students to engage with each other around the application of assessment criteria, to further develop their evaluative judgement, and ability to explain and justify their judgements of texts and their revisions.

Task: The initial base tasks in design 2 were adapted, such that in in one group of students they were asked to work as dyads, submitting a single revised text, and in the other group they worked individually.

Tools/materials and participant structures: In this design, the participant structure varied by group, with some working in pairs and others individually. When students work in dyads, they involve in discussion consisting of reflection and critique on the structure of essays and the application of automated feedback. The materials and tool for this design are the same as those in design 2.

Iterations and Augmentation: This task design developed from that described in design 2. A key concern in this design was that peer discussion may mediate the understanding and use of the augmented feedback provided by AWA; that is, this task may develop students' abilities to – critically – use such feedback.

Design Representation 3A: Design pattern of the task design for iteration 3 pilot

4.3.2 RQ1a) Student Perceptions of the writing intervention

Since the iteration was a pilot run before the next main iteration, data analysis from this iteration included studying student perceptions of the intervention only to identify any limiting factors that can be rectified in the next iteration. To study student perceptions of the intervention that answer RQ1a, data from 124 students who completed the activity was analysed. 55 of them were from the AWA feedback group, and 69 were from the No feedback group. The perceived usefulness of the activity score of students across the two groups (awa vs none) is shown in Figure 30.

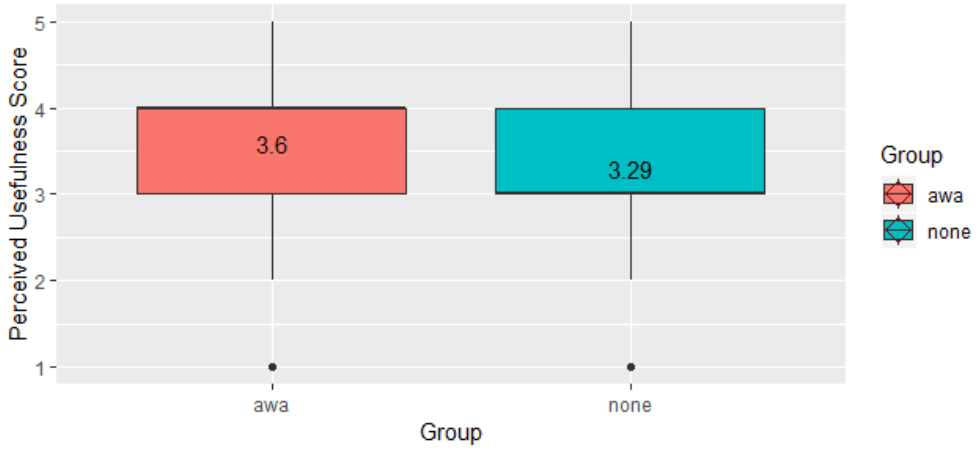


Figure 30: Perceived usefuless of the activity score across groups in iteration 3 pilot

The perceived usefulness of students increased in this iteration compared to the previous iteration (Figure 19), and the AWA feedback group found the activity to be more useful (M = 3.6, SD = 0.95) than the No feedback group (M = 3.29, SD = 1.16). However, while an encouraging trend towards awa, no significant differences between the group conditions were found in a Welch two sample t-test: $t(121) = 1.63$, $p = .11$. Also, there was no significant difference between the individual vs dyad setting: $t(97) = -.28$, $p = .77$.

The perceived usefulness of students increased in this iteration compared to the previous iteration. The feedback comments of students were studied to understand student perceptions of the new changes in this iteration and the collaborative setting in more detail. Students generally had an improved understanding of the purpose of the task and the use of rhetorical moves and discourse markers in their essays. This can be seen from the comments of students from both the AWA feedback group and the No feedback group exemplified below, where students demonstrated an improved understanding of rhetorical moves. This result can be partly attributed to the reading added in the current iteration in the beginning of the task which explains students what these rhetorical moves and discourse markers are, and how they add clarity to academic writing.

"There was a variety of feedback that was useful in revising the essay, including the importance of using rhetorical moves to enhance the quality of the text, deepen the argument and allow for the reader to obtain a more developed understanding of what the essay is trying to convey. This drew my attention to how using discourse markers improve the essay and allow for the reader to be able to flow on with the essay with ease, comfort and guidance." – Student 55, AWA feedback group

"It is useful as the use of this tool allows me to pinpoint what pieces of writing can be considered good pieces of analytical writing. It is often difficult when simply proofreading your work to understand the mechanics of your writing. For example, understanding how to effectively use rhetoric moves through a piece of writing to guide the reader through your argument and to be aware of what discourse markers you have used in your essay." - Student 52, AWA feedback group

"I think it was good to pinpoint what is considered good use of language and what was not helpful. I also liked the fact that they categorised the use of language that is expected in the essay to be able to keep this in the back of mind when writing essays as we know what to expect." – Student 21, No feedback group

"In previous essays in both law and other disciplines this concept of engaging with discourse markers which I believe now that I am conscious it will be important in portraying my own voice and opinions with further analytical writing. It will also force an evaluation of evidence used which previously my essays may have lacked." – Student 75, No feedback group

"The task reminded me of the importance of structure, signposting and clear language so as to engage the reader." – Student 60, AWA feedback group

However, it was observed in the classroom that many students wanted to complete the activity on their own laptops and did not prefer working in the collaborative dyad setting for the whole task. This might be because of their motivation to complete the task by themselves, or due to the difference in the pace of students for different tasks (reading for example) as in the comment below:

"A bit too much reading to conduct the task in pairs - but I found it helpful :)" – Student 7, No feedback group

4.3.3 Discussion

Preliminary analysis from this pilot data showed increased acceptance and understanding of the activity among students. In general, the AWA feedback group students found the activity more useful than the No feedback group in this iteration, probably due to an improved understanding of the purpose of the activity and the usage of automated feedback. However, no statistically significant difference was found

between the conditions. Some students did not like sharing work with peers and wanted to have their own versions - so they still completed two copies on their own laptops. This led to incomplete datasets, and a further decision to keep the overall activity individual with the peer discussion as an embedded component of that individual task.

As discussed in Section 3.2, a feature of design-based educational research is the lack of control over authentic teaching and learning conditions in which interventions are studied, compared to laboratory experiments. A good example of this was the variability in how different tutors explained the purpose of the task to students, using the given initial reading, leading to varied levels of understanding in students belonging to different tutorial groups. This finding prompted the decision to create for the next design iteration a standardised video by the lead Law academic, explaining the relevance of the activity, and importance of rhetorical moves, in the context of legal writing.

4.4 Design Iteration 3

Design iteration 3 was the main iteration where the intervention design was stabilized for Law with different elements. It contributed to the development of theoretical principles in learning analytics from practice by aligning it to learning design, formalising them using a conceptual model explained in Section 5.1. It emerged from the process of implementation of LA in authentic writing practice described in the iterations of this research by identifying key elements that contribute to the development of context-sensitive learning analytics tools.

4.4.1 Task Design and rationale

The task design of this iteration was similar to the previous iteration but made four key changes based on the feedback from the previous iteration. Firstly, it introduced AWA-Tutor writing activities in-class as an individual activity but with peer discussion as one component of the design, removing the collaborative setting for the whole activity. Secondly, the example text provided to students was revised to include only two paragraphs removing references, to focus students on key revisions based on rhetorical moves. Thirdly, on the technical front, a major development was the introduction of a new, faster and more robust version called AcaWriter. In addition to the highlighted rhetorical reports on students' writing given by AWA, AcaWriter provided additional contextual feedback specific to the writing context. Finally, a video introduction was

played at the start of the task where the lead Law academic of the subject explained the purpose of the task and the concept of rhetorical moves, to minimize tutor differences in explaining the purpose of the activity. The changes in task design for this iteration 3 are represented in Design Representation 3B.

DESIGN 3: Benchmarking, Text-Revision, Peer-Discussion, and Automated Writing Analytics with Contextualized Feedback

Problem: We wanted students to engage with and make sense of the automated feedback better around the application of assessment criteria, to further develop their evaluative judgement and revision skills, and the ability to explain and justify their judgements of texts and their revisions.

Task: The initial base tasks in pilot design 3 were adapted, but students completed the task individually with an embedded peer discussion component.

Tools/materials and participant structures: In this design, the participant structure was individual for completing the task. The revision task included a peer discussion component after their individual revision activity to share feedback and learn from each other. When students involve in discussion, they involve in reflection and critique on the structure of each other's revised essays and the application of automated feedback, and through observation of this dialogue research and implementation, data is obtained to study impact. The materials for this design are the same as those in design 2 and pilot design 3.

Iterations and Augmentation: Based on the student feedback from previous designs, a new tool called AcaWriter was introduced. The improved automated feedback on rhetorical moves that was contextualized for the subject can aid in better revision by improving students' understanding of the assessment criteria. It can also scaffold the peer

Design Representation 3B: Design pattern of the task design changes in iteration 3

4.4.2 Developing a formalised LA-LD model using the Law context

Iteration 3 highlighted more clearly than previous iterations what the key elements are that contribute to better contextualized writing support. The first step to contextualize AcaWriter for the Civil Law context was to start with the assessment. Rubric elements from the assessment criteria were mapped to AcaWriter's rhetorical moves to ensure that its automated feedback was congruent with the assessment regime. This led to identification of features that were useful for the context at the tool level, while also feeding back to the LD where the assessment criteria needed to be better defined to capture the intended learning outcomes. *Thus, LA and LD shaped each other to achieve*

better alignment. The features for the LA tool might also need to be freshly created for the learning context if desired, or tuned at the system architecture level to modify existing features for the context as demonstrated in an earlier example (Knight, Buckingham Shum, et al., 2018). Sample sentences of this mapping of rhetorical moves to elements of the assessment criteria provided to students as examples in AcaWriter are in Figure 31.

Sample Sentence	Essay Assessment Rubric Element	AcaWriter Move
The concept of good faith **has previously been thought** to be a work in progress in Australia.	Engagement with the law and scholarly literature	Background
This article will trace the origins of good faith and its development in the common law. **This essay contains three parts. The first part will talk about** the origins of good faith.	Statement of thesis, Essay plan	Summary
However, where the obligations are found in statute and they conflict with contractual obligations, **it is important to note that** the former must prevail.	Identification of relevant issues, Critical analysis and original insight	Contrast,Emphasis

Figure 31: Assessment criteria from the Law essay writing context mapped to rhetorical moves from AcaWriter

In this learning context, the analytical report highlighted all the rhetorical moves identified in the students' text (Figure 32), and the feedback messages prompted reflection for students to focus on the key moves that should ideally be present in their writing. Based on the alignment to assessment, feedback from AcaWriter was tuned for the context to provide feedback messages on missing key rhetorical moves, and suggesting sample phrases to use as shown in Figure 33. While these samples were provided as exemplars from which students can learn how to express the important points in their essay better, it was by no means a concrete rule that prompted a fixed style of writing for all students. There are many ways which in which the moves could be expressed and students were explicitly prompted in the feedback to use their human

judgement when it comes to evaluating the automated feedback and structuring their own writing.

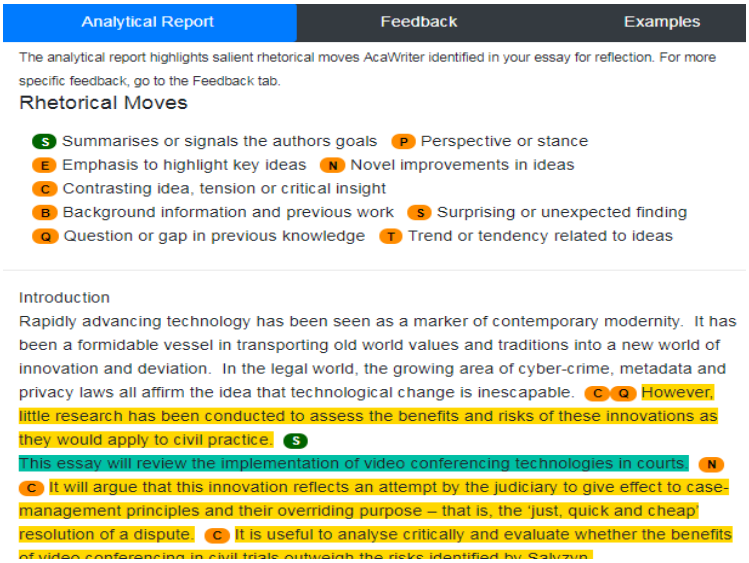


Figure 32: Report highlighting rhetorical moves automatically identified by AcaWriter as part of the feedback from AcaWriter

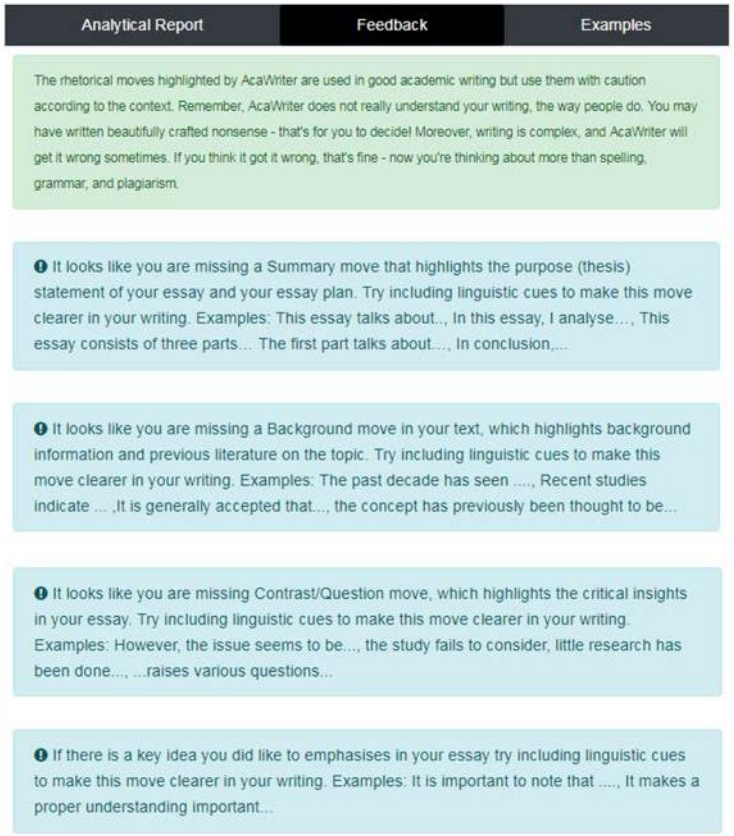
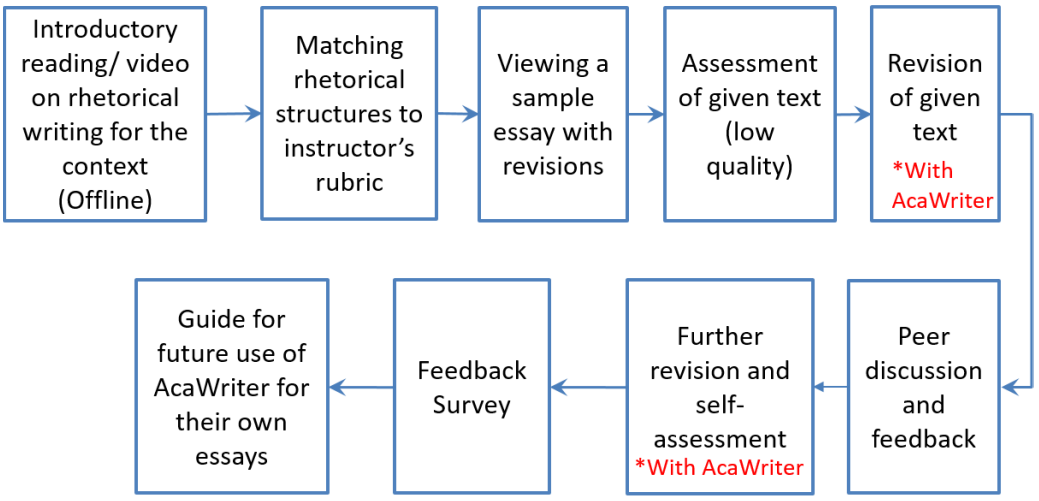


Figure 33: Feedback messages from AcaWriter tuned for the Law essay writing context

To introduce the AcaWriter tool to students in a way that they can make sense of it in its pedagogical context, the task design helped to coherently integrate the LA tool into the curriculum. This task was implemented as a pedagogic intervention that included a number of online and in-class tasks for students to write better essays for the subject and use AcaWriter to get feedback on their drafts. All of these elements were co-designed by LA researchers with the instructors who were the subject experts familiar with the learning contexts working directly with students.

The writing intervention developed for students in the Civil Practice subject cohort was a one-time activity that was facilitated in class in a tutorial session in authentic settings. Students were divided into two groups for us to evaluate the use of contextualized tool in this scenario: AcaWriter feedback group, No feedback group. The design of the task is shown in Design Representation 3C.



Design Representation 3C: Workflow of tasks in iteration 3 where the writing task design integrates AcaWriter for Law context

First, an induction was provided to the task using an introductory reading or video which was created by the instructor to explain the goals of the activity. Then students logged in to the AWA-tutor online platform which facilitates the writing tasks online and collects activity data. The first task was a matching exercise where students were asked to identify sample sentences from an essay that would match elements of the

instructor's marking rubric, so they have a better understanding of the rubric facets in the assessment criteria by learning from exemplars. In the next task, students were provided with a sample improved essay pdf where the instructor highlights what changes can be made to the text and rhetorical moves can be added to improve the structure and clarity of the essay and its concepts. Then students had to assess the given sample essay by applying the assessment criteria, to enhance their ability to self-regulate their work by practicing the skill of self-assessment.

The main task followed, in which students were asked to revise this given essay which they had assessed earlier, to improve its quality. For this activity, students in the AcaWriter feedback group used AcaWriter to receive automated feedback to revise and improve the given essay, while students in the No feedback group did not get access to AcaWriter and revised the essay based on the guidance provided in the previous tasks only. Then students had a discussion with their peers on the changes they made on their improved essay and provided further feedback. Peer feedback has been seen to be effective in writing instruction since students learn from each other, and by providing feedback they become better judges of their own work. For groups receiving automated feedback for revision, this peer feedback provided additional contextual feedback and sense-making to interpret and augment the feedback from AcaWriter (Shibani, 2017b). Except for this revision task where the AcaWriter feedback group students were provided additional access to AcaWriter feedback, the two groups were involved in the same activities for comparing their effectiveness. Finally, students completed the feedback survey where they answered questions on the usefulness of the intervention. For future use of AcaWriter for their own draft essays, a guide was provided. The earlier iterations explain the rationale behind the tasks designed, and preliminary results. In those interventions however, students only received a generalized analytical report with the identified rhetorical moves highlighted in the text with no contextualized writing feedback from the automated tool.

4.4.3 Data Analysis

To evaluate this contextualized approach to provide writing support for learners, data from this pedagogic intervention was analysed as in the previous iterations. However, there were some technical issues in the working of the new tool AcaWriter during the initial sessions, which led to loss of data. The results discussed next are based on responses from 90 students who completed all parts of the activity in this learning

context and did not face technical issues. This cleaned data included 44 from the No feedback group and 46 from the AcaWriter feedback group.

Data analysis in this iteration includes studying the perceptions of students on the writing intervention with and without automated feedback to address RQ 1a. The revisions made by the students in the revision task are studied using various metrics to answer RQ1b on the impact of automated feedback on revisions. To dig deeper into the process of revision that students engage in using automated feedback, their stages of revision are studied using automated revision graphs to address RQ2a. Additionally, the scaffolding of automated feedback with peer discussion is studied to address RQ2b.

4.4.4 RQ1a) Student perceptions of the writing intervention

Students perceptions of the writing intervention with and without automated feedback were studied from students' responses to the survey questions on how useful the activity was for them to improve their essay writing on a scale of 1 (not useful at all) to 5 (very useful) and why they thought so. Figure 34 shows the difference in the perceived usefulness score between the AWA feedback group and the No feedback group.

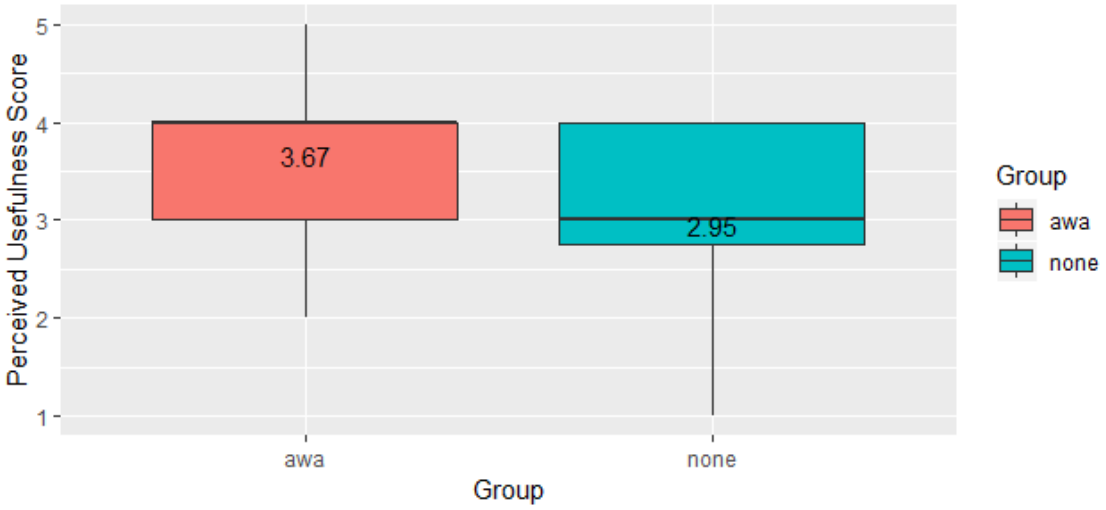


Figure 34: Perceived usefulness across groups in Iteration 3

The mean perceived usefulness score for students who received AcaWriter feedback (M = 3.67, SD = 0.79, N = 46) was higher than those who did not receive AcaWriter feedback (M = 2.95, SD = 0.96, N = 44) during the revision task. The effect of AcaWriter feedback in the activity was found to be significant in a Welch two sample t-test: $t(83) = 3.9$, $p < .001$ with a large effect (Cohen's d = 0.82, 95% CI [0.38, 1.25]). This result was in contrast to results from the earlier interventions with generalized

feedback, where no significant difference was found between the control groups. While there are different elements that play a role in the contextualization, the contextualized application of the LA tool by tuning the feedback seems to have an impact on students' perception of the activity due to its direct interaction with the students. Qualitative responses from students also showed that many students found the activity useful to learn skills that can help them self-assess and improve their writing by making them self-aware of their writing.

"When you're editing your own writing, you automatically think that your work sounds good and that all your ideas and views have been clearly conveyed. This exercise was useful in the sense that it indicated areas where I needed to be more explicit, which on my own I would not have noticed." – Student 40, AWA feedback group

"allowed me to understand how to clear up writing style and language in providing a more succinct and cohesive argument, especially in the introduction, which reverberates throughout the essay in aiding structure and clarity of argument" – Student 21, AWA feedback group

"I think what is being taught is something I was already aware of. However, by being forced to actually identify ways of arguing, along with the types of words used to do so, it has broadened my perspective. I think I will be more aware of the way I am writing now." – Student 55, AWA feedback group

"Made me think about the structure of an essay more and how to make the essay more persuasive." – Student 27, AWA feedback group

"I have realised you need to use more adjoining words as well as rhetorical questions, something I don't use often in my essay writing." – Student 76, No feedback group

There was also some scepticism regarding the extent to which artificial intelligence can help in a nuanced language understanding problem, especially when the right terms are not explicitly used in the writing.

"A good reminder of important elements of essay writing. However, I am not sure how useful AcaWriter actually is other than providing some general feedback." – Student 22, AWA feedback group

"It made me think about the importance of having clear writing with a clear stance and structure. The technology helped somewhat with this, but I found it was limited to

recognised extremely plain language. Language which was more complex but accomplished the same thing was not picked up on." – Student 2, AWA feedback group

The writing task embedded the learning analytics tool in the curriculum, by closely aligning it to the learning design. It helped students understand the usage of rhetorical moves, and the usage of AcaWriter feedback to write better essays for their subject and improve their writing skills. AcaWriter capabilities are seen to augment the design with better results, suggesting the appropriateness of large-scale deployment of the tool for all students in future implementations.

4.4.5 RQ1b) Impact of automated feedback on student revisions

To study the impact of automated feedback in terms of the actual revisions made, revisions made by students on the given essay were studied using the below metrics.

4.4.5.1 *Scores Analysis*

The revised essays of students were marked for improvements by the same tutor similar to the marking done in iteration 2. Two outlier records of students who wrote a completely new essay on the topic instead of revising the given essay were removed, resulting in a total of 88 improved essays, 45 from AWA feedback group, and 43 from No feedback group. The final score attained by students across the two groups is shown in Figure 35.

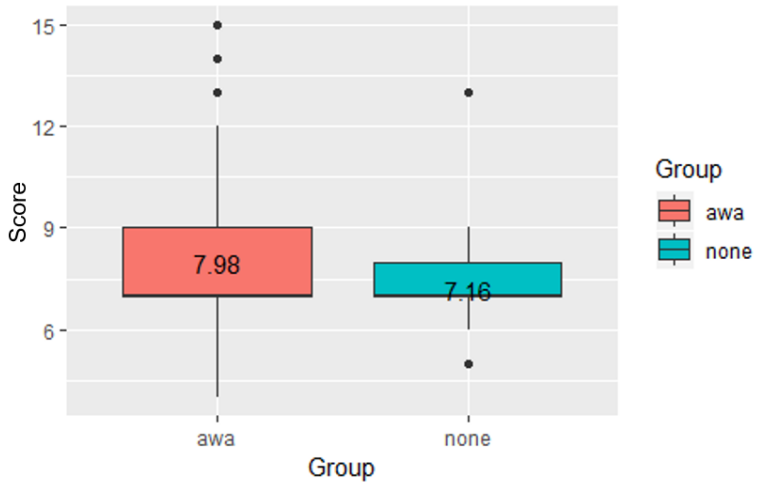


Figure 35: Scores of students' improved essays across comparison groups in iteration 3

The AWA feedback group obtained higher scores for the revised essay on average (M = 7.98, SD = 2.49, N = 45) than the No feedback group (M = 7.16, SD = 1.21, N = 43).

Welch two sample t-test showed a difference between groups: $t(64) = 1.96$, at a $p = .055$ level, with a small observed effect (Cohen's $d = 0.41$, 95% CI [-0.02, 0.84]).

From the scores of students' valid improved essays ($N = 88$), performance bands were created as Improved, Degraded and Neutral, based on the lower and upper quartiles of score summary statistics. If the score was greater than 8, it was classified as improved, if less than 7, it was classified as degraded, and neutral otherwise. The proportion of students in the feedback groups falling in those categories is shown in Figure 36.

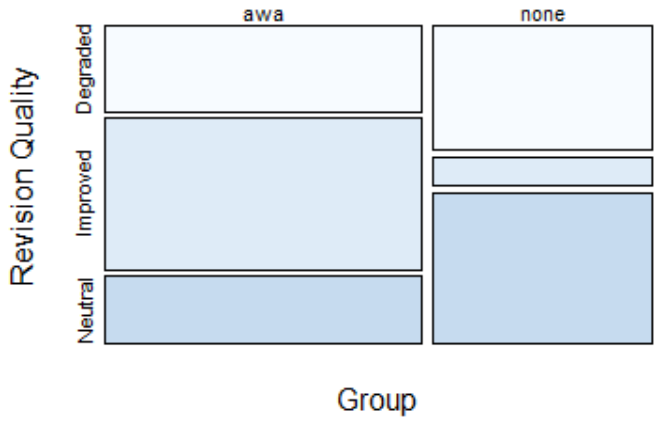


Figure 36: Proportion of students in awa and No feedback groups categorised by revision performance

A significant difference was found between the two groups' revision performance in a chi-squared test: $\chi 2(2) = 12.75$, $p < .001$. This effect can be attributed to the automated feedback from AcaWriter in the AWA feedback group which helped students to improve the essays using the suggested rhetorical moves to produce significantly more improved essays. Students in the No feedback group produced more degraded or neutral essays than improved essays.

4.4.5.2 *Rhetorical Moves*

Since the main goal of the task was to teach students how to write better essays with the use of rhetorical moves, the revised essays of students from both groups were analysed in terms of their inclusion of rhetorical moves to study their differences. This was done by running their final improved essays through AcaWriter and counting the rhetorical moves found in the essay. The number of rhetorical moves introduced in the improved essays compared across groups us shown in Figure 37.

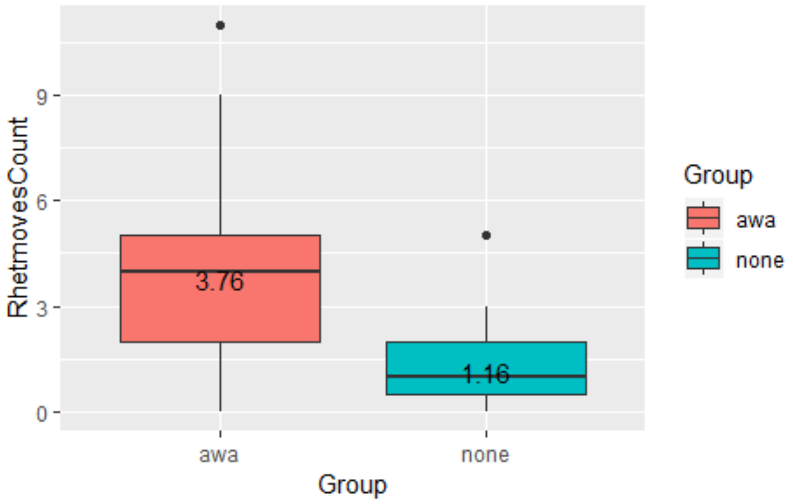


Figure 37: Rhetorical moves introduced in students' improved essays across comparison groups in iteration 3

Students in the AWA feedback group introduced a statistically significant higher number of rhetorical moves (M = 3.76, SD = 2.8, N = 45) than students in the No feedback group (M = 1.16, SD = 1.02, N = 43) in their improved essays: $t(56) = 5.81$, $p < .001$. The effect size was large (Cohen's d = 1.22, 95% CI [0.75, 1.68]). This indicates that using the automated feedback from AcaWriter helped students to understand the use of rhetorical moves mapped to their assessment criteria, and to apply them to the given text for improvements. These rhetorical moves also had a moderate positive correlation with the scores discussed earlier: r = .46, n = 88, $p < .001$, indicating a relationship between the presence of rhetorical moves and essay quality.

4.4.5.3 *Revision Metrics*

To study the revisions made by students across the two groups, metrics trialled in iteration 2 were first applied on the improved essays of students. There was a moderate positive correlation between the simplest word count measure (change in the number of words) and the score: r = .56, n = 88, $p < .001$, indicating that students who added more words scored more on their improved essays. The change in word count across the two groups is shown in Figure 38.

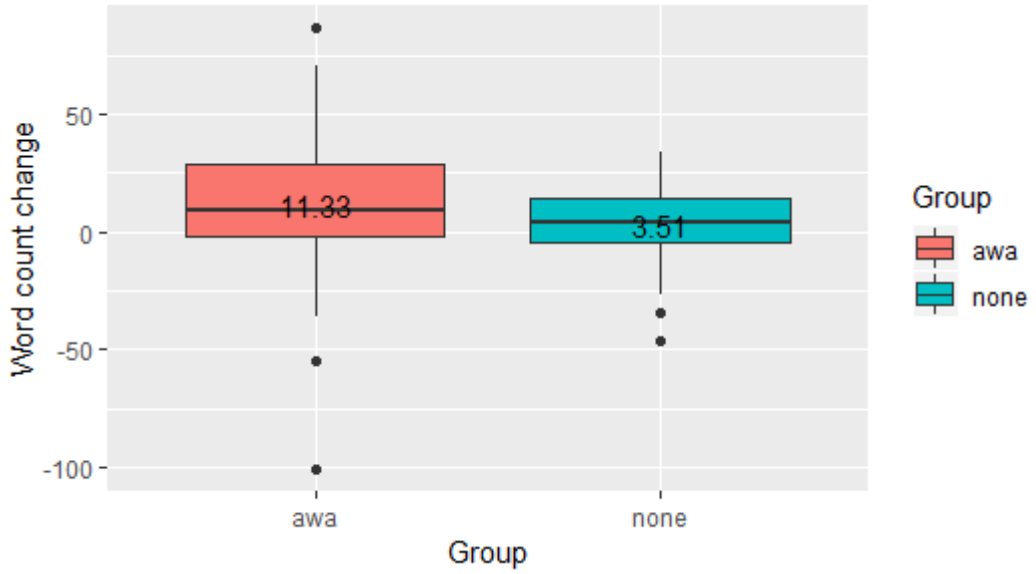


Figure 38: Change in word count across groups

The mean word count change was 11.33 for AWA feedback group (*SD* = 32.2), and 3.51 for No feedback group (*SD* = 17.2). The actual values of word count ranged from -191 to 379 for AWA feedback group and 246 to 326 for No feedback group. However, no significant differences were found between the groups in a Welch's two sample t-test: $t(67) = 1.43$, $p = .16$.

The next measure was the cosine distance, which did not have any relationship with score this time: $r = -.14$, $n = 88$, $p = .18$. The mean cosine distance was 0.98 in AWA feedback group (*SD* = 0.02) and 0.99 in No feedback group (*SD* = 0.01) as shown in Figure 39. No significant differences were found between the groups: $t(68) = -1.8$, $p = .07$; a small effect size was observed (Cohen's $d = -.38$). The reduction in the number of words in the text provided to students for revision resulted in the cosine distance measure being unable to account for small changes made at the document level. This shows that cosine distance can only be used when there is a considerable number of words in the text for analysis. Because the above-mentioned revision metrics are calculated at a document level, they do not capture nuanced changes made by the students in individual sentences from the small example text provided to students for the revision exercise. This led to the development of more detailed metrics by studying sentence-level changes in the text.

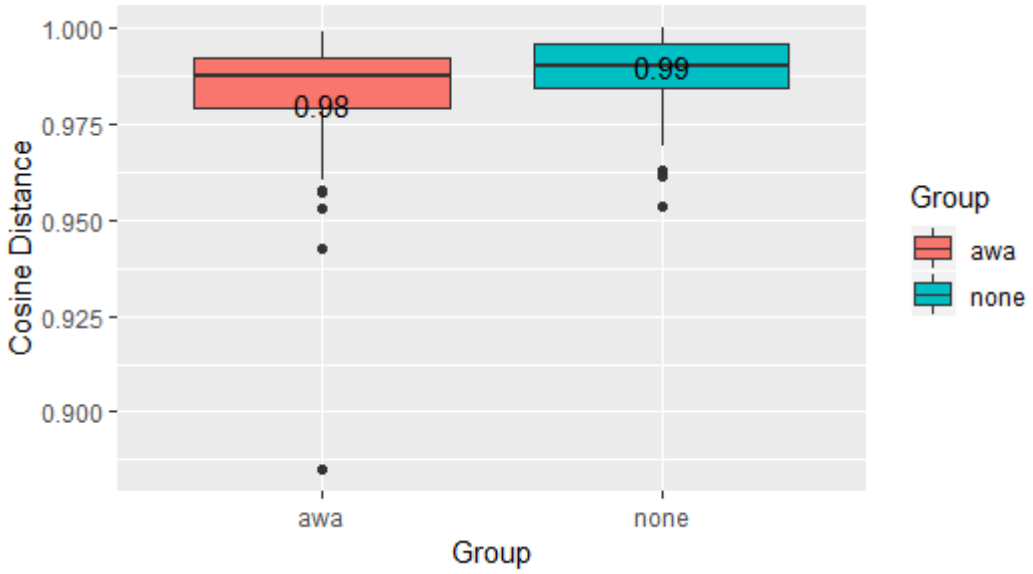


Figure 39: Cosine distance of improved essays compared to the given essay across groups

4.4.5.4 *Automated Revision Graph for Studying Sentence-Level Revisions*

In addition to the document-level revision metrics studied earlier, a version of automated revision graph was developed to study revisions at a sentence level (recall that this was first introduced as a manual analysis in Section 4.2.5.5). This helps quantify the revisions made further by comparing the original sentences in the given text with the revised sentences in the improved essay of students. The first step in this analysis was to construct the revision graph for the improved essay using similarity metrics and rhetorical moves.

The revision graph was developed in a Jupyter notebook in Python 3.5 for automated revision analysis at sentence level. The inputs are the two texts to be compared (text 1 and text 2): in this case, the given essay which is the same base text for all students (first column of the revision graph), and individual students' final improved essays (last column of the revision graph). The nodes of the graph represented as circles denote individual sentences presented in the order that they appear in the texts. To understand how the revision graphs represent changes made to the text, we now need to take a closer look at an example in Figure 40 with additional descriptions.

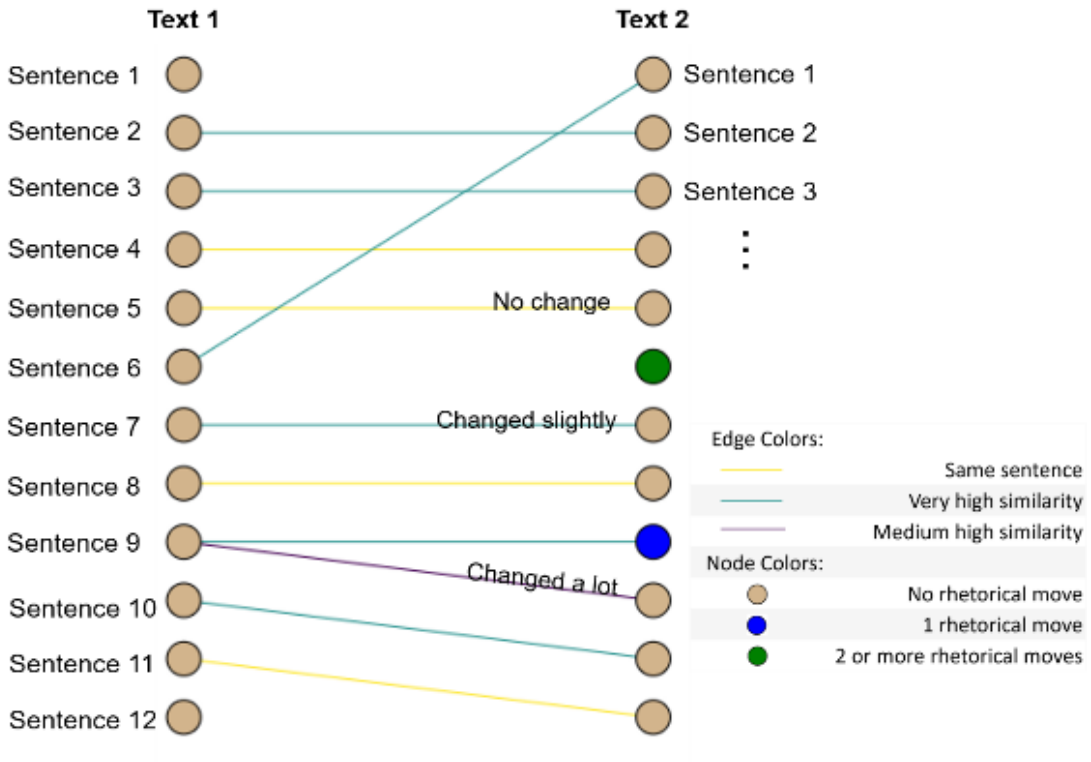


Figure 40: Revision Graph with descriptions using a sample improved essay (text 2) compared to the given essay (text 1)

In the example shown in Figure 40, both the given essay on the left and the improved essay on the right contain 12 sentences each, represented as nodes. The colour of the node represents a measure of the text feature we are interested in – in this case, the presence of rhetorical moves in a sentence, since that was the focus of the revision task for students. Text 1, which is the given essay in the exercise did not have any rhetorical move in its sentences, so all the nodes are light brown in colour. This default colour of the node is changed in text 2 of the revision graph if the student introduces one or more rhetorical moves in a particular sentence. The node is blue if there is one rhetorical move, and it is green if two or more rhetorical moves are present in the sentence. In the sample revision graph shown above, sentence 6 of the improved essay (text 2) has two or more rhetorical moves, and sentence 9 has one rhetorical move. To investigate what the actual sentence was, a mouse hover can be made over the node in the html version of the graph as shown in Figure 41.

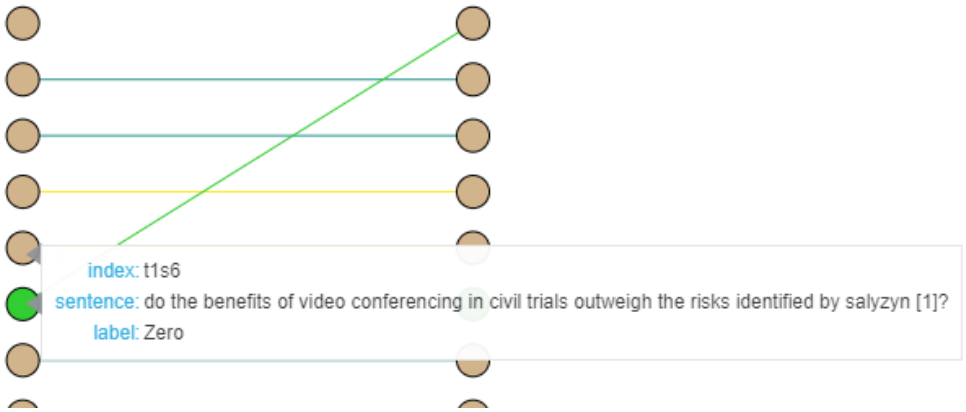


Figure 41: Html version of the revision graph showing the actual sentence for reading, on mouse hover over the node

The edges represent how similar the sentences are in text 2 compared to the sentences in text 1, showing the changes the student made to the text, and where they came from. If there is no outgoing edge from a node in text 1, it shows that there is no sentence in text 2 similar to it, which means that the sentence has been removed from text 1. For instance, in the sample provided in Figure 40, the two sentences in the given essay, sentence 1 and sentence 12 have been removed by the student in their final improved essay. The yellow coloured edges from text 1's nodes indicate that the two nodes are exactly the same, meaning that the no changes were made in that sentences. Examples from the sample revision graph in Figure 40 include sentences 4, 5, 8 and 11 from text 1. The light blue/ teal coloured edge indicates that the two nodes are highly similar, which means that the sentence has been altered, but to a small extent (Examples are sentences 2, 3, etc. in text 1 of the sample graph). The purple coloured edge represents medium similarity, meaning that the sentence in text 2 has evolved from another sentence in text 1, but has changed a lot. So, the darker the edges are, the more changes the student has done to the given text. Students might also introduce a new sentence that is not very similar to any sentence provided in the given text. Such newly added sentences will not have any edge from text 1, E.g. Sentence 6 of text 2 in Figure 40. Sentences in the revised text can come from a combination of sentences in the given text which will show two or more incoming edges in a node in text 2 of the revision graph. In contrast, there can also be cases where one sentence from the given text is split into two sentences in text 2 (see Sentence 9 of text 1 in Figure 40 showing two outgoing edges).

Construction of the revision graph involved several steps, making use of technical elements from previously available Natural Language Processing models and graphical visualization packages in Python. Figure 42 provides an overview of the steps - a detailed explanation of how the steps were implemented is given next:

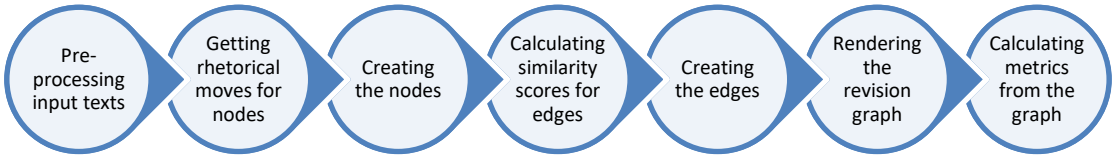


Figure 42: Steps involved in the automated construction of revision graphs

Step 1: Pre-processing the input text files, and sentence parsing

The first step in the automated construction of the revision graph was to pre-process the input files to get them in the desired format. The input files in our case were html files because they originated from Json files stored in the database, which were then converted into html files. The html style elements were stripped, and the text was extracted from the input html files. This text was cleaned further to remove unwanted characters and converted to lower case. Finally, the text was parsed to sentences using the Text Analytics Pipeline (TAP) API call. TAP is an open-source software that provides text analytics services using a GraphQL API (More details at: https://github.com/heta-io/tap). This step did not use the open corenlp sentence parser to make sure that the sentence parsing is consistent with the rhetorical move parsing later done by TAP. Using the cleaned text, overall metrics of the revised text like sentence count, word count etc. were calculated.

Step 2: Getting rhetorical moves for all sentences

The next step was to call the Text Analytics Pipeline with a rhetorical moves query to get the rhetorical moves present in each sentence. This information is used to colour the nodes of the revision graphs, as it is the main text feature we are interested in studying in students' improved texts.

Step 3: Creating the nodes

By making use of rhetorical moves from the above step, the nodes of the revision graph were created. A nodes csv was created with three columns: an index for each node to indicate the sentence of the text it denotes (e.g. t1s1 where t1 is text1 and s1 is

sentence 1), the actual sentence's text, and the node category for specifying the colour (zero for no rhetorical moves, one for one, and more for two or more rhetorical moves).

Step 4: Creating text vectors and calculating similarity scores

Now that the nodes of the graph are created, the edges need to be created based on how similar the sentence in the revised text is corresponding to sentences in the previous text. Finding this similarity was made possible using a cosine similarity score (Note that this score is similar to the Cosine distance measure discussed earlier, but instead of calculating how different the texts are in the former, the latter score calculates how similar they are). To calculate this score, the text was converted to a vector, and compared with the given text to get the score. Each sentence in the revised text was compared with every sentence of the given text to calculate the similarity scores.

Step 5: Creating the edges

Based on the similarity scores calculated above, edges for the revision graph were created between the nodes of the given text and the revised text based on thresholds. If the similarity score was equal to or greater than the highest similarity threshold (0.99 for the same sentence), an edge was added between the nodes of the two sentences with a weight of 1. If the similarity score was equal to or greater than the next highest similarity threshold (0.80 for highly similar sentences), an edge was added between the nodes of the two sentences with a weight of 0.8. The same process was repeated for medium similarity nodes (0.6) with an edge created between the nodes and a weight of 0.6. Values for these three columns: startnode, endnode, weight were appended for each edge to form the edges csv.

Step 6: Rendering the revision graphs

The next step was to create and render the revision graphs using the nodes csv and the edges csv created earlier. This was done using network graphs from a python library called 'HoloViews'. HoloViews is an open-source Python library (https://github.com/pyviz/holoviews) that provides the ability to represent and visualize graphs easily with facilities for interactively exploring the nodes and edges of the graph, using the bokeh plotting interface. It helps us to explore the actual sentences further by hovering upon the nodes of the graph. The rendered revision graphs were also saved as html files in the specified output folder.

Step 7: Calculating metrics

From the revision graph generated, features of the graph were extracted to come up with metrics that quantify some of the changes made by students at sentence level. This final step aggregated details from the revision graph to calculate metrics of the nodes and edges by counting the number of rhetorical moves in each node, counting the similar edges etc. - more details are provided later with examples.

The revision graphs help in studying the different kind of changes students make at a sentence level on the given base essay, which is the same for everyone. They can visually represent the revision actions like minor changes, major changes, additions and deletions made in the sentences of the given text, and the presence of rhetorical moves in the revised texts. There were many kinds of revision graphs that emerged from the improved essay of students: ranging from very few changes made and no rhetorical moves introduced in the revised essay (A), to major revisions made to the given text and introduction of rhetorical moves in many sentences (C) as shown in Figure 43.

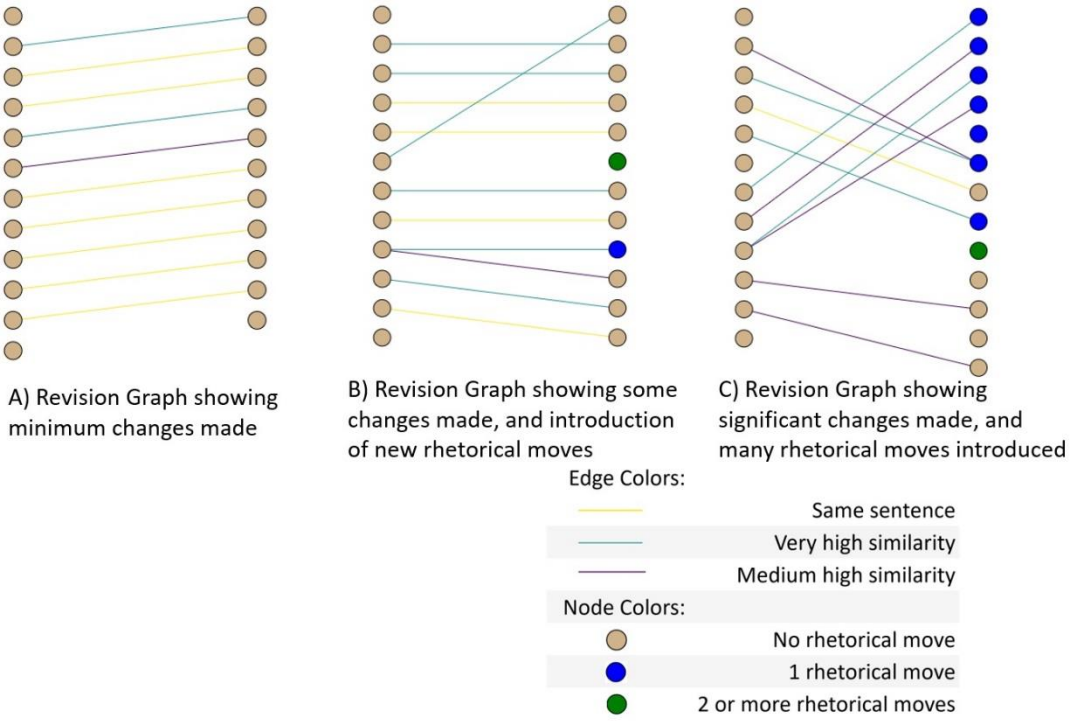


Figure 43: Sample revision graphs demonstrating different kinds of revisions made by students at the sentence level to improve the given essay

To build revision graphs and calculate metrics for all the students' revised essays in iteration3, the steps in Figure 42 were repeated in a batch processing script by

iterating them over a folder containing all the improved essays of students. This was done using a module that runs an iterative process to analyse all the improved essays of students present in a folder and store the measure of the nodes and edges of their individual revision graphs as numerical indicators in a csv file. The metrics from the revision graphs included the following:

- sentcount – number of sentences in the revised text (count of nodes in the text)
- zerorhetnodes – number of nodes (i.e, sentences) in the revised text which did not have any rhetorical move
- onerhetnodes – number of nodes in the revised text which have one rhetorical move
- morerhetnodes – number of nodes in the revised text which have two or more rhetorical move
- rhetnodepresence – a composite measure from the two indicators before which counts the number of nodes in the revised text having at least one rhetorical move
- sameedges – number of edges which show absolute similarity between the two nodes showing no changes made
- highsimedges – number of edges which show high similarity between the two nodes showing minor changes made
- medsimedges – number of edges which show medium similarity between the two nodes showing major changes made

The examples provided earlier will have the metrics calculated as following (shown in Table 5) based on their revision graphs:

Table 5: Sample metrics for the revision graphs in Figure 43

Revised Text	**sent count**	**zero rhet nodes**	**onerhet nodes**	**morerhet nodes**	**rhetpre sence**	**same edges**	**highsim edges**	**medsim edges**
Sample A	11	11	0	0	0	7	2	1
Sample B	12	10	1	1	2	4	6	1
Sample C	13	5	7	1	8	1	4	5

Revision Graph metrics to study revision performance (quality)

From the revision graph metrics shown above, correlations to score and other metrics were calculated (given in Figure 44, visualized using the R *corrplot* package). The presence of rhetorical moves (number of sentences containing at least one rhetorical move) was positively correlated to score (r = .57), but no other metrics had strong relationships with score. The rhetpresence metric also had a strong negative correlation (r = -.59) with the sameegdes measure which shows the number of edges which did not have any changes made from the given text. This rhetorical presence metric is different to the total number of rhetorical moves measure discussed earlier in Section 4.4.5.2, because the total number measure counts the moves at the overall document level (this includes two or more moves in a sentences as separate moves), while the presence counts at a sentence level (counting one or more rhetorical moves in a sentence as '1', thus measuring the quantity of changes affecting more sentences.

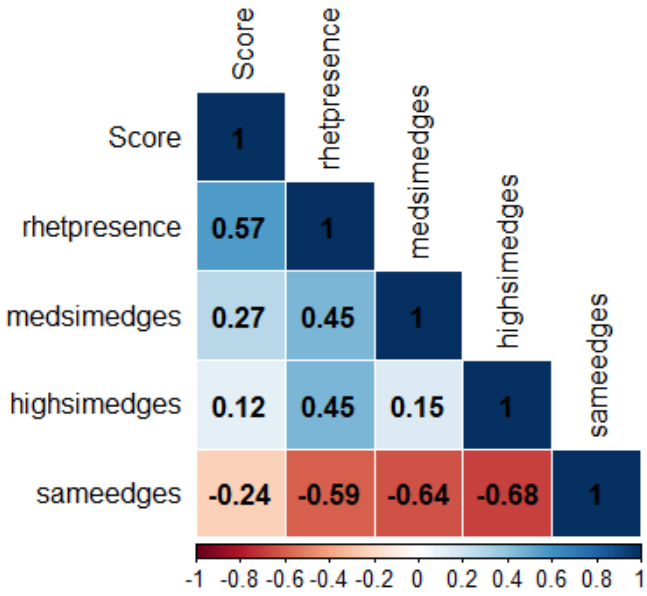


Figure 44: Correlation of revision graph metrics to scores and other metrics (For description of metrics, see Section 4.4.5.4)

While these metrics are not absolute indicators of quality of the revision in terms of the actual content, they are still helpful in quantifying the changes made by students to the given text. In this way, they help us study students' engagement with the writing task, and the revisions impacted by automated feedback from AcaWriter, even without having a human marker score the texts for improvements.

Within the same automated feedback group, it was also observed that the revisions made by students were somewhat different. This might be because of the way students engage with automated feedback and apply it on their writing. To explore the different kinds of revisions made by students within the automated feedback group, their engagement with AcaWriter feedback was studied.

4.4.6 RQ2a) Studying student engagement with automated feedback

To further study students' engagement with automated feedback, student interactions with automated feedback and revision stages were analysed from students of AWA feedback group. This helped answer RQ2 by studying the behaviours of different users while interacting with automated feedback in terms of how often they requested for feedback, and how they applied this feedback to make revisions on their writing. First, the interactions with feedback across the students were studied using their feedback requests, followed by a detailed examination of their stages in the revision process using a version of revision graphs.

Studying student interactions with automated feedback based on feedback requests

From the 45 students who belonged to the AWA feedback group and wrote valid revised essays, varied interactions were found in terms of their feedback requests. AcaWriter recorded the instances when students clicked on the 'Get feedback' button by saving their text at that point in time as a log, which is analysed as follows. Table 6 below shows the summary statistics of the frequency of students automated feedback requests.

Table 6: Summary statistics of the frequency of automated feedback requests

N	Mean	SD	Min	Max	Median
45	6.13	3.06	1	14	5

The number of times students requested for feedback to improve the same base text varied widely as denoted by the standard deviation, min and max values in the summary statistics. This measure of students' interactions with automated feedback was split into three categories based on the quartiles of the frequencies: highInteraction if frequency was greater than 8, lowInteraction if frequency was less than 4, and moderateInteraction otherwise. The distribution of interaction categories among the 46 students who used automated feedback to improve the given text is provided in Figure 45.

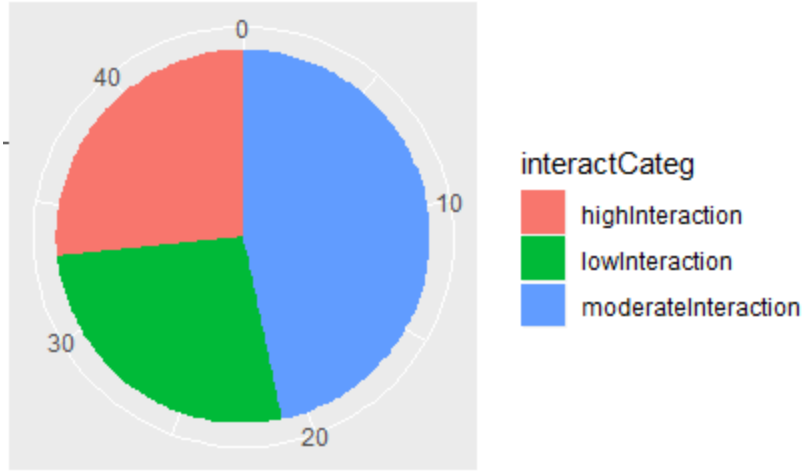


Figure 45: Distribution of interaction with automated feedback across students

From the figure above, we can observe that most students had a moderate number of interactions with AcaWriter i.e, requested for feedback about 4 to 7.75 times. This

interaction category when compared with the revision performance category discussed earlier is shown in Figure 46.

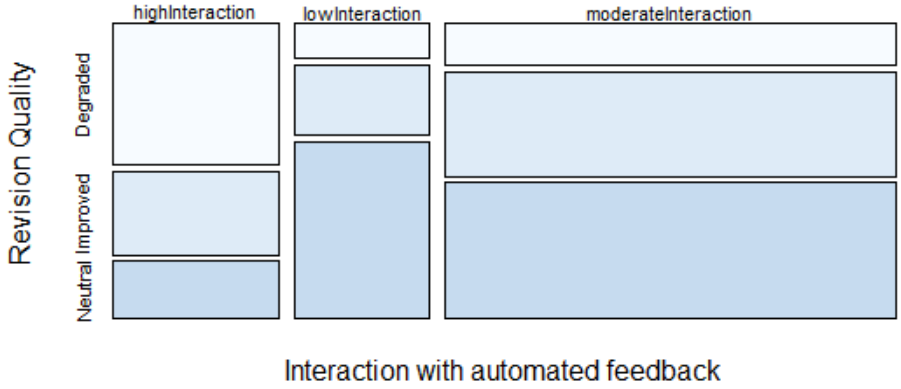


Figure 46: Plot of Interaction with automated feedback Vs Revision quality

From Figure 46, it can be seen that most students in the 'Neutral' revision quality category had low interactions with automated feedback. This might be because they were not very involved in the task and hence not did not make changes to the text by requesting feedback. Students who had moderate interaction with automated feedback were mostly in the improved or neutral category. On the other hand, many students having high interaction with automated feedback degraded their texts, which might be because they try to get all rhetorical moves by playing the system rather than actually trying to improve the quality. However, these observations were preliminary and no significant effects were found in a chi-square test: X-Squared (4) = 4.79, p = 0.31.

Multi-stage revision graphs to study engagement and revisions with automated feedback

In addition to the lower level frequencies of student interaction with automated feedback, we can also study students' stages in their revision using the revision graphs in much detail. To do this, a version of the revision graphs explained earlier, but comparing multiple drafts over time called *'Multi-stage revision graphs'* were created. Using these revision graphs, we can study the underlying processes that were involved in the stages of revision of the text after receiving automated feedback.

The multi-stage revision graphs were useful to observe how students engaged with automated feedback in different ways and applied the feedback on to the writing to make revisions to the given text in many ways. As we saw earlier, students had varied interactions with AcaWriter in terms of the number of times they requested for

feedback and were categorised into High interaction, Low interaction, and Moderate Interaction categories. In this section, we will take a look into their stages of the revision process to study these interactions in more detail, and see how they led to improved, neutral, or degraded texts.

First, the revision process involved in some students who had high interaction with automated feedback is described. Figure 47 displays multi-stage revision graphs of three students who had high interaction with automated feedback but scored differently on their revised essay. These revision graphs are similar to the previous version of revision graphs explained earlier, but instead of comparing only the initial and the final text, it also compares the many drafts in between them. The base text given that students start with is the same for all students. Each draft was stored when the student clicked on the 'Get Feedback' button in AcaWriter.

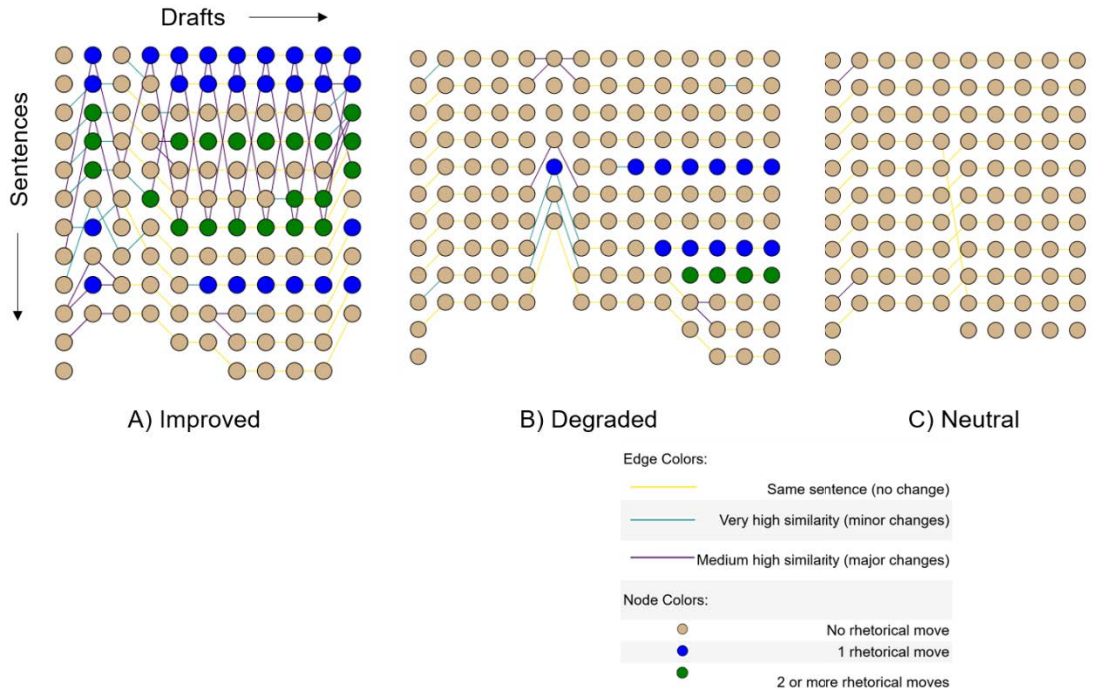


Figure 47: Multi-stage revision graphs of students who had *high interaction* with automated feedback and their score categories

As shown in Figure 47, the first revision graph (A) with an improved text's revision stages shows high levels of engagement, and the student taking into account the tool's feedback to include rhetorical moves from the very first draft created. A number of changes have been made in many sentences as indicated by the blue and purple edges, and the student has requested for feedback 10 times to see if the changes

have been detected by the system as a rhetorical move. We can also see that the student has moved around the sentences to change their positions in the paragraph, and split some sentences into two. Some sentences in the text have similarities between them as indicated by the crossing edges with the previous draft – this shows internal cohesion in the text. In the final text, the student has added rhetorical moves in 7 sentences, and overall the effort shown by student to revise the text to high quality is evident leading to an improved essay. In the revision graph B, we can see that the student has interacted with automated feedback but did not make significant changes, other than tweaking some sentences to get the rhetorical moves highlighted by the tool. The student has also removed some key sentences, leading to a degraded essay. This might be because the student tried to get the rhetorical moves as a priority rather than focusing on improving the quality of the given text as a whole. In the later stages of the revision, the student has not made significant changes even after requesting for feedback. The student who got a neutral score for the revised essay has revision graph C shown in Figure 47, and it shows that even though the student requested for feedback many times, he/she did not make much changes to the text in response to the feedback. Most sentences remained the same as shown by the yellow edges leading to a neutral score.

Students who had low interaction with automated feedback also engaged with the feedback in different ways, and brought about revisions which led to improvements, degradations and neutral scores as shown in their multi-stage revision graphs in Figure 48.

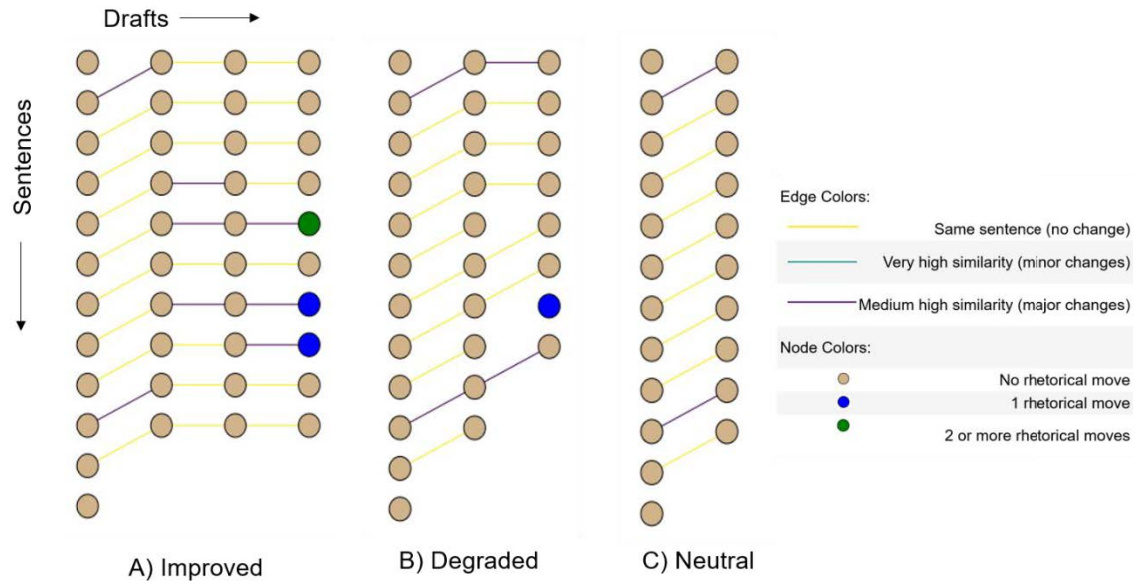


Figure 48: Multi-stage revision graphs of students who had *low interaction* with automated feedback and their score categories

The revision graph A in Figure 48 shows how the student requested for automated feedback only three times, but still managed to take into account the feedback and apply it to the text by making major changes and including rhetorical moves in three sentences leading to an improved essay. The degraded essays' revision graph shows that the student had removed some key sentences in the given text. The feedback from the tool has not been taken into account much to revise the text for including rhetorical moves. These factors have led the student to degrade the given text, even with the use of automated feedback.

Students who interacted with automated feedback a moderate number of times also came up with improved, degraded and neutral essays depending on their engagement with the feedback, and their revisions in response to it. Figure 49 shows samples of improved, degraded, and neutral essays with their revision graphs, illustrating the stages of revision and interactions with automated feedback in each case.

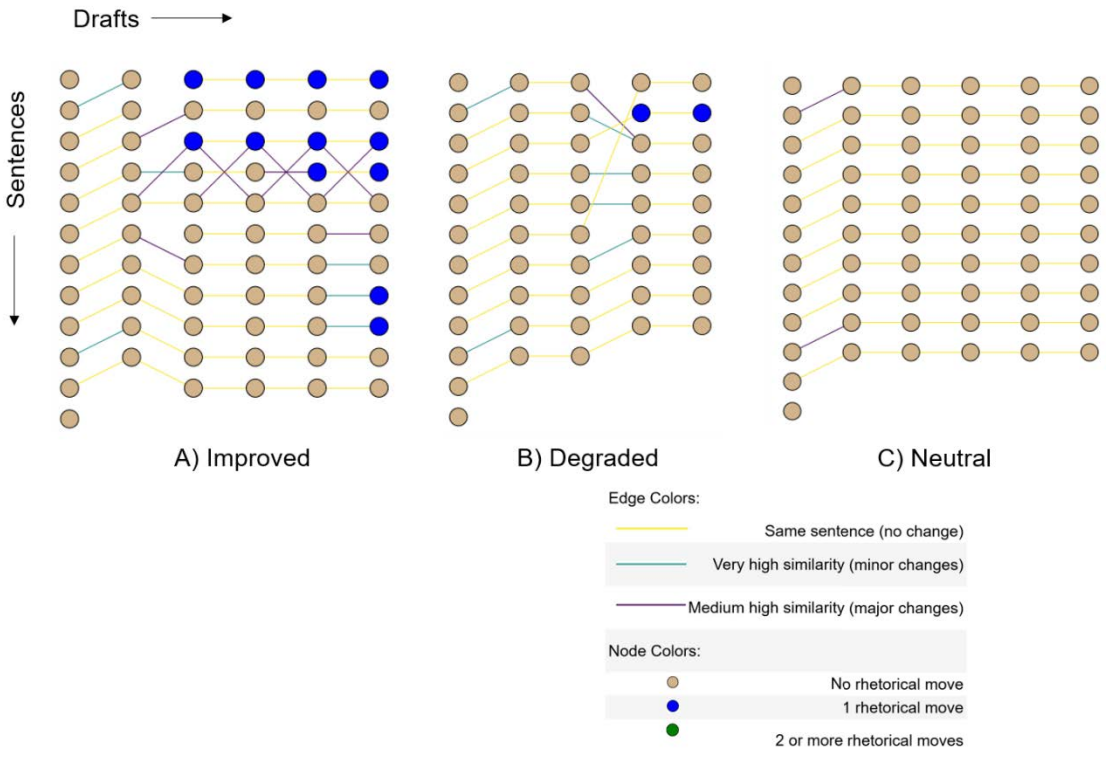


Figure 49: Multi-stage revision graphs of students who had *moderate interaction* with automated feedback and their score categories

In revision graph A showing the stages of revision of an improved essay with moderate interaction in Figure 49, we can see that the student made considerable changes to many sentences in the given text. The student has applied the feedback and introduced rhetorical moves in the sentences from one draft to the next, and validated it by asking for feedback. In the final improved essay, the student has used rhetorical moves in five sentences by making changes to the given text. Some sentences (E.g. 3 and 5 in the final text) also have internal cohesion by containing similar words. We can observe that at each step after receiving automated feedback, the student has made some changes to the text showing engagement with feedback and its application to the text. In the revision graph B, although the student has made some changes to the text, it led to a degraded text. The student has also included one rhetorical move, but only by making a tiny change showing that the move was added just to cater to the feedback message, and this might be a reason for the degradation. In the neutral scored essay's revision graph at C, very few changes were made after the first request for feedback, and none afterwards. This lead to a revised essay which was almost identical to the given text, leading to a neutral score for changes made.

From the multi-stage revision graphs discussed above with examples, we can observe that students undertake different behaviours while engaging with automated feedback and revising their texts. To our knowledge, there has been no prior work exploring students' engagement with automated feedback and how they apply the feedback on their writing to revise it in this level of detail. This contributes to the body of research on automated writing feedback to study how different students apply it on their writing, and hence study how it impacts students with different levels of engagement.

4.4.7 RQ2b) Student engagement with automated feedback scaffolded by peer feedback

To study student engagement with automated feedback using additional scaffolding for RQ2b, the first scaffolding explored was the use of peer feedback and discussion in iteration 3. This included studying student perceptions of peer discussion in combination with automated writing feedback from their survey responses, and a content analysis from sample peer discussions to learn more about the actual content of their discussion.

4.4.7.1 *Student perceptions of peer discussion*

To study the impact of peer discussion with automated feedback, student perceptions of peer discussion were first analysed. This analysis was based on student responses in the feedback survey where they answered questions on what aspect of peer feedback they liked, and what they did not like so much based on their engagement in peer discussion in this activity. Student perceptions of peer feedback and discussion included the following:

Many students mentioned that the peer discussion helped them gain other perspectives on the essay, and find new ways to approach the task:

> *"Can read from someone else's perspective and see how someone else will perceive your writing, and how this is different from my own perspective."*

> *"Good to hear other people's perspectives and find news ways to approach the task."*

> *"I find hearing different perspectives very helpful, especially in essay writing as it is good to have a fresh pair of eyes and insight."*

"Peer feedback is always helpful because it helps to understand things from a different perspective."

Students also thought that the peer discussion helped them to learn from each other as it made them aware of additional perspectives:

"My peer discusser helped me identify things that I had skimmed over and missed."

"I miss little details when editing by myself. So to have a peer edit it, they pick up on the little things that I missed/ errors."

It helped them interpret and clarify the purpose of the task and the use of the AcaWriter software tool:

"Clarifying the task itself, and how to use the software"

"Provides different interpretation and perspective"

They thought that the discussion helped them to reflect on their thoughts and discuss it with their peers to understand their opinions as well:

"Discussing with my peer allowed us to mutually agree with the issues that we found with the paper, and confirmed my thoughts and concerns."

"It was really helpful to see what other people thought of my words, maybe it gives some insight into the markers mind when someone that isn't you is reading your words."

"It's nice to have a sounding board and encourage reflection."

A student also mentioned about the content of the discussion saying that the automated feedback from AcaWriter gave them ideas to discuss on in relation to the given essay:

"The automated feedback we were given gave us key points to discuss and give feedback on."

Furthermore, students noted limitations in peer discussion, which are discussed next. Students thought that the value and depth of the feedback and discussion they had depended on the ability of the peer. They thought that some peers did not have the expertise or knowledge to provide useful feedback, and were hesitant to criticise.

"It is going to vary on the peer and their ability to convey knowledge. If you have an absolute gun that is great at teaching to provide feedback, then it is going to

be way more beneficial than with somebody lacking an understanding and/or inability to communicate what they know."

"They are not often as harsh or knowledgeable as a tutor"

"People don't want to criticise, pick too much at the essay"

"The feedback not as in-depth whereas the AI technology recognised specific points of improvement in my writing."

They said that that the writing styles of the students might be different, leading to conflicting ideas or disagreement in the discussion:

"Everyone has their own style of writing, and it is difficult sometimes to mesh two different styles. People also might have different attitudes towards a topic, as was described in the introductory video, which could cause conflicting ideas in an essay."

"Everyone thinks and write differently and it takes time."

"Everyone has different styles of writing and opinions so it can be difficult to discuss when we have somewhat different goals."

Some students also thought that the discussion would be more engaging if they had different essays to share with each other. Future work could investigate this potential by providing students an option to engage in peer discussion on their own writing from assignments where they work on different topics.

"Because it was the same essay, the adjustments were very similar. Perhaps different texts would be more engaging to discuss."

"Discussion is quite limited as we both changed similar aspects of the extract."

4.4.7.2 *Qualitative analysis of the content of peer discussion*

To further study the actual content of peer discussion i.e., the kinds of discussion that students have with their peers during the task in detail, volunteers were requested for observation by the researcher. Students who volunteered for this study were asked to move to a separate room. A video recorder was placed over the shoulder for each group to record their conversations and interactions as shown in Figure 50.

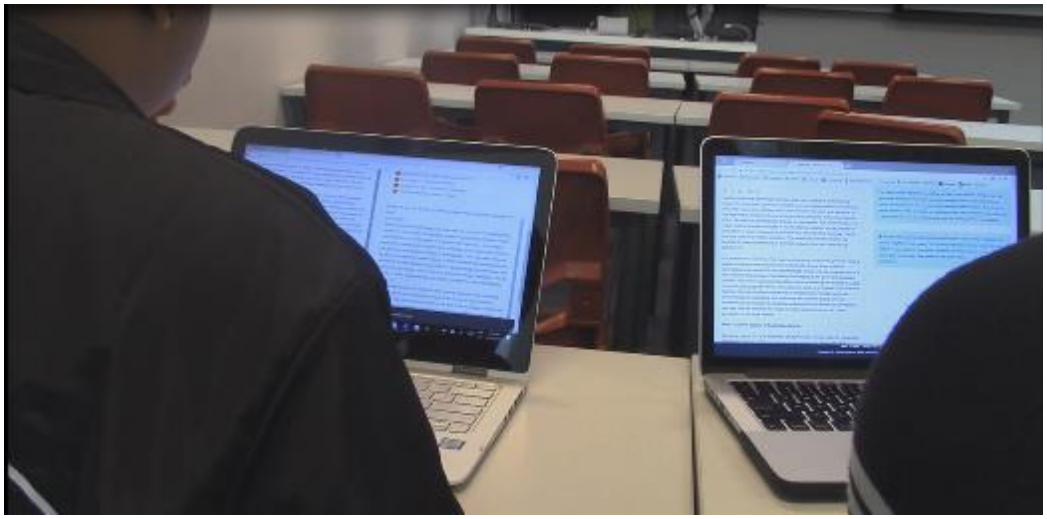


Figure 50: Data capture using over the shoulder video recording for capturing peer discussion

The recording was started after students completed the revision task individually, and when they were asked to discuss with their peers on how to improve the provided sample essay further. Three groups who received AcaWriter feedback were video recorded and transcribed for instances of engagement with automated feedback and revisions in the given sample writing. After this discussion, students were asked to go back to the task and make any additional changes if needed for the final submission of their revised essay. Participant details are provided below:

P1 – Student 1 of Group 1 (Male)

P2 – Student 2 of Group 1 (Female)

P3 – Student 1 of Group 2 (Male)

P4 – Student 2 of Group 2 (Female)

P5 – Student 1 of Group 3 (Male)

P6 – Student 2 of Group 3 (Female)

R – Researcher who was present in the room

The conversations were generally sparse due to the limited time allotted for this particular part of the task (10-15 minutes) during their tutorial hour. From the three groups that had their actions recorded during peer discussion, instances of engagement with automated feedback are discussed below:

Suggesting improvements referring to automated feedback

Students suggested improvements to be made to the given essay to their peers by pointing out possible revisions that can be made, and included references to automated feedback. This shows how students have engaged with their writing and their peers aided by automated feedback from AcaWriter. In the example shown in Figure 51, the student from Group 3 uses a reference from the feedback message from AcaWriter which is to use the phrase "In this essay, I will discuss" to explain how the essay can be improved to the peer.

P6: yeah are you supposed to put a question in the introduction?
P5: that's true
P6: I don't do that
P5: I don't do that either
(both laughing)
1:20 P6: and like "In this essay, I will discuss"
(puzzled look)
P6: When you are writing an essay
P5: Definitely, I
P6: It's what ..... to do
1:36 P6: (looking at her screen) In this essay, the reason is
do you feel like talk about the .. like how we were gonna set it out
like I'd just say, for these reasons
P5: I'll write.. erm
1:45 P5: this essay will discuss as a thesis statement, so it is clearly articulated in the talk
P6: cool
1:52 P5: And I say reasons why
P6: yeah

Figure 51: Excerpt of peer discussion demonstrating suggestions using automated feedback in Group 3

Similarly, in Groups 1 and 2, students are referring to automated feedback when suggesting phrases like "in summary', and "it is important to note" to their peers as shown in Figure 52 and Figure 53.

P1: oh.. So we revise
P2: Change the whole thing
P1: get more orange
P2: get more orange (laughing)
P2: yeah, yours is pretty ?..
P1: I need more green
P1: In summary (typing that on screen adding "in summary" to a sentence)
P2: Ultimately
P2 (looks at P1's screen, highlighted text)
P2: what?
P1: the first two paragraphs of content

Figure 52: Excerpt of peer discussion demonstrating suggestions using automated feedback in Group 1

P3: I didn't get it
P4: The current stance of… uh.. No, not that one, the one before it
(P4 goes back to her screen to check something)
P4: what about the integrity of the law
(P4 looking at P3's screen)
P4: ya, alter examination, there (pointing to screen)
P3: ah that specific section
P4: yeah, i think that's it, but I'm
…..
P3: I think that I should get rid of that "It is important to note" (laughs)
P4: nah i don't think so, this is like that however

Figure 53: Excerpt of peer discussion demonstrating suggestions using automated feedback in Group 2

Showing their interactions with automated feedback to the peer

Students shared with their peers the changes they made in their text, and the interactions with automated feedback. Excerpts from Group 1 and Group 2 are shown in Figure 54 and Figure 55. They seemed to share this to their peers to get clarification on the feedback, or to simply show the other what they attempted to change in the given text. This allows students to learn from others' work and view different perspectives in approaching the same problem.

P2: (points to P1's screen) mine says.. S is summarising.. Is there a reason why it is (pointing to his screen S highlighting)
so i was trying to change that. But it cant. Coz its
P1: It's a half summary
P2: yeah
P2: Its because it is examining.. (pointing to his screen)
(both looking at their own screens)
P2: Did you add that yourself?
P1: which one? (looking at her screen)
P2: This … discussion
P1: I should've,.. probably
P2: yeah, you did
P1: that whole part 1, I didn't touch it at all
P2: yeah, same

Figure 54: Excerpt of peer discussion demonstrating the sharing of their interactions with automated feedback in Group 1

(swapping laptops to show their revisions to each other)
P4: Oh I just changed, I don't know
P3: It's better, what I think about. It just makes it
P3: see what you got with this … (points to screen)
P4: yeah. I think it's good
Feel like its a personal writing assistant
P3: ah I get it
P3: yeah
……
(both swapping their laptops to get their own laptops back)
P4: How come yours captures the changes but mine
P3: Because I've did those discourse markers, like I would change this to point (pointing to an example on his screen)
P4: oooh
P3: coz that's the whole work

Figure 55: Excerpt of peer discussion demonstrating the sharing of their interactions with automated feedback in Group 2

Interpretation of automated feedback

The peer discussion further facilitated students' interpretation of automated feedback as it allowed students to clarify with their peers what the feedback meant. The excerpts from Groups 1 and 3 shown in Figure 56 and Figure 57 show instances where students learn more about the tool, that it annotates phrases and identifies key markers in their writing. This helps students to make sense of the feedback and learn how to use the tool better to improve the given writing.

P2: so mine is very .. like it just kept saying I'm summarising. It makes sense , but I don't think I..
P1: I don't know, I got P, E, C in one sentence and I'm like
P2: What's P?
……..
P1: ...little bit of green, little bit of orange
P2: (laughing) Got a mix of colors
Okay, i think.. (reading the peer's text inaudible)
P2: (pointing to screen) Have you changed this?
P1: (leaning over to see her screen) erm, I just ran, like, say a little phrase to it
P2: (leaning over to see his screen) Can I see?
P2: Okay. (pointing to text on his screen above summary sentence) (reading the text ...inaudible)
P1: I think its annotated phrases

Figure 56: Excerpt of peer discussion demonstrating interpretation of automated feedback in Group 1

P5: this is actually pretty amazing
P6: (laughs)
P5: (pointing to awa highlighting in his screen) like it gives you every... the meaning of every sentence
P6: yeah
……..
P5: that's a yellow sequence for a reason
P6: (pointing to his screen)
P5: this is highlight yellow
P5: (asking R) What does the yellow say again?
R: so these are .. summary in green, orange
It has detected that you are re talking about some novel improvement
P6: oh
R: so what it does is, it picks up places where you have included markers that clearly show something to the reader
P5: oh
P5: I did very well for the first sentences, I'm very glad that I wrote.. It just goes.
P5: that's actually really interesting

Figure 57: Excerpt of peer discussion demonstrating interpretation of automated feedback in Group 3

Critiquing automated feedback

It was also observed that students from Group 1 critiqued the automated feedback while discussing changes with their peers. The excerpt in Figure 58 shows how a student critiques the tool that it did not register the change (recognise a rhetorical move with a tag), and how they tried to get the 'S' (summary) tag. This shows that the student had tried revising the text expecting a tag, but was disappointed to see that it was not picked up by the tool.

F (pointing to her screen) what did you change here? Let's discuss
P2: Did i change this? Yeah I did. This part, (pointing to P1's screen) the benefits of video conferencing, and it didn't even register it
P1: Oh there was a surprise, with the finding.. sorry
P2: (laughing) erm
P2: (reading from screen) this essay will discuss"
P1: (laughing) trying to get the S
P2: did you use in a whole sentence or is it actually set?
P1: uhm
P2: very hard and critical … i don't know why..

Figure 58: Excerpt of peer discussion demonstrating critique of automated feedback in Group 1

Helping with tool usage

There were instances where students provided instructional feedback in terms of how to use the tool, and what was expected in the task. The examples in Figure 59 and Figure 60 demonstrate students showing the peer how to get feedback from the automated tool. Even if this peer feedback doesn't help them improve their writing directly, it helps them make use of the tool better.

P3: (showing the feedback messages page) for me this doesn't make sense
P4: look at yours
P4: you fixed all
P3: (both looking at P4's screen) Did you press something? Get feedback? That's why
P4: I don't know, I haven't seen anything changed

Figure 59: Excerpt of peer discussion demonstrating peer helping with automated feedback tool usage in Group 2

P5: I don't get anything (pointing to feedback tab on the other screen)
P6: Did you edit anything? (shrugging & laughing)
P5: I changed it and then it said… oh yeah
P6: okay, alright. I change the graph. It's not coming up
let me show haha

Figure 60: Excerpt of peer discussion demonstrating peer helping with automated feedback tool usage in Group 3

In addition to the above themes of students' engagement with writing using automated feedback, students also discussed about writing in general. This included references to sharing their thoughts on parts of writing like heading organization and referencing which were not the focus of their current activity, but are important aspects of writing. A sample is shown in Figure 61.

P6: (pointing to her screen) do you set it out like this? like part 1, part 2 (looking at him)
P5: No i don't write like this
P6: yeah
P5: its like A, Big a, and then dot and then sub points little i
P6: yeah, okay
P5: but I don't make it to parts
P6: yeah, I don't think i can do it like that... referencing that there are extra developments basically

Figure 61: Excerpt of peer discussion demonstrating other writing-related discussions in Group 3

4.4.8 Discussion

Results of data analysis from this design iteration showed that student perceptions of the writing intervention were generally positive, and students appreciated the contextualized feedback on their writing from the new AcaWriter tool. There was a significant difference between the AWA feedback group and the No feedback group in terms of how students perceived the activity – AWA feedback group students found it more useful than the No feedback group. Students in the AWA feedback group also scored significantly higher than students in the No feedback group on their scored improved essays and introduced more rhetorical moves in their writing. These effects observed showed that the automated feedback from AcaWriter impacts student perceptions and their writing, and hence led to a decision to use automated feedback for all students without study conditions in future iterations so that some students are not deprived of its usage.

Document-level revision metrics like number of words and cosine distance were not able to differentiate the quality of revised essays and no significant difference were found between the groups, so sentence-level metrics were developed using revision

graphs. The first version of automated 'Revision Graph' was constructed by parsing the revised text and finding similarities with the given text at sentence level using text and graph analysis methods. Useful metrics from the revision graph like the number of sentences unchanged in the revised text and the number of sentences with the presence of rhetorical could be derived from the revision graph. Such metrics can help researchers quickly quantify the revisions made to a given text without having a human marker to read and assess it.

Student interactions and engagement with AcaWriter feedback were further explored for those who worked with automated feedback in the revision task. Most students had a moderate number of interactions with automated feedback based on their feedback requests, but their amount of interaction did not have a significant effect on the score of the revised essay. To explore this relationship further, multi-stage revision graphs were constructed using intermittent drafts of students writing after each automated feedback request. Within the groups of students who had high interaction, low interaction, and moderate interaction with automated feedback, varied score categories were found, and the interactions did not have strong effects on the score categories. However, findings from the multi-stage revision graph analysis show that the level of engagement with automated feedback, and the methods that students use to apply this feedback to revise the given text impact their quality of revision to an extent. An implication of the result is that these factors should be considered by writing analytics researchers when studying the effectiveness of automated feedback on students' writing quality. It might not be an effective method to pool all students using automated feedback into one group while evaluating the impact of automated feedback in detail, as the individual capabilities and differences also matter. These individual differences could be studied by finding ways to assess the feedback literacy of students while interpreting and making sense of automated feedback to apply it on their writing (Carless & Boud, 2018). This could lead to further study of the automated feedback literacy and application among students, and possible ways to personalize feedback for students to use them better. The revision graphs are currently designed for researchers to study revision patterns across students, and their interactions with automated feedback at a preliminary level. Future work can also explore the option of improving the readability of the graphs and evaluating it with instructors for their use.

To study how additional scaffolding can help students engage with the automated feedback, the effects of peer discussion and peer feedback were evaluated in this iteration. From the survey responses, students generally found the peer discussion useful to learn other perspectives and encourage reflection, but also reported that its effectiveness depended on the peer's expertise and ability to provide feedback. To take

a deeper look at the actual content of students' peer discussion to study the value it adds to automated feedback, a content analysis was done on sample discussion transcripts. Initial findings from the discussion analysis show that peer discussion facilitates students' engagement with automated feedback to some extent. In particular, it helped students identify ways to improve their writing, and suggest improvements to their peers. It has also helped them to share their interactions with automated feedback and get clarifications if needed from their peers. Some students have further used peer discussion to critique the automated feedback, leading to deeper engagement. However, more studies need to be done to validate these findings with a bigger sample. Since the time allotted for this sub-task was limited, student engagement can be studied in more detail if more time is allotted in future studies combining automated feedback and peer discussion. Furthermore, the data analysis explained in the above section was affected by a few issues due to the quality of the collected data. Students were sometimes not audible and the video recorder could not pick up all the conversations clearly, which led to difficulties in the transcription process. Students used gestures, and pointed to texts on the screen which were not legible at times and were hidden from the shoulder level. They also did not use complete sentences while talking to their peers in the face to face setting since their gestures helped them convey meaning. The transcription of this peer discussion was hence completed by the researcher who was in the data collection environment so as to incorporate some context in this analysis.

Results from this study show students' engagement with a given piece of writing using automated feedback; this effect should be tested in other scenarios where students use their own writing as well. One problem however is that the quality of peer feedback depended on the peer's ability and this affected the discussion in some groups. To overcome this problem of feedback quality, previous studies have suggested the use of guiding prompts and training for assessors to provide effective feedback (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).In addition, there were differences in how the different groups interacted with the automated feedback while discussing the writing task from low levels to high levels of reflection. Group 1 expressed a deeper understanding of the automated feedback and engaged with it at a higher level by going beyond surface-level changes and challenging automated feedback, when compared to groups 2 and 3. This aligns with prior research on using peer feedback on writing revision tasks which found that novice revisers were not able to do more than superficial grooming when asked to revise a peer's draft, and while their edits were helpful for their peers, they did not necessarily help them to improve the structure or content of their drafts (Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010). Some students are more capable of engaging with the automated feedback meaningfully than

others when discussing with their peers. To make it effective for all students to engage with automated feedback in a meaningful way, more scaffolding might be required to teach students how to reflect on their writing and use automated feedback efficiently.

Findings from the iteration support the usefulness of AcaWriter feedback to improve writing for the Civil Law subject, with possible avenues for supporting students better to engage with automated feedback. The significant differences in perceived usefulness score observed between the control groups led to the decision of providing automated feedback to all students from the next iteration. The design and theoretical principles can now be transferred to a new subject context to study its generalisability by transfer, explored in detail in the next Chapter.

Chapter 5: Conceptual Model and Design Transfer to Learning Context 2

This chapter introduces the conceptual model that emerged from the previous iterations and explains its application in a new learning context to which the design was transferred to in the next iteration. Building on the frameworks reviewed in Section 2.6, and further contributing to learning analytics theory by bridging research and practice, a conceptual model called *Contextualizable Learning Analytics Design (CLAD)* was developed from the iterations discussed in Chapter 4: Design Iterations in Learning Context 1. The model combines elements of learning analytics (LA) and learning design (LD) by involving educators and researchers in the core design of learning analytics for learners. The key components include LD elements: Assessment and Task Design, and LA elements: Features and Feedback/UI that flexibly shape each other to bring context for learning analytics. The CLAD model thus builds the theory that LD and LA elements should synchronously align in particular contexts to support that learning, rather than being rigidly fixed. It will be explained in this chapter, followed by its exemplification in a second learning context demonstrating the transfer of design principles from the previous Law learning context[15].

5.1 Conceptual model: Contextualizable Learning Analytics Design (CLAD)

The lack of explicit pedagogical theory is a challenge in the field of Learning Analytics, which can limit the potential of LA approaches and cause misaligned practice (Koh, Shibani, et al., 2016). To contribute to the theory of contextualized LA and to further support theory-guided practice, a conceptual model is proposed, that brings together elements of LA and LD for contextualized support for learners. This contribution of my thesis builds on existing work to develop a model grounded in theory for the wider learning analytics community, with validated empirical evidence for developing contextualizable learning analytics applications.

[15] Sections of this chapter have been published in conference proceedings of LAK2019 (Shibani, Knight, et al., 2019)

Knight, Buckingham Shum, and Littleton (Knight et al., 2014) argue that the way we deploy learning analytics as forms of assessment implies particular epistemological and pedagogic assumptions. In this view, how we understand learning or 'success' is fundamentally bound up with the learning context and intended learning outcomes. Thus, learning analytics under a 'pragmatic maxim' bases the quality of analytics on its practical consequence on learning, asserting the need for contextualization of LA in a learning context (Gibson & Lang, 2018). Taking this perspective, a conceptual model for learning analytics is defined to align with pedagogical contexts. The conceptual model builds on existing work connecting LA and LD to provide a usable model for learning analytics researchers and practitioners for implementing contextualizable learning analytics applications in their pedagogic contexts.

Contextualizable Learning Analytics is defined as *"the pairing of learning analytics and learning design that can be flexibly adapted for different learning contexts"*. The core of the concept stems from the link between two types of components: Learning Analytics (LA) and Learning Design (LD), which mutually drive each other during the course of contextualization Figure 62. LA here includes the tools and technologies to be contextualized, and LD denotes the teaching and learning context whose pedagogical intention is the context in hand. These key components come from existing work on aligning LA and LD so that the pedagogic intent of analytics applications can be maintained, interpretation of analytics enhanced, and the assumptions of the learning design tested using data and tools provided from learning analytics (Gašević et al., 2016; Lockyer et al., 2013; Shibani et al., 2017).

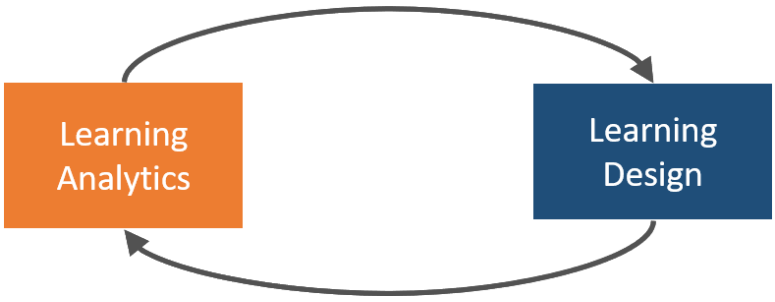


Figure 62: Link between Learning Analytics and Learning Design in Contextualization

To unpack how these components can work together in practice, we introduce a conceptual model for Contextualizable Learning Analytics Design (CLAD) (see Figure

63). The model includes elements for contextualizing learning analytics to different learning settings. Educators who are the subject experts in the context drive the contextualization process in the model. As identified by previous work, educators play a key role in contextualizing LD and teacher-inquiry processes are critical to authentically embed LA in practice (Alhadad & Thompson, 2017; Bakharia et al., 2016). But in addition to the educators, the model also includes other stakeholders like LA designers who work hand in hand with the educators for effective contextualization. This communication among stakeholders bridges the knowledge gaps about the potential and trade-offs of LA innovations and manages expectations among the different stakeholders (Prieto, Rodríguez-Triana, Martínez-Maldonado, Dimitriadis, & Gasevic, Forthcoming; Thompson et al., 2018). The productive dialogue between stakeholders in interdisciplinary teams facilitates an improved understanding of the potentials and pitfalls of analytics, which can lead to pedagogic impact and increased adoption (Thompson et al., 2018).

The model proposes four main elements represented as gear cogs in Figure 63 for contextualizing learning analytics design (evolving from the implementation discussed in Section 4.4.2), so that LA can be coherently integrated in authentic classroom settings. The learning design (dark blue) brings the educators to contextualize learning analytics through the key elements of Assessment and Task Design. The key elements of learning analytics (orange) are Features and Feedback and User Interface (UI), developed by LA designers which are adapted for the LD. Features refers to those attributes in the data that are considered significant enough to identify, and bring to the user's attention via the Feedback/UI. These LD and LA elements should synchronously align in particular contexts to support that learning, rather than being rigidly fixed. When used effectively, LA plays a supporting role in LD. When used ineffectively, the two mechanisms are either not fully engaged, or are driving in opposite directions, with LA resisting the LD, or vice-versa.

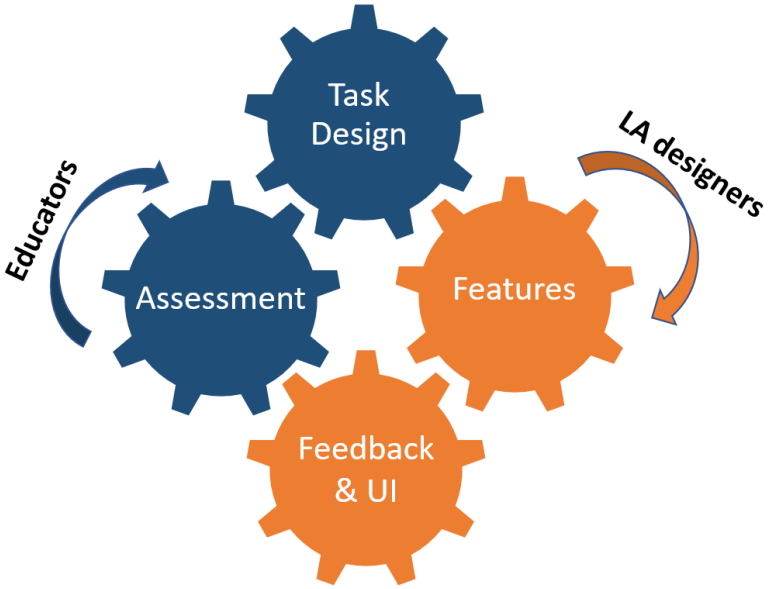


Figure 63: Conceptual model for Contextualizable Learning Analytics Design (CLAD)

Assessment plays a key role in contextualization, since it should align with the intended learning outcomes of the curriculum (Biggs, 1996), without which the learning analytics would remain distant from the pedagogy. This is important because the uptake of LA by learners and educators depends on how well its usage is mapped to their curriculum and adds value. The task design ensures that the LA is relevant to the learner and is grounded by pedagogic theory (Shibani et al., 2017; Wise, 2014). The construction of complex learning activities that make use of analytics to encourage further metacognition requires pedagogical expertise. Such effective implementation of LA can be supported by LD (Kitto et al., 2017). For LA to align to LD, features of the LA system should be adaptable to the learning context to provide useful analytics. That is, the features should be pedagogically salient, and not simply technical representations of low-level trace data. This modeling might require changes to the underlying technical infrastructure for introducing new features identified as a requirement for the context, or the possibility to amend existing features in order to address the LD context. The feedback and the User Interface (UI) that the learner ultimately interacts with also should be contextualized for the pedagogical context, so that the relevance and usage of the technology for the context is well understood.

In addition to being tuned for the LD, LA also shapes LD. The elements of LA contribute to the LD in two ways. First, LA and the data it is based in create technical affordances for the LD. Second, LA can challenge the theoretical underpinnings of the

originally designed learning activities. LA and its intervention design can help learning designers with improving the alignment of: learning tasks, assessment criteria, and LA data for reflective purposes; bringing these into 'constructive alignment' towards the learning outcomes (Biggs, 1996). Figure 64 illustrates the interplay between LA and LD elements to show how each element should be able to flexibly adapt to a context to align with the other elements.

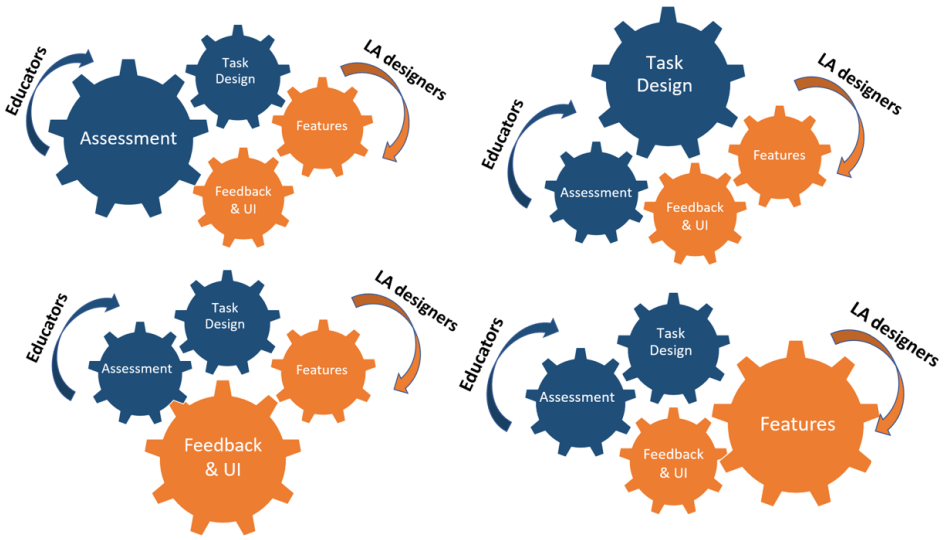


Figure 64: Flexible interplay between LD and LA elements in the CLAD model for contextualization

The CLAD model thus accentuates the concept of *malleable* LA and LD, where the contextualization elements shape and drive each other to align in the pedagogical context. This process of developing contextualizable Learning analytics by aligning the elements of LA and LD can help spot issues and requirements to make them flexible enough to improve each other when desirable. Such alignment draws on the notion of flexibility in scripts for computer-mediated interactions so that the strict design choice does not affect the dynamics or over-constrain the pedagogic setting (Dillenbourg & Tchounikine, 2007; Knight, Shibani, et al., 2018). The model emphasizes an equal contribution of LA and LD for contextualization, so that they do not crowd out each other with more weight given to one over the other. This task-oriented approach is granularized enough to study how students are using LA to study its effectiveness, in contrast to other work which focuses on LD-LA alignment at an overall course level (Gašević et al., 2016).

The contextualizable LA design model may be applied both top down and bottom up. First, top down, large scale generalized LA can apply CLAD to provide support for particular learning contexts. In this way, mature LA tools that are proven to work in general settings can implement CLAD to add specificity to particular LDs, to provide contextualized support. Second, bottom up existing LA for specific learning contexts can be scaled up to other learning contexts. This transfer of design across learning contexts can be enabled by the use of abstractions like 'design patterns' that can share the general principles of task design and purpose including theoretical foundations and the practical implementation of designs for educators (Goodyear, 2005).

5.2 Design Iteration 4: Transfer

The fourth design iteration implemented a transfer of intervention design from the Law context to a new discipline – an Accounting discipline where students wrote business reports. This second learning context provides an example of transferring a design that worked well in one context to another, using the Contextualizable Learning Analytics Design (CLAD) elements discussed earlier.

5.2.1 Task Design and rationale

The new pedagogic context here is an Accounting subject, where students are required to write business reports as part of an assignment. While accountants mainly deal with numbers, they also have to create and communicate information to aid decision-making, and their main form of communication is a business report. Some key rhetorical moves were identified as important constituents in good business reports by instructors, which could be taught with the help of AcaWriter. The pedagogical activity design from the Law essay context discussed earlier was held as a baseline with few design elements tuned for contextualizing it to this business report writing context. The changes in task design are shown in Design Representation 4A.

DESIGN 4: Benchmarking, Text-Revision, Peer-Discussion, and Scaffolded Automated Writing Analytics with Contextualized Feedback, Transferred to a New Discipline Setting

Problem: We wanted to transfer the existing design to another discipline, but with contextualized support for the new discipline's writing context. Similar to the previous designs, we wanted students to engage with each other and with automated feedback around the application of assessment criteria, and to develop their evaluative judgement. We also wanted them to apply this knowledge on their own writing.

Task: The base tasks in design iteration 3 were adapted for the new subject context. The online tasks were provided as a homework activity, and the peer discussion was facilitated in class.

Tools/materials and participant structures: Students completed the online task individually, and examined the automated feedback further during peer discussion in class. New materials similar to the previous law context including an introductory video explaining rhetorical moves in the context of business report writing, exemplars, degraded text, a sample improved text were created.

Iterations and Augmentation: This task design developed from that described in design 3, introduced a new feedback module in AcaWriter to provide context-sensitive feedback for business reports. An additional scaffolding provided to students to apply AcaWriter feedback on their own writing may develop students' abilities further to critically engage with such automated feedback.

Design Representation 4A: Design pattern of the task design changes in iteration 4

5.2.2 Exemplifying CLAD model using the Accounting context

The CLAD elements implemented in this context are explained in detail next using the new Accounting context. Here AcaWriter rhetorical moves were first mapped to rubric elements of the business report's assessment criteria. From this mapping, several AcaWriter features were identified as important for the context which shaped the tool feedback and made the assessment criteria more robust. The assessment criteria for the report's content mapped to AcaWriter rhetorical moves are shown with examples in Figure 65.

Organisational analysis
Where does your report provide contextual information about the organisation's objectives, strategy, structure and activities?
Defining performance
Where does your report provide your perspective P about how to define performance or success for the organisation? Where does your report provide emphasis E to highlight the most important aspects of performance for the organisation?
Justification of your definition of performance
Where does your report provide convincing, persuasive justifications for your definition of performance by proposing novel N or critical insights, contrasting ideas or tension C? Where does your report justify your definition of performance with reference to prior work or background B?
Written communication
Where in your report do you use appropriate summary statements S to signal the content, sequence and goals of the report?

Acawriter Move	Sample Sentences
Summary	S This report defines performance for Nike Inc as a whole from the three main perspectives of economic, social and environmental performance. S This section will explain how sustainability can result in the achievement of organisational objectives and contribute to company success.
Perspective, Emphasis	P Importantly, research has suggested a link between stronger brand perception and customer loyalty. E Therefore, to effectively measure performance for the Canadian Head Office of Lululemon it is essential to consider how the transformational self-improvement ethos of the company is achieved by analysing non-traditional metrics.
Background	B Previous market analysis shows that Nike is the world's largest supplier of athletic shoes, equipment, and apparel. B Such an evaluation by a global organization has been observed previously.
Contrasting ideas and Issues	C These requirements maintain product quality as unethically produced garments could be of lower quality, damage its reputation and ultimately contradict the companys objective. C While the companys corporate mission is to maintain its market position as a leading brand for an active and mindful lifestyle, a holistic approach to defining performance is necessary given that Lululemons strength lies in its premium image
Novelty	N This closer connection with their customers allows Nike to create new ideas and convert them into products quicker while also being in touch with their demands. N Nike believes that through investing in employees and communities, they can inspire while also creating a new method of growth (Nike 2018c).

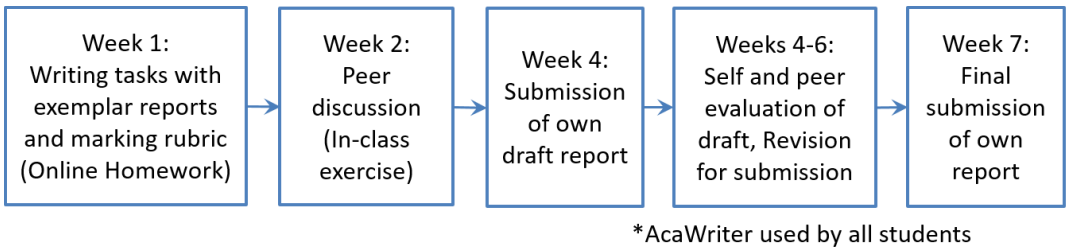
Figure 65: Mapping of assessment criteria in Accounting business reports to rhetorical moves in AcaWriter

Only the relevant rhetorical moves were highlighted in the analytical report in this context, and similar to the Law context, the feedback messages were tuned in AcaWriter based on the missing key moves for this context as shown in Figure 66.

Figure 66: Feedback messages from AcaWriter contextualized for Accounting business reports

To design the pedagogically grounded learning activity, design patterns from the existing Law design were transferred to this context. The learning design spread over several weeks co-designed with the instructors is shown as a workflow of tasks in Design Representation 4B. The core of the writing activity involving introduction to rhetorical moves and AcaWriter, self-assessment and revision tasks remained the same, with few modifications done for the context. The online writing activity was designed for Week 1 as a homework activity, so students can learn from past examples and practice revision on the given sample writing in their own time. In week 2, they continued this writing activity in class with peer discussion, where students could interpret, clarify, and learn from each other. In the following weeks, students used AcaWriter for their own draft reports which they submit for assessment. The design hence involved both practice with given samples on AcaWriter to learn writing skills, and its application for their own written reports. While working with their own reports for their assignment, students were provided an additional scaffolding to be more engaged with AcaWriter feedback as part of their self-evaluation exercise. They were provided prompts to answer questions on whether they agree or disagree with the

automated feedback and why. They were also asked to annotate the rhetorical moves report provided by AcaWriter to encourage their reflection, and critique the feedback provided. All students completed the same writing activity with access to AcaWriter - no variable study conditions were defined; this change was incorporated due to results noted in the previous iteration that students with AcaWriter feedback find the intervention more useful (significant) than those who did not get the feedback.



Design Representation 4B: Workflow of tasks designed for Accounting context in iteration 4

5.2.3 Data Analysis

To evaluate how the transfer of pedagogic intervention with AcaWriter worked in a different learning context, student responses for the survey questions were analysed from the online task to study their perceptions, answering RQ1a. Their engagement with automated feedback on their own writing using the additional scaffolding provided using prompts was also studied to answer RQ2b.

5.2.4 RQ1a) Student perceptions of the writing intervention

Data from 302 students who completed the activity showed that students generally perceived the task to be useful to improve their report writing ($M = 3.8$, $SD = 0.9$). Figure 67 shows the distribution of rating across students.

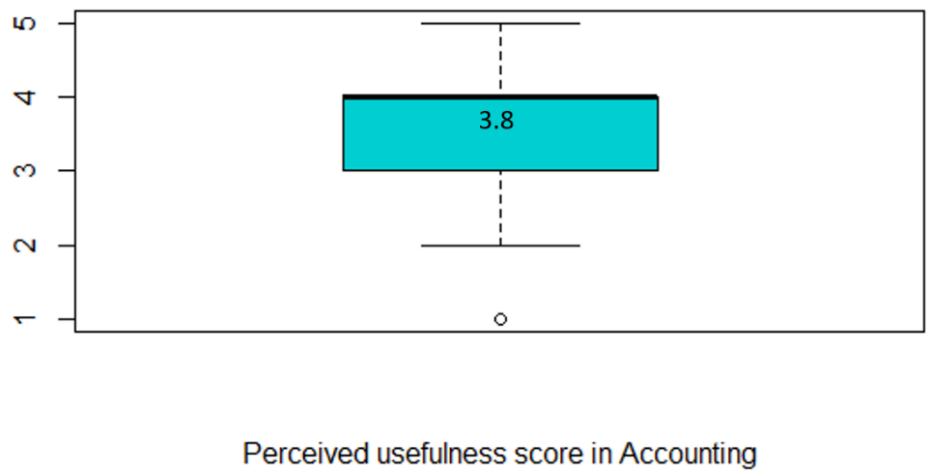


Figure 67: Perceived usefulness of the writing intervention in iteration 4 (Accounting context)

66% of students selected 4 or 5 (highest) in relation to the perceived usefulness score, which shows that many students found the intervention useful to improve their report writing. Further qualitative responses showed that many students found the activity helpful and appreciated the feedback provided, while few others believed that human feedback would be more effective compared to the automated feedback.

"By doing this exercise, it shows me a clear idea on how we could approach our writing for the upcoming individual report assessment task. Not just that, but will generally help in gaining confidence in writing a clear and concise business report future uni projects or for work." – Student 286

"It's like having a tutor or another person check and give constructive feedback on your work." – Student 54

"I have always been quite good at writing I think, however, I never approached it in this way and I have found it quite helpful." – Student 55

"It was interesting to see what AcaWriter was able to pick up and the feedback it gave allowed me to edit the sample doc a lot faster than normal." – Student 57

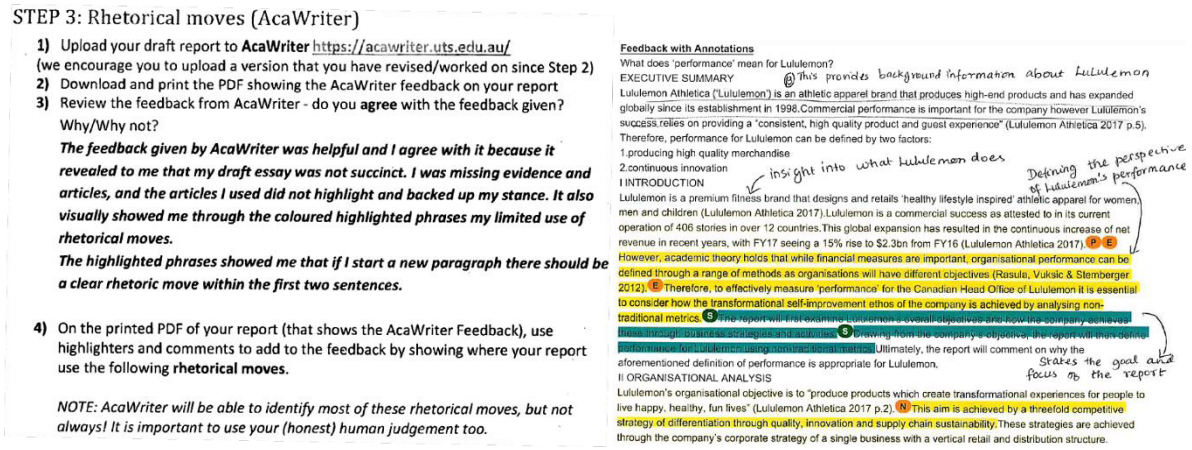
"I believe this exercise may be of better use to some than others, and that it offers good information that could be of use, for me personally, the program would need to be able to help me to better understand what I'm doing incorrectly than correctly, and as such, I believe that a human reading through it is still more effective in that regard." – Student 131

Overall, the intervention contextualized for the subject was found to be helpful for students to improve their writing skills, with the transfer of design patterns successfully implemented and tested. However, the impact of this activity on the writing of their own had to be studied to see if students apply the skills they learnt from the activities onto their own writing to reap long term benefits.

5.2.5 RQ2b) Student engagement with automated feedback scaffolded by additional instruction

To study how students engaged with automated feedback using the additional scaffolding to answer RQ2b, the self-evaluation sheets of students where students answered prompts about their engagement with AcaWriter feedback were analysed. This task was part of a mandatory self-evaluation exercise (SEE) created for assessing their assignment by the instructor, and involved students in applying automated feedback for their own assignment writing. The self-evaluation exercise included several components including general formatting of the report, referencing and plagiarism check, automated feedback on rhetorical moves, and a peer evaluation, of which we are interested most in how students engage with the automated feedback from AcaWriter.

The AcaWriter feedback's SEE section consisted of a few questions and asked students to annotate the AcaWriter's report to better engage students with the automated feedback provided on their writing. A sample SEE sheet and annotated AcaWriter feedback is shown in Figure 68 (Rewritten to ensure handwriting is not identifiable).



STEP 3: Rhetorical moves (AcaWriter)

1) Upload your draft report to **AcaWriter** https://acawriter.uts.edu.au/ (we encourage you to upload a version that you have revised/worked on since Step 2)
2) Download and print the PDF showing the AcaWriter feedback on your report
3) Review the feedback from AcaWriter - do you **agree** with the feedback given? Why/Why not?

The feedback given by AcaWriter was helpful and I agree with it because it revealed to me that my draft essay was not succinct. I was missing evidence and articles, and the articles I used did not highlight and backed up my stance. It also visually showed me through the coloured highlighted phrases my limited use of rhetorical moves.

The highlighted phrases showed me that if I start a new paragraph there should be a clear rhetoric move within the first two sentences.

4) On the printed PDF of your report (that shows the AcaWriter Feedback), use highlighters and comments to add to the feedback by showing where your report use the following **rhetorical moves**.

NOTE: AcaWriter will be able to identify most of these rhetorical moves, but not always! It is important to use your (honest) human judgement too.

Feedback with Annotations

What does 'performance' mean for Lululemon?

EXECUTIVE SUMMARY

This provides background information about Lululemon

Lululemon Athletica ("Lululemon") is an athletic apparel brand that produces high-end products and has expanded globally since its establishment in 1998.Commercial performance is important for the company however Lululemon's success relies on providing a "consistent, high quality product and guest experience" (Lululemon Athletica 2017 p.5). Therefore, performance for Lululemon can be defined by two factors:

1.producing high quality merchandise

2.continuous innovation

insight into what Lululemon does

Defining the perspective of Lululemon's performance

I NTRODUCTION

Lululemon is a premium fitness brand that designs and retails 'healthy lifestyle inspired' athletic apparel for women, men and children (Lululemon Athletica 2017).Lululemon is a commercial success as attested to in its current operation of 406 stories in over 12 countries.This global expansion has resulted in the continuous increase of net revenue in recent years, with FY17 seeing a 15% rise to $2.3bn from FY16 (Lululemon Athletica 2017). P E

However, academic theory holds that while financial measures are important, organisational performance can be defined through a range of methods as organisations will have different objectives (Rasule, Vuksic & Stemberger 2012). E Therefore, to effectively measure 'performance' for the Canadian Head Office of Lululemon it is essential to consider how the transformational self-improvement ethos of the company is achieved by analysing non-traditional metrics. S The report will first examine Lululemon's overall objectives and how the company achieves these through business strategies and activities. S Drawing from the company's objectives, the report will then define performance for Lululemon using non-traditional metrics. Ultimately, the report will comment on why the aforementioned definition of performance is appropriate for Lululemon.

States the goal and focus of the report

II ORGANISATIONAL ANALYSIS

Lululemon's organisational objective is to "produce products which create transformational experiences for people to live happy, healthy, fun lives" (Lululemon Athletica 2017 p.2). N This aim is achieved by a threefold competitive strategy of differentiation through quality, innovation and supply chain sustainability. These strategies are achieved through the company's corporate strategy of a single business with a vertical retail and distribution structure.

Figure 68: Sample SEE sheet and annotation of AcaWriter report completed by students in Acounting

Students were asked explicitly to apply their human judgement too while using the AcaWriter tool for self-evaluation. The prompts provided to students to engage them with automated feedback on rhetorical moves for improving their business report is provided in Appendix B. Students submitted these completed SEE sheets in hard copy format along with their assignments for marking. Data from the scanned SEE sheets of students submitted as part of their assignment were transcribed for a subset of ***114*** students (based on their availability from the instructors) in an excel file. The file was transcribed for themes emerging from the types of student responses to scaffolding and engagement with feedback.

Analysis of their engagement with the automated feedback helps us study the *'feedback literacy'* of students while working with automated feedback. Feedback Literacy as a named concept has been recognized recently as a key academic skill for students (Carless & Boud, 2018; Sutton, 2012). Extending Sutton's work which defines feedback literacy as the ability to "read, interpret and use written feedback", Carless and Boud (2018) redefine feedback literacy as *"the understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies.* This section will explore how different students engage with AcaWriter feedback and demonstrate feedback literacy by making use of their response to prompts. By using prompts, we force students to take a closer look at the feedback and record this process and their thinking using those scaffolds, which otherwise would not have been made visible when engaging with automated feedback.

From students' responses to question 3, their agreement with automated feedback from AcaWriter was coded as one of the following categories:

1. Agree: If they completely agree with the feedback, without any critique of the feedback provided

2. Disagree: If they don't agree with the feedback provided, expressing some form of rebuttal to the feedback provided

3. Neutral: If they agree with some parts of the feedback and not others, or agree to some extent with rebuttal to few instances of feedback

4. NA: Did not answer the question

Examples from the coding of main categories from above are shown in

Table 7.

Table 7: Coded Examples of students' agreement positions with AcaWriter feedback

Example	Student Id
Code 'Agree'	
Agree, because I missing a background move in to next text, and missing a contrast sentence, which I did not realise before.	114
yes, AcaWriter appeared to recognise that I included all of the required rhetorical moves in my report	8
Yes, the feedback seems to be correct and has correctly highlighted rhetorical moves.	54
Code 'Disagree'	
No. I can see multiple places where I have made a rhetorical move and it wasn't highlighted	6
I do not agree with the feedback given because some sentences about background, emphasis and perspective are not highlighted by the software.	50
In general I don't agree with the feedback. I believe I have employed enough rhetorical moves. The issue is that they have not been worded in the way AcaWriter prefers.	80
Code 'Neutral'	
I do agree with the feedback given to the information that was picked by AcaWriter. However, I do not agree with the information it did not pick up. I believe I did use more rhetorical moves that AcaWriter did not pick up, whether it was because there were not the appropriate/signal words or phrases used.	12
I mostly agree with the feedback. However some of my rhetorical moves were not detected. Despite this, the feedback were still useful as it correctly identified some areas of my work that have applied moves.	100
Yes - the feedback regarding the rhetorical moves used in the first paragraph was helpful, however AcaWriter did not register all examples of moves I think.	111

The next prompt asked students to spend some time working on their reports by running AcaWriter analysis, and enquired about the effect that AcaWriter feedback had on their writing. The following prompt also asked students to report on the changes they have made/ will make to their report after using AcaWriter. Student responses to these prompts were analysed to study the extent to which students understood AcaWriter feedback and applied it to improve their writing. They were categorised as one of the two categories described below:

1. Deep: If they demonstrated an understanding of how AcaWriter picks rhetorical moves to provide feedback and/ or applied the feedback to reflect on their writing and make significant revisions.
2. Shallow: If they failed to exhibit a deep understanding of AcaWriter feedback by critically engaging with the feedback to make considerable revisions.

Examples of students' responses coded as **'Shallow'** are shown below:

- *"I updated summaries but the software did not pick up some of the moves made. 1) Provide further critical insight 2) Make further word improvements in ideas. 3) Show more personal perspective." -69*
- *"AcaWriter shows me I am missing the background information and previous work, so I accepted this suggestion and add to my final report. 1) I fixed some grammar and vocabulary errors. 2) I added some sentences to make report more reference. 3) Trying to add more evidence to support my report." – 41*
- *"I learnt to add more emphasis statements and emphasis statements. 1) Make emphasis paragraphs. 2) Summary statements." -58*
- *"There is nothing much that changes as AcaWriter managed to detect most of the aspect. 1) More summary statements presented as it lacks initially. 2) I will provide more critical ideas." - 73*
- *"I fixed some reference and structure of my report. 1) confused sentence. 2) reference." -87*
- *"No changes. I have 'perspective'. I've provided 'emphasis'. I am sure that I have core idea and background explanation. 1) I have the emphasis of writing and detailing discussion. 2) The system failed to recognise my writing. 3) There is no point to make any changes follow the 'feedback'." -102*

Example of students' responses coded as **'Deep'** are shown below:

- *"The changes that were made created more areas of background information as well as greater emphasis of ideas. Once completed, the feedback determined that all of the key rhetorical moves had been addressed. This allowed the report to be more analytical to provide a greater justification of the definitions of performance. 1) Many changes were made by re-wording some sections to create more emphasis. This occurred particularly within the strategy section providing greater emphasis of the importance of the strategies that are employed by Rio Tinto in allowing the Company to achieve its overall objectives. 2) Stronger background information was provided within the areas that defined performance. This was to provide a greater justification of the definitions of performance, particularly within the innovation and environmental sustainability section, which now provide a greater justification of the definition, rather than just a description. 3) ln reformatting the sentences to meet the requirements of the rhetorical moves, the overall persuasiveness of the definitions of performance was improved. This was done to provide a stronger link between the company's recognition of performance and the achievement of its mission, to justify why the said definition of performance is important. " – 60*

- *"I used the linguistic cues recognized by the AcaWriter more frequently, then more rhetorical moves are highlighted by the system. For example, I used "although", the tool spotted the whole sentence as a "contrast". I also added "it is widely recognised that", then the tool spotted it as a "Background" move. 1) I added emphasized words such as "the key focus areas of performance" to make my arguments stand out. 2) I added coherent words such as "therefore", and "in summary" to specifically indicate that the following sentences are summary statements. 3) At the beginning of subsection 2.2 and subsection 3, I added sentences to show my perspectives. 4) I further divided the block of text into small paragraphs to make my essay more clear and easy to read. 5) I found that after using AcaWriter, it not only improve my written communication, but also enhance the progression of my idea and analysis for Westpac performance. For example, I realised that I need provide more perspectives in the justification part so that my report shows a deep engagement with academic literatures." -52*

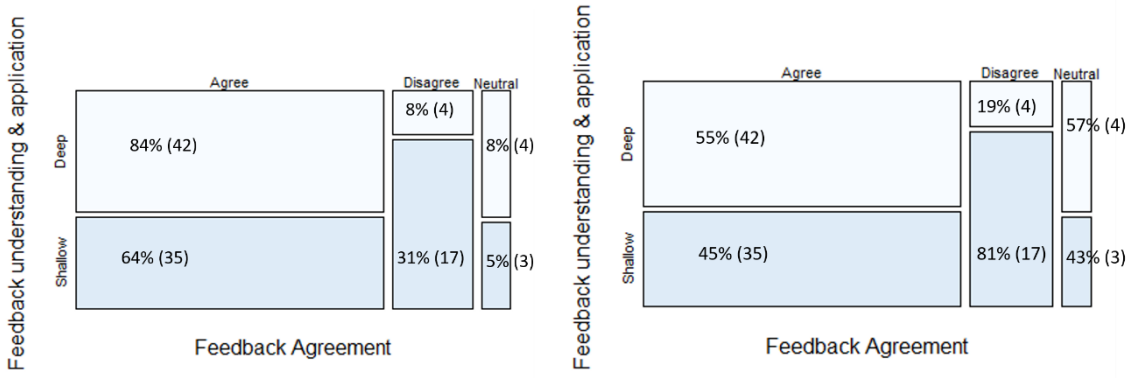
- *"By adding clearer justification elements and refering to additional research, improvements were made to N and B. 1) Including more background information, particularly when defining performance. 2) making reference to pressures work - look*

into more sources. 3) Develop stronger justifications and critical insights on performance." -93

- *"After my second draft, my essay was visually more colourful. AcaWriter spotted the key moves: Summary, Background, Perspective on Contrast. This signals to me that my essay should be clearer. After re-reading it again my language was also clearer and it developed my definition of performance clearer. I was able to fix my definitions of performance. From there I had a better idea of what t wanted to say without the irrelevant and messy thoughts. 1)My report was missing research and academic articles to bock up my points. This is critical in answering the question of why this performance aspect in important to my company Australia Ethical Investments. Therefore, I plan to use more articles and references to reline my research and improve my "Background move". 2) I also realised after getting feedback from Acawriter that t did not hove a Contrast move. This is important in showing the issues and different perspectives on a topic. Therefore, I plan to consider other conflicting academic articles on the issue of ethical investing to further my argument and show my deep engagement into this topic. 3) Although I have defined my key performances, my perspective and stance was not clear" My definition of performance was not well defined and there were little explanation to emphasis on the important performance aspects. Therefore, I will adjust my essay so thot my definition of performance is much more clearer and insightful." -109*

- *"1) After my first submission to AcaWriter, I changed the general structure of my report significantly. This was due to the database bringing to my attention that I was not clearly defining the organisation with nothing being highlighted in the organisational analysis. 2) The tool allowed me to identify areas where I was waffling and using an excess amount of words. Through the tool highlighting effective sections of the report, it allowed me to review my submission to ensure the remainder of my report was written in a succinct and precise manner. 3) After using the tool, I will still need to focus on strengthening my justification arguments, so they become clear and easy to pick up on when being read over. At the stage of my final AcaWriter review, some definitions of performance have clear justifications however others are weak and need improvement." - 39*

From the student responses coded based on agreement and understanding/ application of AcaWriter feedback as above, quantitative analysis and visual representations were done to get more insights. The cleaned dataset with inconsistencies and missing data (unanswered NAs) removed contained 105 records. Preliminary results from the analysis of this data are discussed next. Figure 69 shows the proportion of students engaged in deep and shallow ways categorised by their agreement with AcaWriter feedback (numbers indicate row-wise and column-wise proportions % (n) in each category).



(A) (B)

Figure 69: Row-wise (A) and Column-wise (B) proportions of students with shallow and deep engagement with AcaWriter feedback categorised by agreement categories

Among the group of students who demonstrated *deep* engagement with AcaWriter feedback by understanding and applying it on their writing, the majority (84%) *agreed* with the feedback. In the shallow engagement group, more students tended to disagree with the feedback (31%), compared to the disagreement percentage in the deep engagement group (8%). Also, in the group of students who disagreed with the feedback (shown in B of Figure 69), most students (81%) demonstrated shallow engagement. This shows that many students who blatantly disagreed with AcaWriter feedback did not engage with the feedback in a meaningful way that the prompts intended to support. However, no association was found between feedback engagement and agreement in a chi-squared test: $\chi 2$ (2) = 8.6, p = .01. A more detailed analysis of these categories with respect to the scores obtained by students for the assignment is provided in the next section.

The coding of these categories also uncovered certain problematic aspects in interpreting student responses. When student responses to the prompts were coded, it

was seen that students demonstrated varied levels of understanding and incorporation of AcaWriter feedback into their revisions. While few instances were recognisable in terms of student understanding and use of feedback to engage with their writing as deep and shallow, few others were hard to code, particularly if their wordings were not explicitly indicative of their understanding. Examples of instances below were hard to code conclusively as 'deep' without elaboration, but demonstrated possible deep engagement, and hence were provided the benefit of doubt and coded as 'deep':

- *"1) Added background reference to Taxation Taskforce - (B) achieved 2) Increased justification to 'performance' definition - more (N) and (c) appeared. 1) Greater reference to background/ previous work. 2) Improved justification to Otley's (1999) "performance" definition. 3) Increased reference to academic sources." -30*
- *"AcaWriter recognised where I made changes and I was able to integrate all rhetoric moves into my report. 1)found more academic relevant articles 2)rearranged sentences as to improve structure 3)Rephrase the last section since no moves were highlighted despite editing it" - 3*

This is a challenge in coding such text without the context of knowing how the student engaged with the feedback. It is hard to identify how well the student engaged with the feedback if their responses do not elaborate their interactions in detail. Furthermore, there were differences in how students applied and understood the feedback even within the deep and shallow categories as we can see from the examples listed above – while few demonstrated better understanding of how to effectively apply the feedback and disregard its suggestions if appropriate, few demonstrated better changes made to their writing considering other factors not directly related to the automated feedback provided, but still deep in terms of the revisions made. A finer coding scheme will be developed in the future to capture this range within deep and shallow understanding. Furthermore, this coding scheme will be validated with examples coded by more than one coder with interrater reliability to generalize findings. It will explore feedback literacy at more detail, allowing conclusions to be made from the data using individual differences among students. Future work can include additional think-aloud strategies to capture the cognitive processes behind student engagement and application of automated feedback. By understanding students' feedback literacy and the types of feedback that best help students using their

engagement, future work can design personalized writing support for students that adapt to their needs.

5.2.6 RQ1b) Impact of automated feedback on student performance

To study the impact of student engagement with automated feedback in terms of more tangible results, marks that measure student performance in their writing assignment were obtained from the instructors. Since this design iteration of the project was the only one where students used AcaWriter feedback with their own assignments (in addition to working with exemplars), their marks were used to assess their final performance. Based on the assessment criteria for business reports, the assignments were scored for written communication by the instructors and tutors. The marks scored by students in this assignment contributed to their overall credits to pass the subject.

These marks were available for all students of the cohort, and the dataset contained 403 records in total. The assignments were scored from 0 to 30 for written communication and this mark was used for analysis. Summary statistics of the marks are shown in Table 8.

Table 8: Summary statistics of marks scored by students in the submitted assignment

N	Mean	SD	Min	Max	Median
403	20.67	4.87	5.5	29.5	21.5

This includes the marks of students irrespective of whether they completed the first online activity with AcaWriter or not. So firstly, the effect of students' completion of the online activity on their marks was studied. The marks scored by students who completed the online activity vs the ones who did not complete the activity is shown in Figure 70.

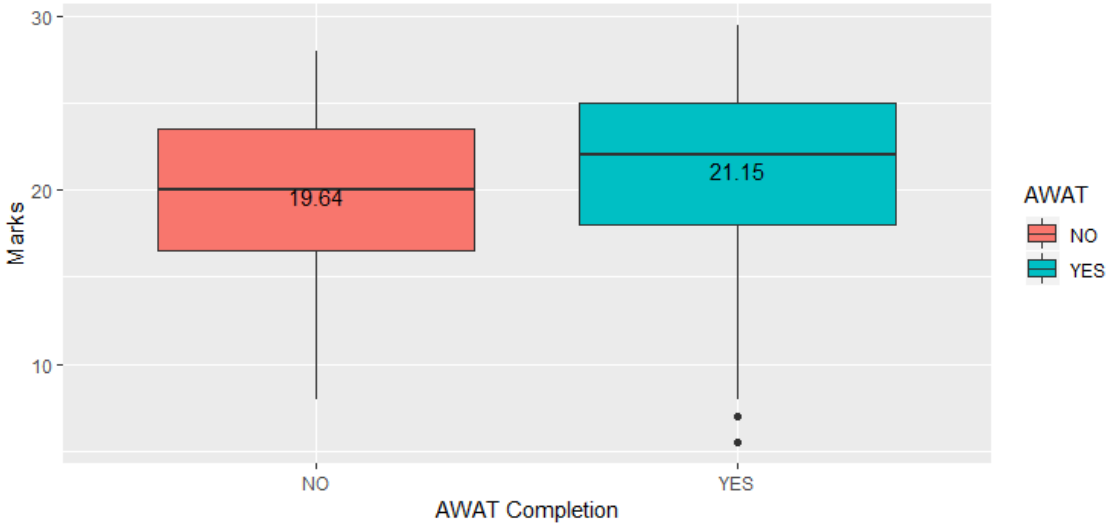


Figure 70: Student marks across groups who did (YES)/ did not (NO) complete the online activity

Results show that students who completed the online activity with AcaWriter scored higher marks for their assignments on average (M = 21.15, SD = 4.87, N = 274) than students who did not complete the activity (M = 19.64, SD = 4.72, N = 129). Welch two sample t-test showed a significant difference between groups: $t(258) = -2.98$, $p < .001$, with a small observed effect (Cohen's d = -.31, 95% CI [-0.52, -0.10]).

Next, to study the impact of student engagement and agreement with AcaWriter feedback on their marks, the clean data set was considered. This consisted of 105 students who either agreed, disagreed or maintained a neutral position on the automated feedback from AcaWriter. The difference in the marks scored by students belonging to these three groups is shown in Figure 71. Students who disagreed scored the highest (M = 22.81, SD = 3.33, N = 21), followed by students who agreed with the feedback (M = 22.04, SD = 4.17, N = 77) and the neutral group (M = 20.79, SD = 2.23, N = 7). However, statistical tests found no significant difference across the groups - a one way Analysis of Variance showed that the effect of groups on marks was not significant, $F(2,102) = .74$, $p = .48$. This indicates that whether the students agreed or disagreed with the automated feedback from AcaWriter had little effect on their writing performance, but it might be the case that their engagement with automated feedback mattered the most.

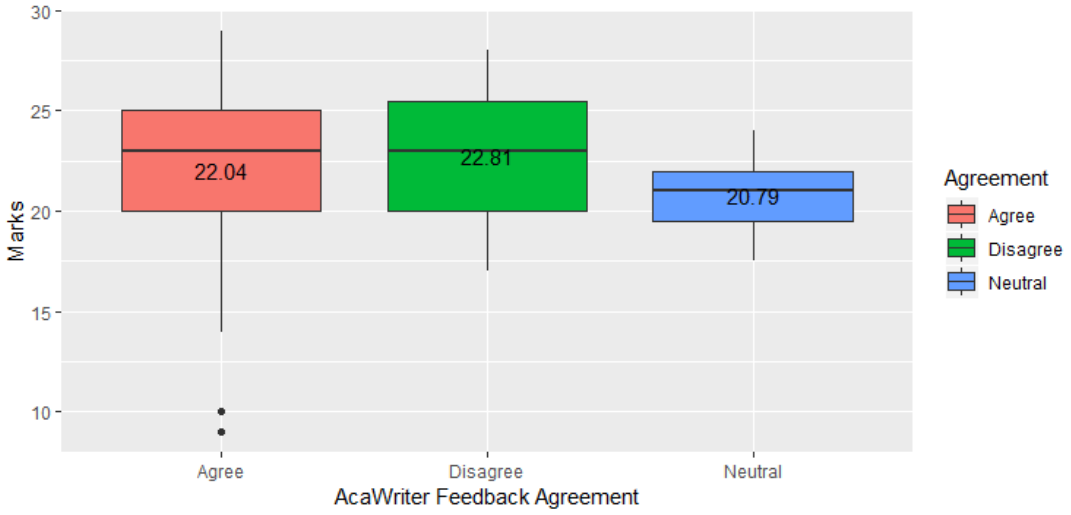


Figure 71: Student marks across different groups of agreement with AcaWriter feedback

To further explore student engagement with automated feedback with respect to their performance, their understanding and application of feedback categorised in the previous section was studied. The difference between the shallow and deep understanding groups in terms of the marks scored is shown in Figure 72.

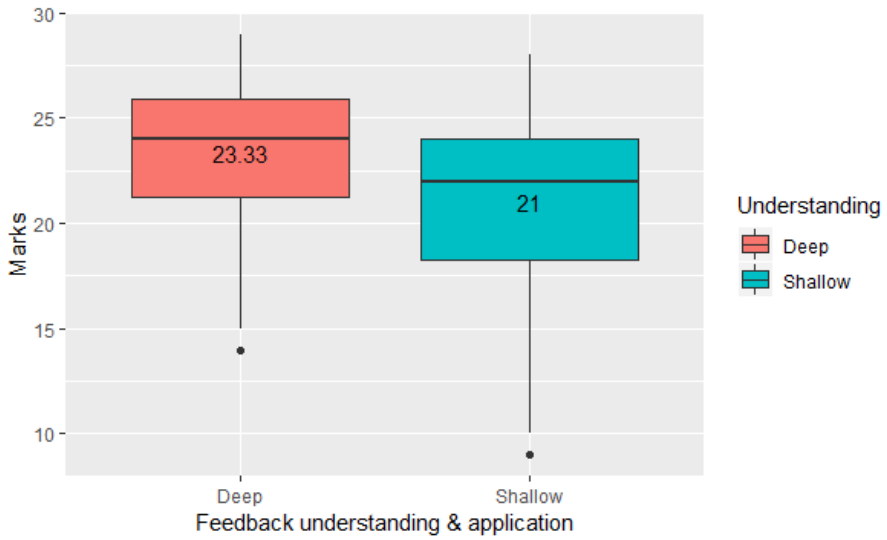


Figure 72: Student marks across deep and shallow groups of AcaWriter feedback understanding & application

There were differences in student marks between the shallow and deep engagement groups based on their understanding and application of automated feedback from AcaWriter. Students who had deep engagement with AcaWriter feedback scored higher (M = 23.33, SD = 3.29, N = 50) than students who had shallow engagement (M = 21, SD = 4.14, N = 55). A significant difference was found between the

groups in a Welch two sample t-test: $t(101) = 3.21$, $p = .002$, with a medium observed effect (Cohen's $d = -.62$, 95% CI [0.22, 1.02]). This suggests that students' level of engagement with feedback is related to their writing performance, and supports the case for the need to improve students' feedback literacy (Carless & Boud, 2018) when using automated feedback for effective writing support.

5.2.7 Discussion

This iteration validated the transfer of the existing writing intervention using AcaWriter from the Law context to an Accounting context, exemplifying the Contextualizable Learning Analytics Design (CLAD) model. The CLAD elements Assessment, Features, Task Design and Feedback played a role in contextualizing learning analytics using AcaWriter for this new context. The successful implementation of design transfer demonstrated the use of such design representations to reuse and adapt existing patterns for other learning contexts. This iteration exemplified the use of AcaWriter over a longer period of time for students to apply on their own writing assignments in addition to the provided exemplars. An additional scaffolding was also provided to students in order to better engage with AcaWriter feedback using prompts as part of their written assignment.

Results from student perceptions of the online writing intervention were generally positive and showed that students found it useful to improve their writing. Prompts provided to students to enhance their engagement with automated feedback from AcaWriter made visible aspects of students' understanding of feedback and their application of feedback. Tools like these can not only aid students to better engage with automated feedback but can also help instructors and researchers uncover processes like student engagement to make student thinking visible. Most students tended to agree to AcaWriter feedback when using it for their written assignments, but a mix of shallow and deep engagement with the feedback were observed among the students.

In studying students' performance in terms of the marks scored for their assignments, a small effect was found based on their online activity completion. Also, whether students agreed with AcaWriter feedback or not had no effect on their performance. What affected student performance was how engaged they were with AcaWriter feedback by understanding and applying it on their writing. Students who demonstrated deep engagement with the feedback scored higher marks than those who demonstrated shallow engagement. This shows that regardless of students' agreement/

disagreement with AcaWriter feedback, what mattered the most for students to perform well in their written assignment was their ability to engage deeply with the feedback. This concept of automated feedback literacy by making sense of the feedback, and critically engaging with it to apply it on their writing needs to be explored in more detail in future work in order to make the use of automated feedback effective for all students.

Findings so far from the implementation of AcaWriter in two learning contexts have illuminated the impact of automated feedback on *students*, answering the first two research questions. The next chapter aims to examine the third research question which focuses on insights from *practitioners* in these implementations.

Chapter 6: Practitioner Perspectives

As discussed in the research goals of the thesis in Section 3.1, this section will explore practitioner perspectives and outcomes in addition to the results from students examined earlier[16]. To do this, a post-implementation case study with the instructors and tutors was carried out to explore educator perspectives in the design and implementation of AcaWriter at the end of all design iterations. This chapter will explain the rationale behind this study and its findings to answer the third research question.

6.1 Rationale

The bulk of empirical LA studies focus on student outcomes and performance measures using tools with little focus on the barriers to adoption of these tools like practitioner involvement (Klein et al., 2019). Among the key groups of stakeholders, including Learners, Educators, Researchers, and Administrators (Romero & Ventura, 2013), there is limited research on the role of educators in integrating learning analytics in authentic practice. Among the notable studies that explore teacher interaction, two approaches have been taken. A set of studies have investigated teacher design and inquiry processes, for example in a professional development design workshop on designing for learning using learning analytics (Alhadad & Thompson, 2017), while the 'completing the loop' project investigated how learning analytics from Learning Management Systems (LMS) could be delivered to support teachers (Corrin et al., 2016). This was done by interviewing university teachers to understand what analytics they would find useful, and implemented a tool to deliver meaningful analytics.

Other studies have investigated educator perspectives and sense-making on learning analytics, for example regarding information from LA focused on online collaborative learning (van Leeuwen et al., 2017). Similarly in a study by Holstein et al, teacher expectations were explored to see how Intelligent Tutoring Systems (ITS) with real-time dashboards could be designed for blended classrooms (Holstein et al., 2017).

[16] Parts of this chapter have been submitted for publication in a journal article (Shibani et al., In Submission)

This study used design interviews including generative card sorting exercises, semi-structured interviews, directed storytelling, and Speed Dating sessions to understand teacher notions before designing an ITS for them. Other studies have examined teacher views to qualitatively evaluate the usefulness of LA applications after they are implemented (Echeverria et al., 2018; Koh, Shibani, et al., 2016). More generally, the sub-field of teaching Analytics has focused on capturing and analysing teacher actions to help teachers improve educational designs prior to their delivery (Sergis & Sampson, 2017).

While these studies give insight into how educators reflect on specific LA applications, they generally do not provide detail about long term usage, or the ways that educators adopt and adapt learning analytics to their specific context. Thus, across this body of work the perspectives of teachers regarding learning analytics and their use in existing practice is limited. Addressing this gap involves bringing relevant LA to practitioners at an implementation level, and supporting them to contextualize the LA to their pedagogic context (Shibani, Knight, et al., 2019). This can be supported by inter-stakeholder communication during the process of adopting an LA innovation (Prieto et al., 2018). This participatory approach emphasizes the need to take into account practical considerations of different stakeholders to bring LA in authentic practice for making it effective for learners (Greller & Drachsler, 2012).

To enable greater opportunities for cross-fertilization between research and practice in learning analytics, this study focused on the perspectives of educators, who are one of the main stakeholders to effectively implement learning analytics. It aimed to uncover the key motivations, implementation process, outcomes, challenges and support required for educators from the writing analytics example of the thesis. By bringing to notice these important, but underrepresented views, this study will expand our understanding of learning analytics in authentic practice and answer my third research question on practitioner perspectives of writing analytics applications in the classroom.

6.2 Methodology

To investigate this research question, qualitative research methods are used. There is a concern that the favouring of quantitative approaches in learning analytics might lead to a focus on finding generalizable structural relations rather than understanding

nuanced processes in learning (Wise & Cui, 2018). The value in qualitative research methods is that they provide insights that are deeper in nature with interpretive rich descriptions (Erickson, 1985). Such rich insights are hard to obtain from quantitative methods by analysing large number of participants who provide short responses.

Data for this study comes from semi-structured interviews conducted with educators who worked with the writing analytics application in their classrooms (Interview guide approved by the institutional human research ethics committee is provided in Appendix C). These educators were involved in the co-design of specific feedback modules in AcaWriter, and writing tasks for students to make use of the automated feedback on their writing. The interviews were conducted after in-class implementations of the tool over a number of semesters, and invited the instructors to reflect on the whole process and bring forth practical issues and considerations. Ethics approval was granted for the study, and participants consented to their data being used for research purposes, including acknowledging that their quotes would be used in a way that would be re-identifiable (because it is public knowledge which academics have worked on the project, including through co-authorship of papers). Descriptions of the interview participants are provided below:

I1 (First Instructor) – Law Academic

I2 (Second Instructor) – Business School Academic

I3 (Third Instructor) – Business School Academic

I1 was the first educator who co-designed the intervention in class with researchers, over five semesters to develop a stable learning design for Law students. I1 was interviewed via videoconference. I2 and I3 worked together to implement an adapted version of I1's learning design by tuning it for their classes in Accounting. They were interviewed together in a face-to-face session.

The interviews lasted approximately one hour, which were video and audio recorded, and transcribed for analysis. This interview data was qualitatively analysed by coding for key themes that emerged from the data. It followed the method of extracting key issues and insights related to the specific research questions deductively and deriving themes from each participant's interview to create a shared understanding of collected interview data inductively. The aim of this approach is not to create an exhaustive coding scheme, but to provide an interpretive typology to understand the interview case data. The main themes identified from the data addressing each research

question are discussed in the following sections. Excerpts of the quotes from the instructors are provided as exemplifications of themes identified[i]. Additionally, the tutors who helped facilitate the intervention in classrooms along with the instructors were sent a survey for their feedback.

6.3 RQ3a) Factors influencing adoption in authentic classrooms

To investigate factors influencing learning analytics (in this case, AcaWriter) adoption, instructors' motivations, challenges and support required were studied. The findings are explained as follows.

6.3.1 Motivation

The main motivation for instructors to use writing analytics in their classroom fell under five main themes. They are explained in detail below:

6.3.1.1 *Improving students written communication and self-assessment*

The instructors were keen to improve students' disciplinary written communication. They wanted to provide support to students to develop this skill and to learn to self-assess their writing better. One instructor was particularly keen to teach students how to assess their own work to build their evaluative judgement (Boud et al., 2015) and reduce the number of remark requests received; aims that are not typically explicit targets of LA interventions: *"part of the issue was throughout the accounting degree they don't necessarily get support and so they land in the subject without having lots of practice and developing their [writing] skills. So, we wanted to really push that." [I3]*

"their writing skills are generally pretty high, but the reason that I wanted to use AcaWriter was because I found, when I first started teaching this really large cohort, that there were an extraordinary number of requests from students after they had received the mark for the essay for either a remark or an explanation as to why they had not achieved a better mark [...] I just thought wouldn't it be great if we could find a way to get them to do some self-assessment and reflect on their essays a little more thoughtfully so that they could reach the conclusions themselves as to how they could have improved." [I1]

"we are responsible for improving students' written communication skills, and so we're really open to working with a technology to help us do that [...] we are responsible for assuring one of the programme learning objectives [...] which is around assuring that students can have the skills and capabilities to be able to communicate in a written form in a way that's appropriate to a business audience as well." [I2]

6.3.1.2 *Direct feedback for students*

The instructors valued the provision of automated feedback which AcaWriter offered. Students could get this automated feedback directly from the tool anytime, freeing the tutors from providing formative feedback to individual students. They were ready to test the ideas that there were certain aspects of writing amenable to immediate, automated feedback.

"we wanted to provide students with the formative feedback before their summative assessment, so that they had an opportunity to recognise parts that they can improve and build on [...] In engaging with AcaWriter we've helped the students get feedback directly on the piece of work they're working on and they can use the tool directly. [...] the broader motivation was [...] to provide feedback to students on their written communication that did not require the tutors to have to mark-up reports and provide that back." [I2]

One instructor also felt that this would remove the need for tutors to develop expertise in giving feedback on writing:

"And, also wasn't necessarily reliant on the tutors having to develop expertise around providing feedback around written communication because that's not necessarily their core expertise. So, I was really open to the idea of using a tool that could actually do that for us and that the students could use themselves at any time" [I3]

They particularly thought the automated feedback would be useful to teach students how to better structure their writing, while the content could be taught by the tutors or subject experts. They also thought that it would be complementary to the other kinds of feedback students get on their writing from other tools, peers or tutors.

"assessing the merits of their arguments is something that is very difficult to automate. But assessing whether or not the essay has certain features and follows a certain structure to me is more mechanical. So just as you might use Grammarly or a grammar checker with Word or you turn it in to check the originality score of your essay, I thought wouldn't it be great if we could use this tool, this writing analysis tool, to automate some of the feedback." [I1]

"in terms of written communication you've got content versus structure and I think Aca Writer is bringing more of the structural improvements, whereas we have other activities in-class so we talk about ideas around the topic and that's more content." [I3]

"...compared to other writing analysis tools that are available freely online or even things like spellcheck and grammar, the appeal of Aca Writer is that it does deal with more structural elements that sit within students' writing and so it can act as a complement to other tools that can provide feedback as well." [I2]

6.3.1.3 *Saving time*

A major motivation for instructors was to save time when supporting large student cohorts. Requests for further feedback and remarks require considerable resource. The instructors believed that the intervention and the tool would help reduce these numbers and save time.

"But we can't afford to do that [giving formative feedback] when we have 400 students because it already takes us maybe about 20 hours to mark one class [~35 students] of these assignments and so we can't have the tutors spend that time again giving formative feedback. So, we had to do it in a way that is time-efficient." [I1]

"The sorts of numbers we're dealing with can be anything between 280 students to 420 [...]. And I would say prior to introducing AcaWriter in the most meaningful way, which was in the last year or so, we would have up to I would say 20% wanting more information, wanting a remark, complaining about their mark, wanting their essay redone. [...] maybe as few as a tenth, but never fewer than one in ten wanting more information. In a cohort of 280 students, you've got 28, often 35 students wanting more information." [I1]

6.3.1.4 *Instructor knowledge of writing and motivation to support it*

The instructors who trialled writing analytics in their classroom were generally particularly self-motivated to improve student experience. Their interest in delivering student-centred learning put them in the forefront of other innovative practices when opportunities arose.

"it's really left to us to drive that and it's left to individuals to drive it. We weren't given any encouragement to go or push to go do this. An opportunity came up and we thought it was a good idea, so we went for it." [I2]

"It's nice to get some external recognition [a teaching award] but that's not really part of the motivation to do it. The motivation is really student-centred, trying to figure out how best to develop students' written communication and support that in different ways. And being just really open to technology as a potential solution." [I3]

"we've got an award for stuff that we've done in the past but that's just a by-product. The reason we have done all this stuff is because we are really passionate about student-centred learning and so that's really where our strong motivation lies." [I2]

6.3.1.5 *Openness in the role of technology*

The instructors were curious to explore the potential of writing analytics technology to develop student writing. They had prior experience in using technology in the classroom, and were comfortable in using educational technology to innovate in their classrooms. This made them more open to trying new techniques to improve their teaching practices.

"That's the first time I've used a writing analysis tool, but I have used a lot of technology in the classroom for various purposes, sometimes to give the students an opportunity to engage in or practice authentic way with the kind of tech values in practice […] I just wanted to convey here that I'm pretty techy, like I'm comfortable with tech" [I1]

"I think we've got a long history in our subject in trialling and experimenting various different innovations in our teaching. And that could be in terms of activities or it could be in terms of trying different ways to teach particular concepts […]. we're definitely very open to trialling different technologies and to putting them in place and seeing how they go. […] I was really curious about the role of writing analytics" [I2]

Some of the innovations they had previously employed in their classrooms included the following, ranging from low tech to highly sophisticated technologies:

- Google Docs to build on peers' works
- Running a negotiation activity with an online dispute resolution tool
- Tracking feature in Word to mark-up amendments
- Audience response systems like Mentimeter
- Online quizzes
- Designing apps for social injustice
- Wikis to support in-class activities
- Using business intelligence systems

6.3.2 Challenges

The instructors also noted challenges in implementing the writing intervention. These are useful to take note of so that the educators are better supported to overcome these challenges.

6.3.2.1 *Effort involved*

The instructors thought that the setup of the intervention, and related work took a lot of time and effort. This was in addition to their normal teaching load, and added up work with a number of things they had to prepare before, during, and after delivering the intervention.

"It took a lot [...] I would say the main challenges were the ones that were at the real cutting edge of what we were trying to achieve, and that's where the excitement was and that was where the magic was, so that was a labour of love [...]. There was a lot of work involved in the design process and dealing with the students, dealing with emails, recording instructions, writing them up, trying to get buy-in from my tutors, from the tutors who also teach this subject [...] And I would say I spent at least six hours just on the proposal, but I would say no less than about 40 hours a semester was spent on this, which, if you think about it, is huge." [I1]

For the Accounting instructors, a number of materials had to be prepared to adapt the Law intervention for their context, the set-up cost of which was high. This included:

- Finding samples of student work, getting permissions to use them as exemplars, and de-identifying them
- Marking the samples and providing example feedback for students to learn from
- Creating a sample text that the students could improve through redrafting
- Conceptually aligning AcaWriter feedback and the assignment needs
- Preparing a script to explain the relevance of technology to students

"I probably underestimated the setup cost [...] if I think about all the things that an academic has to do to adapt to do" [I2]

It is to be noted that much of this work would be required if instructors were aiming to develop students' writing and assessment criteria literacy, regardless of whether an automated feedback tool was used. However, the key difference is that additional materials were required from the instructors for this validated task design.

6.3.2.2 *Tutor involvement*

The instructors noted that tutor involvement could be a potential factor that affected how the intervention was delivered to students. This is because the tutors are the people who facilitate activities in some classes, and their involvement thus plays a part in students' engagement. Thus, in analytics interventions, engagement and support beyond the primary instructor may be important (but is understudied).

"I'm just wondering to what extent do we, maybe need to get the tutor team more involved in the sell. I don't know, I always just go back to that level because that's the intermediate level that we forget quite often. But if the tutors can't sell it, given that they're going to be the face to face, more so than even a lecture. We did obviously let them know but to the extent of whether that could have perhaps been stronger, I don't know." [I3]

6.3.2.3 *Imperfect tool*

The instructors noted some flaws in the automated feedback from AcaWriter in terms of correctly identifying rhetorical moves in student writing.

"I don't think it does it particularly well yet. I still think it's flawed, but I think that the fact of it as an intervention and the fact that it does it partially well is pretty amazing, and I think it's really profound and impactful. So, yes, it's a really valuable tool and I think if you took it away, we would lose something of value in the step towards improving student writing. But, obviously, it's not perfect. I actually think the fact that it's not perfect, which, let's face it, spell check isn't perfect, Grammarly isn't perfect. All they ask you to do is think about it [...] And I know what Grammarly's doing, and I know why I would override what Grammarly suggests. Now if that's what the students are doing, well, more power to them, but at least they understand what their text is doing and how it's behaving." [I1]

A specific example provided by an instructor was AcaWriter's inability to distinguish between use of the word 'innovation' with reference to content (e.g. "Their major innovation was…") and rhetoric (e.g. "We provide an innovative tool").

"I am still slightly concerned with how Aca Writer recognises things and so I think one of my feedback was one of the rhetorical moves when it recognises you've got a new idea in your writing. So, the things that were highlighted as new in our sample reports, the thing that I found common with all of them is that they all had the word innovation in them and so Aca Writer is going they must be talking about a new concept. But Aca Writer,

in my opinion, was failing to distinguish between a new concept as a structural element versus innovation as the topic and the academic content of the report." [I3]

Even though the instructors recognize the flawed nature of the tool, they believed that the imperfect tool facilitated students' understanding of their own writing and taught them to think about it critically. This ability to critique automated feedback was found to be of value to students because it led to a deeper understanding of writing concepts. It supports the argument of giving tools to students to help with their learning even if the underlying algorithms are not perfect by clearly explaining its usage (Kitto, Buckingham Shum, & Gibson, 2018). This is because students then learned to look for these concepts in their texts, and identified ways to signal elements of writing using those newly learnt concepts.

"I think the real value for us is actually telling the students that it's an imperfect tool. It's a tool that can provide them with one source of feedback, but they need to use that feedback critically. And I think the temptation to just use it as an algorithmic assessment takes away the critical distance they have from the feedback." [I2]

"And part of what we want to do in the subject is for them to develop their critical thinking skills of questioning what's in front of them and saying actually I don't agree with that. And so, I think it's really important to keep that level of thinking." [I3]

6.3.2.4 *Teaching appropriate use of feedback*

A related problem with having an imperfect tool was for instructors to teach students how to effectively use the feedback from the tool. The appropriate use of feedback was not observed in all students since there was a varied level of understanding. An instructor noticed that some students engaged with the feedback at a surface level, spotting the presence or absence of moves only, and did not engage at a deeper level with a critical eye.

"they're engaging with the highlighting that AcaWriter can give them in terms of the different rhetorical moves. But they didn't really engage much with the additional feedback that AcaWriter popped out" [I2]

There were also students who were too critical and dismissive of the feedback. So a balance is needed to explain the value of automated feedback, even if it is imperfect.

"they're probably still learning how to. It's a tricky thing to balance in terms of we want them to actually be critical of the feedback they get from AcaWriter, because by being

critical of the feedback we actually force them to justify their position even more. Yet, we don't want them to be dismissive of the feedback they get from AcaWriter and I think we have to strike that balance [...]. So, I think there's still a bit of work to be done in terms of students actually engaging with the written communication, the value of the product." [I2]

Another instructor thought that the associated risks and imperfections of automated feedback should be explained to use it with caution, and emphasized that students should learn how to apply their human intelligence. They noted that the students should be taught how to properly engage with the feedback for them to fully understand the value of the tool to improve their writing.

"I would say to students in the future that they should see it as one of the four main tools that they would use in technology, always thinking about writing analysis and how if you were going to rely on it for producing an outcome that was deterministic rather than probabilistic, that you've got to think about the risk associated with that because of its imperfections at this time, but that it should be used. My idea would be that it should always be used in conjunction with human intelligence." [I1]

6.3.2.5 *Disrupting disciplinary teaching and research*

One instructor noted that the intervention was disrupting the teaching of a core discipline-related subject, by increasing its focus on writing.

"As to whether it made things more streamlined, whether it improved efficiencies, no, it is the biggest disruption. Because you're talking about taking a cohort of students and basically having to fill in their learning with almost a different subject. It's like I'm having to teach them English and English language parts of speech." [I1]

This concern was observed in a new subject co-ordinator as well who preferred to not run the Law writing intervention during class hours, leading to its delivery in a fully online setting. However, due to low uptake for the optional homework activity in that context with only a small number of students completing the task, it is not discussed and the data is not evaluated in the current study. This shows that instructor involvement, and incentivising students to complete the activity for improving their writing skills reflects on student uptake of the intervention.

Relatedly, within the academics' own disciplines the work they undertook to implement AcaWriter – as educational research – was undervalued, providing educators less academic credit compared to discipline-specific research. Encouraging

learning analytics innovation and implementation may be supported by faculties that support and incentivise teaching and learning innovations and scholarship.

"there is a very strong sentiment in the law schools in Australia [...] against writing up teaching pedagogic, teaching research, educational research. We are supposed to be writing up our research as experts in our particular substantive area of law." [I1]

6.3.3 Support for academics

The instructors talked about support mechanisms which helped them in the process of designing and implementing the intervention in their subjects.

6.3.3.1 *Collaboration & support*

The instructors felt supported by researchers and valued the collaborative nature of the research project. They believed that this support encouraged them to build authentic writing support for students with enthusiasm.

"I felt enormously supported [...]. The entire [analytics team...] were absolutely instrumental in solving some of the key problems I faced. But I just want to be clear that I was supported the whole way in keeping this really authentic for the students."[I1]"it was really good. So, I think it was really collaborative and I think what was really encouraging was the enthusiasm to get these things into place. I found working with the researchers, they were really responsive, we could have weekly meetings. We divided up the tasks and so it was a pretty ambitious project." [I2]

6.3.3.2 *Responsiveness*

One instructor thought that it was instrumental to have a responsive team to remedy the glitches as and when they occur. These quick responses helped them increase the motivation of students, and maintain credibility.

"One of the strengths of the project team was how responsive the researchers were, particularly in troubleshooting. [...] We've got 400 students who we're trying to maximise their motivation to do their homework and to do their assignments. And any glitch needs to be remedied straight away, otherwise we risk losing credibility and the motivation of our students." [I2]

6.3.3.3 *Agency*

The instructors thought that they had a lot of agency and power in designing the different parts of the intervention. This is important because the end goal was to support their students, in a subject they taught with first-hand experience.

"I always felt as though the design, the content, the pace, the way we engage with the students, I felt I was deferred to incredibly respectfully the whole way as the lead on this even though in truth we were all collaborators of equal input. But I felt I was given an enormous amount of agency because these were my students and my subject, and I really appreciated it. And I felt the outcomes were better, all the better for that." [I1]

"Yes, absolutely [had enough agency in how the intervention was designed for our class]" [I2]

6.3.3.4 *Wider adoption*

As instructors who were early adopters of the AcaWriter tool and having experience in implementing writing interventions for students, the instructors suggested possible routes for wider adoption of the tool by other academics. One instructor thought that it was important to explain what AcaWriter does and make it clear that it is available for use, which would then encourage academics to try to solve their problem using the tool. It was also suggested to figure out disciplines where such technology might already be in use in industry, so an authentic practice could be delivered using the tool.

"I think the first step is to first of all, all of those academics who don't know what it does or how it does it need to have that explained. And then it's important [...] to listen to those academics and find out what's the particular problem they would like to solve and then take it from there. [...] I think it's more important to say to as many academics as possible, we've got this tool. This is how law used it. This is what it's done for law, but there are many other problems it could solve. Do you want to go away and think about whether you could use a writing analysis tool? [...] I would also try and find out how the particular industry that they support, that that faculty delivers graduates into is already using writing analysis software to give it some practice or authentic meaning." [I1]

Another instructor suggested the creation of adoption packages for academics so they could easily use different versions of AcaWriter – with or without the fully designed intervention. It was also recommended to explain the requirements with a list of items the academics needed to prepare, so they would be aware of their commitment.

"I think you could put together a couple of options in terms of the packages, what it would mean to adopt AcaWriter..... Because I think probably the biggest hurdle for adoption is people not having a sense of what it is they're committing to in terms of getting it in place. So, if you can be really upfront in terms of, okay, if you want to have it just sitting there for an assessment task then they key thing you're going to have to do is just help us with the mapping and the providing of feedback. But if you want to have an online tutorial then you're going to need to do all this. And so, then they have a sense of what it is exactly they need to do in order to customise it for them." [I2]

An instructor mentioned that the benefits for academics to be involved should be explained. The academics who are responsible for improving students' written communication can be tapped into to provide a solution to their problem.

"And also just make sure that you explain the benefits, both for the academic and the student, because obviously we're here to benefit students but if there's nothing in it for the academic they're like, why bother? Why should I do it? [...] I think you could tap into academics that do have to, like as a first port of call, that are responsible for written communication [...] And so, you go to the ones that have that need." [I2]

The instructor also added that the academics would be interested in being part of a broader research project. Being able to view how the trials work across faculties would help academics be aware of what others are doing, and encourage them to use it for their subject.

"The other thing would be potentially bundling this all up as part of a broader project or research project that's looking at how we support written communication in different disciplines. And so, for them to see the broader project and to see the value of that, I think would be attractive to some academics so they feel that they're not just doing this by themselves but they're doing it as." [I2]

6.4 RQ3b) Implementation by practitioners and outcomes

To answer the second part of RQ3, this section explores the implementation strategies and processes used by the instructors and the outcomes observed in their classrooms with automated writing feedback.

6.4.1 Implementation

While implementing the intervention with the use of writing analytics tool for students, the instructors were involved in a co-design process with the researchers over a period of time. The instructors talked about the key points to be taken note of in this process of designing and implementing the intervention in their subjects and the support mechanisms which helped them.

6.4.1.1 *Design process*

The instructors were involved in a design process that iteratively and incrementally tested and evaluated new ideas. They identified areas where they could improve their existing pedagogic writing practices using the analytics offered by AcaWriter. This follows the strategy of augmenting existing pedagogically sound good practice with affordances of LA for better adoption, rather than revolutionizing those practices (Knight, Shibani, et al., 2018). The design process of the intervention was different for Law and Accounting disciplines, with more iterations of the current task design in Law when compared to Accounting. There were a few failures in the initial Law trials, which led to a stable design iteration. These iterative design stages – rarely reported in research literature – are an important and understudied component in understanding the practice of designing LA for authentic scenarios.

In the first trial without the tool, the students were simply asked to self-assess their writing at submission. However, many students did not participate:

"I knew that some 10% did it, seriously 10%, so it was pretty hopeless. It was all a bit of a disaster, but that was just the first semester." [I1]

In the following semester, a voluntary group of students were introduced to AcaWriter, to support this self-assessment:

> *"I incentivized that by giving them a promise of a little script of text [for their CV] […] So that control group was fantastic. I had about 30 students and they were very critical. They were really critical of the whole app, the use of it." [I1]*

To improve students' understanding of the tool, the instructor then created a tutorial which explained rhetorical moves by taking into account feedback from linguistic experts.

"So, then what we did was had a bit of a discussion. I was a bit disheartened. I presented an initial set of results to my colleagues […] she said, […] the reason they don't understand what the tool is doing is because they don't understand the [rhetorical moves]. She's a linguist and she's an expert in all things linguistic, and I just took it on board" [I1]

However, the instructor explained briefly that this small addition of a short video tutorial on rhetorical moves made little difference. The instructor at this point came up with a document to explain rhetorical moves, and collaborated with the analytics team to co-design a writing intervention, detailed in an earlier work (Shibani et al., 2017). As the instructor noted, implementation AcaWriter in her teaching practice was thus a multi-faceted intervention.

"You've got the intervention of explaining [rhetorical moves], the intervention of demanding the self-assessment in order to get the remark, the intervention of your benchmarking activity, the intervention of using AcaWriter itself." [I1]

The Accounting instructors then took their existing practices, and integrated approaches from Law to implement the intervention based on that evaluation. This method of transfer from one context to another is an effective method for adopting LA to scale to more students (Shibani, Knight, et al., 2019).

"our approach to the subject, is we have incremental changes, that we're always improving and evaluate the things that work and keep the things that work and change the things that need changing." [I3]

As in Law, they had previously trialed the usage of AcaWriter in the following ways:

- As part of a benchmarking exercise teaching self-assessment through assessing exemplars, with AcaWriter feedback marking up exemplars for one group. However, at that stage *"it wasn't actually integrated into their assignment" [I2].*
- This exemplar marking exercise was retained, but without AcaWriter use in subsequent semesters
- Subsequently (for two semesters), they asked students to use a technology to support self-assessment of a draft prior to submission, with AcaWriter as one of the feedback options.

- Most recently, they adopted the approach from Law with writing activities, and the use of AcaWriter as a feedback tool that students use to improve their drafts.

"then at the beginning of this year I think we started to reconnect in terms of some of the things that have been happening in law [...]. It was really pleasing to see that this model of having week one activities in class where students are doing benchmarking. So, they're getting exposed to marking criteria, they're thinking about the marking rubric, they're looking at samples of students' work. And then they're actually using these sorts of tools amongst others to improve a draft of their assignment before they finally submit. That sort of model seemed to fit really nicely around some of the innovation that's happened since then in Law." [I2]

6.4.1.2 *Authentic experience*

A constructive student experience was important for instructors who emphasized not wanting students to engage in activities solely for research purposes. With this motivation, they co-designed authentic experiences for students by aligning formative, writing analytics augmented, tasks to their existing assessments with help from LA researchers. In that way, they thought that students could apply the skills learnt practically to their subject assignments.

"I had to make sure that when the students were using AcaWriter and trying to get feedback to self-assess that it was a genuinely constructive experience for them in the writing process [...]. I did not change the essay or the essay question and I only changed the criteria in the slightest way. The marking criteria only changed so that the explanation and description was clearer but the actual criteria against which they were being assessed did not change." [I1]

"the primary use of Aca writer was to support the students' development of skills inline with an assessment piece [...] the aim is to help students develop their ability to communicate, their written communication, in a professional business manner. " [I2]

6.4.1.3 *Explaining the relevance*

While designing this authentic experience, the instructors thought that it was important to convey the relevance of the technology, and the rationale for the intervention to get buy-in from students. If the relevance was not properly explained, there is a chance that students would not regard it useful. The instructors had created videos explaining this relevance to their discipline of study, and aligned the intervention closely to their

assessment criteria. They believed that if this alignment was not done properly, students would not be engaged and rebel against doing the activities.

"the reason why I had to be so explicit with the students about that is because these guys are very complementary and they're very, very judgmental of the way that they're taught. We're talking about expert learners who think about the way they're being taught and can be very critical and they will complain if they feel that they're being used as guinea pigs in a project that has nothing to do with their learning. They'll quickly rebel, and you don't want to lose them. So that's how I got the buy-in from the students." [I1]

"I think that's sometimes the risk with new technologies coming into the curriculum, is people just go, great, that's an excellent technology. […] The power in using these technologies is having that really close alignment between your outcomes and how your tool helps solve those outcomes. And that thought-process, it takes a while to get that fit. And if there's no fit, the students just go, pffff, what's the point? And so, it's getting that alignment that is the key." [I3]

"we had to figure out what the specific alignment was going to be in terms of the rhetorical moves that AcaWriter could identify and how that married up to the assessment task, to the report we're asking them to produce. As well as what sort of feedback we want to be providing to students. Even on the other side as well, it's preparing a script to explain this to students, having to record that as well and figuring out the sequence of those and then having to fit that." [I2]

6.4.1.4 *Incentivizing students*

Since the intervention collected data for research, it couldn't be made mandatory for all students in Law. In Accounting, all students had to complete one activity reflecting on AcaWriter feedback as part of a self-evaluation exercise which was assessed, but the online intervention task was still optional. In addition to the steps taken above to get student buy-in like creating an authentic experience and explaining its relevance, one instructor also trialled and implemented some ways to incentivize students in the different iterations. This included creating a condition for remarking request, and a providing a script certifying the student assisted in research for their CV.

"There were 280 of them and we said to them, we want you to self-assess and to incentivize the self-assessment process, I attached a condition to it, to the whole process, that if they did not submit a self-assessment, then I would not remark their paper. That

was the ticket to getting a remark. Now, of course, I did remark some of the papers even though they were not self-assessing[…] I would say, yes, you will get a remark, come and talk to me, but bring the self-assessment with you and even if you haven't completed it, we'll complete it together. So, I still made them do the step to buy the privilege of the further discussion and the remark, so that was interesting." [I1]

"Following semester, did the same again but started rolling out a new version with a voluntary group for AcaWriter and I incentivized that by giving them a promise of a little script of text, like three lines of text that they could add to their resumes so that they could note that they'd assisted a researcher in the University of Technology with the development of some writing analysis software. So that control group was fantastic. I had about 30 students and they were very critical." [I1]

6.4.2 Outcomes

The instructors identified a number of outcomes that emerged while implementing the intervention. These outcomes affected stakeholders including their students, themselves, and the research team, and fall under themes (aligned with instructor motivations).

6.4.2.1 *Students' self-assessment and engagement with writing*

Instructors from both subjects mentioned that they saw improvements in students self-assessing their writing, and were more aware of the assessment criteria. Preliminary findings have also shown students' engagement with the activities in the online system, and in the reflective activity, but to varied levels (Knight, Shibani, et al., In submission; Shibani et al., 2017).

"once they truly understood what the tool was doing […], the only comments we were getting in the feedback were things like, oh, it's interesting to note that AcaWriter didn't pick up what I would have thought was quite a good rhetorical shift because I said, it's interesting to note, but AcaWriter didn't pick it up, but I'm not quite sure why. And suddenly you realize, bang, they've got it. They're all over it. And they are now self-assessing. They're now reading their essays really critically looking for something that prior to what we had created they were never looking for and they didn't even really know to look for it at all." [I1]

"they can reflect on the criteria that they're being assessed on or at least be aware of the criteria they're being assessed on, particularly in terms of written communication. So,

they're starting to think about what we're looking for when it comes to good written communication." [I2]

"we can see, and you've shown us some of the data as well in terms of the level of engagement with the system. So, at base level, we're getting the students to do the online tutorials, they're engaging with in class. Most of them are actually engaging with as part of a reflective writing exercise to improve their final reports and most are starting to see some of the value. Although, I think again, they're probably still learning how to…" [I2]

Students also seemed to reflect on their writing more generally, which would lead to improvements in their written communication in the long run.

"they have a more personal, reflective outcome or a personal reflection about their own skills. And about areas more generally where they might look to improve their written communication. So, they might identify a particular weakness in their written communication more generally that they could then think to address over the longer term[…] if we go back to a scenario where you're just asking students to hand in an assignment and hope for the best, absolutely our students are much more reflective about their written communication." [I2]

6.4.2.2 *Improvement in student writing*

One instructor noted that there was an improvement in performance on students' written communication over the semesters as indicated by an overall increase in marks, although the long-term trend needs to be examined further. The instructor from Law noticed an increased use of discourse markers in writing, which helped students provide a well-reasoned view, an opinion or a conclusion using explicit terms (Knight, Shibani, et al., In submission).

"Overall, since we've been working with [the analytics team] around written communication over the course of the last four of five semesters, we have seen marked improvement in students' written communication […] overall their individual assignment pass-rate is going up. Overall their mark is going up, but slowly. Marginally, but slowly." [I2]

"they were now using it in a more meaningful way, they were producing better final versions of their essays. I noticed a change and it was profound that suddenly the discourse markers were everywhere […] And suddenly you can see they're using the language that we were using in the benchmarking activity because they're smart and they're competitive

and they know how to, they're good adopters if they're told explicitly what to do. And suddenly I noticed their essays were better. And they will be better in court and they will be better lawyers for it." [I1]

6.4.2.3 *Direct feedback for students*

The instructors thought they were able to provide feedback to students directly using the tool, without waiting for tutors to respond to them. One instructor also thought that students liked using technology and it helped them to receive feedback at any time of the day.

"In engaging with AcaWriter we've helped the students get feedback directly on the piece of work they're working on and they can use the tool directly." [I2]

"I think that students like using tech because they can do it any time of night, in the middle of the morning [...] And once they understood enough about what it does to find the feedback meaningful, I got a lot of positive feedback from students about the fact that it was automated and I think that's an important part of the story because what we're talking about is saying to the students, paste your essay into this tool and then it's going to give you some feedback because your lecturer doesn't have enough time to do that for you." [I1]

6.4.2.4 *Saving time*

One instructor thought that a lot of time was saved due to the reduction of remark requests from students, attributing it to their improved understanding of the marking criteria and learning to self-assess as a result of the intervention.

"It's just an enormous commitment and, of course, this was all driven by me with two potential outcomes. One, could the students stop asking for remarking? Well, I achieved that [...] We got there. I say it was costing me 40 hours per semester, but if you look at getting, the remarking takes 20 minutes per paper. If you've got 30 students asking for a remark, that's a lot of time [...] But anyway, in an accumulative way, it should, over a 10-year period of teaching, deliver a massive saving on that time" [I1]

6.4.2.5 *Instructor knowledge of writing and motivation to support it*

The instructors felt that they had learned more about the use of rhetorical moves in writing, writing analytics technology, and how students write. They also thought that it led them to reflect more deeply on their domain-specific writing strategies while mapping their assessment criteria to rhetorical moves.

"The other thing it was delivering for me is I was learning more. It was informing my research and my learning, my understanding of how to be explicit with students, how to use writing analysis technology, what it does, how students learn. As an academic, that's a very important process. I realized where the gaps were in their understanding and my explanation [...] I learned to be more explicit and I think I have become better at pausing and asking students the right questions to work out where their learning is up to before I then assume anything or take the next steps, so it's been an intervention for me as much as it has for the students." [I1]

"I guess there was a reflection around the domain specific writing strategies [...] Because we're having to do that mapping between the rhetorical moves and the components of the assessment, that's actually forcing us to think through what are the discipline-specific rhetorical moves that need to be tailored. So, that's one thing that I definitely think I reflected on." [I2]

Furthermore, one instructor mentioned that the process had helped them facilitate a dialogue with students to teach them how multifaceted written communication was.

"It has been helpful in facilitating that dialogue with students [...] Written communication is as expansive as appropriate referencing, spelling, grammar, use of tables, graphs. It's around the overall presentation, it's around the underlying coherence of the arguments, the rhetorical strategies. It is about the appropriate use of terminology. There is a lot in that [...] it's actually difficult sometimes to communicate to students how multi-faceted written communication is [...] And so, by having these different tools and different exercises you can actually start to unpack what good, professional written communication is for students and they can see that they're actually distinct elements." [I2]

It made them more mindful of the considerations in delivering to diverse student cohorts with varying needs.

"up one end we've got students who are fantastic in terms of their vocabulary. But they might write in a way that's just really superfluous and so you're having to communicate to them that they need to be writing more simply and thinking more carefully about their structure. And down the other end you have students that are struggling with spelling and grammar, so whenever we're designing these activities, we need to be also be able to design mindful that we're delivering to a really diverse cohort and we need tools that all of them are going to get value out of." [I2]

6.4.2.6 *Writing research*

The instructors understood the value of research in writing, because they realized the importance of teaching this core skill to students. This outcome, even if not directly impacting their teaching was appreciated by them. They did not have any concerns in the use of collected student data from the current intervention, however expressed concerns in the possible automation of assessment using such tools. They thought that the tools do not have the capability to exercise professional judgement like a human, and can fail to recognize outliers. This recognition of limitations of LA is encouraging to see given that this kind of understanding would transfer to their students.

"For me, it is a core skill. We need to understand as educators what they understand about writing, what they see when they're writing things, what they're thinking about, and what they understand and what they understand are the different parts of a written piece of text and I feel that this has been really valuable, and it has justified the time and effort taken, including all the data gathering we've done. Now, we've de-identified their essays. You can't identify the students. There's nothing there. As a matter of just protecting their privacy and from an ethical standpoint, I have no qualms about collecting all that data and the analysis we did of it I think has been highly technical and quite clever [...] writing analysis is used more and more now across all parts of the law. I have my concerns with the algorithmic bias, but that's just a completely different argument and isn't something that arose with this particular activity" [I1]

"one of the big risks and concerns in this area more generally is that at some point you're going to replace, like this could be used for assessment, you could have automated assessment or algorithmic assessment of students work using some sort of writing analysis tool that actually just can assess the students written communication." [I2]

"Because the thing with algorithms is that they revert to a mean or an average. [...] if you are not average, if you're doing something that's really outside of the box, which might be brilliant, you're going to be cut down by these algorithms that can't necessarily recognise those outliers." [I3]

6.4.3 Tutor feedback

In addition to studying the perspectives of instructors who were the subject coordinators responsible for the design of the subject, feedback was also obtained from tutors who were involved in the subjects. Since these tutors act as middle men in

bringing the writing intervention to life in a face to face setting with students in some classrooms, it was also important to understand their views.

To receive feedback from tutors, survey forms were sent out to tutors who facilitated the writing intervention in the Law and Accounting classrooms with AcaWriter. The questionnaire used to collect tutor responses is shown in Appendix D. Since the survey was optional, responses from 2 tutors in Accounting and 1 tutor in Law who completed the survey are considered for the following analysis.

In the first section of the tutor interview form, tutors were asked about Writing Analytics software and the use of AcaWriter in the writing interventions delivered in their classrooms with a few questions. 2 out of 3 tutors said they were familiar with Writing Analytics software, and one was very familiar. For the next question aiming to understand the role of AcaWriter and the writing intervention in developing students' written communication, 2 out of 3 tutors said that they fully understand its role, and one said that they somewhat understand its role. All three of them said that they do not require more training to facilitate AcaWriter feedback discussion in class. In terms of the time and effort involved in facilitating the intervention in class, one tutor said that it was very easy, and 2 said that it was moderately easy. All three of them said that they required less than an hour to understand and implement the writing intervention with AcaWriter feedback.

The next section asked tutors about noticeable outcomes in students that they observed while in class and/or in marking their assignments with the use of AcaWriter to help their writing. While one tutor rated the writing intervention with AcaWriter to be useful (4), two others gave a neutral rating of 3. The reason behind their ratings were further explored from their open-ended responses. One tutor said that students' paragraphs were structured better, and since it was included as a compulsory part of an assignment (in Accounting), students reflected on and included rhetorical moves in their writing. Another tutor said that it was hard to tell if students improved their writing skills due to the intervention, but it was observed that students reflected more on their writing when examining rhetorical moves, even though AcaWriter did not always identify all rhetorical moves in their writing. The third tutor said that marginal improvements were observed, and there appeared to be more self-assessment in student writing, but students were still not including many rhetorical moves.

All three tutors said that it added value to their teaching, described by the tutors as follows:

- *"It made their assignments easier to read as the communication was more succinct and better. While I consider I am quite good at writing, detailing the rhetorical moves made me understand writing better."*
- *"It definitely added value to my teaching, especially with my marking."*
- *"I am applying rhetorical moves during discussions to provide context and enough justification on a concept."*

Finally, all the three tutors said that they were interested in using the intervention and AcaWriter again in future semesters. While two tutors did not comment on changes to be made in the future, one suggested revising the sample students were asked to review as many students did not make meaningful amendments to it.

6.5 Lessons for LA in authentic practice

Based on the findings from this study, key lessons have been identified for implementing learning analytics in authentic practice more broadly. These lessons apply to designers, developers and researchers in learning analytics to cross fertilize ideas between research and practice. They are discussed as follows:

Lesson 1: Communicate LA in all richness

One of the key lessons for Learning Analytics researchers to bring LA to practice is to communicate the possibilities of LA, and demonstrate what it has to offer to practitioners. For an educator not familiar with the field, LA may mean nothing more than a student activity dashboard in the university's learning management system, which they would be justified in concluding has few insights to offer about improving learning (Jivet, Scheffel, Specht, & Drachsler, 2018). Alternatively, LA may be resented as a new form of management surveillance through tracking instructor activity. Clearly, personalised feedback to students about a higher order competency such as writing, is a very different kind of LA, and there are many other types. By effectively communicating to educators the benefits (and risks) in using LA in training programs/ workshops that present this potential, we can be effective in soliciting collaborations and engage more practitioners (Shibani & Abel, 2018). We have seen that the educators who have effectively used LA in their classrooms become ambassadors to their peers,

who are now approaching us, which is by far the most effective strategy to grow institution-wide interest and impact. Critically, we have demonstrated moving from effective communication, to authentic engagement through co-design (Lesson 3).

Lesson 2: Make LA relevant

While a seemingly obvious statement, many LA initiatives appear to founder because LA is experienced by educators as a top-down 'ed-tech' imposition or fad, with little relevance to their needs. Our experience has been that once an LA application can be tuned to a specific learning context, through coherent alignment with learning tasks and assessment, this leads to improved uptake and effective usage by students and educators (Shibani, Knight, et al., 2019). However, this is only possible when a high level of control over the LA software is possible in the early days, in order to tune it, and when the educators are in turn ready to adjust their learning design to tightly integrate the tool into an assignment. Once that learning design pattern has been evaluated, it can then be adopted/adapted by others with less work. This requires universities to consider capability-related questions such as: (i) what capability do we have to modify the LA software? (Whether home-grown, or an external open source or commercial product), and (ii) how can we resource innovation pilots to couple early-adopter educators with LA teams?

Lesson 3: Provide agency to educators

Trust and ethical issues surrounding the use of data in learning analytics is a known concern for learning analytics practice (Slade & Prinsloo, 2013). In our research, a key finding is that educators greatly appreciate having agency over the design and implementation of the learning analytics application in their classrooms. A lack of agency leads to lack of trust in the LA application, hindering adoption. This becomes particularly important when creating LA for authentic practice, as educators are themselves the frontline representatives introducing LA to their students. Giving educators a genuine voice in the design of relevant LA is now recognised as a first order priority for LA research and practice (Buckingham Shum, Ferguson, & Martinez-Maldonado, 2019). Aiding educators to understand the complexities of artificial intelligence technologies with its advantages and limitations by working together with them, will translate to better understanding by their learners as well, since they will have a much deeper understanding of why the system behaves as it does. The method of co-design which values the views of all the stakeholders involved is deemed to be

useful to bring learning analytics to the classrooms (Holstein et al., 2017; Prieto-Alvarez et al., 2018).

Lesson 4: Provide support and collaborate

Educator agency in implementing relevant LA for their contexts is supported by collaboration and guidance in the co-design process. Quite understandably, educators lose confidence in learning technologies that disrupt their classes/assignments. With many of the most pedagogically interesting LA innovations being technically advanced, and often still lab prototypes, they can be complex to explain, maintain and run as production services. Responsive support from the LA team came through as extremely important in this study, building trust and reassurance for the educators that they were taking a responsibly assessed, ethical step in deploying LA with their students, despite the extra effort required to innovate.

Lesson 5: Engage in Learning Design practices

Engaging in practices to design for learning can support adoption of LA by not only making LA pedagogically relevant, but also by enabling a shared common language to represent good pedagogy. Representations like design patterns can be used to document methods and theory behind implementations in a common structure (Goodyear, 2005). They facilitate the transfer of patterns from one learning context/ discipline to another by preserving the theoretical underpinnings and practical considerations in a LA implementation (Shibani, Knight, et al., 2019). Resources generated as part of the process can be shared via a repository to help other practitioners (http://heta.io/resources/). Moreover, augmenting well-designed student tasks with LA provides a safety net: the students are asked to engage in a meaningful activity even if the technology is imperfect (Knight, Shibani, et al., 2018).Such theory-oriented LA, including using design based approaches helps move LA closer to foundational research on learning (Reimann, 2016).

6.6 Discussion

The uptake of new technology by students is often based on how the main stakeholders (instructors) present to them its benefits and relate it to their curriculum. In this way, instructors make a big difference in bringing those applications for practice in authentic classrooms and ensuring effective use by learners. Their readiness to try the new writing analytics application has been an important factor in our successful

implementations. From the analysis of the tutors' interview data, themes on instructor's motivation to be involved in the intervention design, implementation in their classrooms, outcomes gained as a result, challenges observed in delivering it, and support required for academics to effectively implement it emerged. The motivation of instructors to be part of the process of co-designing and implementing the writing intervention were: improving students' written communication, curiosity in the role of technology, saving time while working with large classes, and their personal interest in improving student experience. The key points they conveyed in terms of the implementation included delivering an authentic experience and explaining its relevance to students, being involved in an iterative design process to sharpen the design of the intervention and the tool, and incentivizing students in different forms to make them be involved as much as possible.

The instructors thought that they achieved the following outcomes as a result of this writing intervention: they were able to provide direct feedback to students, improve students' self-assessment and engagement with their writing, bring about a general improvement in their writing skills, develop their own knowledge, save time and help with research on writing. This showed that most of their expected outcomes at the beginning of this process of co-design which motivated them were met by the time the intervention was designed and implemented. They identified challenges like additional time and effort involved, involvement of tutors contributing to the way the intervention was implemented, teaching students the appropriate use of feedback to work with an imperfect tool, lack of support from faculty, and the intervention disrupting discipline-related teaching. The instructors further discussed the support mechanisms which helped them including: collaborative effort and responsiveness of the research team, agency in how they designed the intervention for their students, and gave ideas on how other academics can be reached and supported for wider adoption of the tool and the intervention.

Furthermore, analysis of tutor feedback did not show any surprising results in terms of how the tutors felt while delivering the AcaWriter intervention by facilitating it in class. Among the tutors who responded to the feedback survey, most found the process easy to understand and follow, and no new limitations were observed in their experiences. Given that the time of these tutors is scare as they are involved in teaching and other workload like marking, it is significant to note that the preparation for

AcaWriter intervention only required less than an hour of their time. The tutors have also reported that students reflected more on their writing, and the intervention added value to their teaching practices. These factors can positively influence tutors and get their buy-in for the implementation of such learning analytics interventions in their classrooms.

Chapter 7: Discussion, Future work, and Conclusion

The overall aim of this research (first introduced in Section 3.1) was to investigate the implementation, and impact on student writing, of an automated feedback tool in higher education teaching practice. This was tackled by exploring three main research questions using the design iterations detailed in Chapters 4-6 of the thesis. This chapter synthesises and discusses the thesis contributions, provides directions for future work, and concludes the thesis[17]. Section 7.1 summarises the findings for the research questions by consolidating results from the study. A summary of the key contributions is provided in Section 7.2 with its implications, followed by limitations in Section 7.3. Section 7.4 discusses how future work can extend the study for additional insights. Finally, the thesis concludes in Section 7.5.

7.1 Summary of findings for the research questions

The thesis investigated three main research questions discussed in Section 3.1 in the design iterations of its study. This section will consolidate findings from the analysis of data from those iterations to answer the research questions. The detailed discussions on these findings are in their respective sections within Chapter 4:, Chapter 5:, and Chapter 6:. The highlights of the findings for all the research questions are mapped to the design iterations in Appendix E.

7.1.1 RQ1 - Writing Products

RQ1 explored writing products and student perceptions, and the impact of automated feedback on them using the following questions:

RQ1 - Writing Products: What is the impact of automated feedback on rhetorical moves in student writing?

RQ1a. What are students' perceptions of the writing task with/ without automated feedback?

[17] Sections of this chapter draw on the following articles: (Knight, Shibani, et al., 2018; Shibani, Knight, et al., 2018a, 2019)

RQ1b. What is the impact of automated feedback on student revisions?

RQ1a investigated student perceptions of the writing task with and without automated feedback in the design iterations over time. The perceived usefulness score rated by students on the writing intervention was studied by comparing the control groups with/ without AcaWriter feedback quantitatively, and their open-ended responses were qualitatively analysed. In design iteration 2 (the first iteration of this study), the mean perceived usefulness score of the automated feedback group (M = 2.8, SD = 0.75) was significantly lower than the other two groups receiving no feedback and instructor feedback ($F(2,198) = 8.32$, $p < .001$). Student responses uncovered useful task and design elements for the intervention for improvements, including student expectation of more actionable automated feedback on their writing. In the next iteration which was a pilot for iteration 3, a peer discussion module was trialled to complement automated feedback and improve sense making. Here, the automated feedback group found the activity to be more useful (M = 3.6, SD = 0.95) than the No feedback group, but no significant differences were found statistically between the two groups: $t(121) = 1.63$, $p = .11$. Student and instructor feedback from the previous iterations fed into the next iteration for Law (Iteration 3), which introduced a contextualized LA design with automated feedback from AcaWriter tuned for the context using CLAD elements. This iteration saw the AcaWriter feedback group rate the usefulness of the intervention (M = 3.67, SD = 0.79) significantly higher than the No feedback group with $t(83) = 3.9$, $p < .001$ and a large effect size (Cohen's d = .82, 95% CI [0.38, 1.25]), the highest achieved so far in the design iterations. The stable design that was developed over a few semesters in Law was transferred to a new Accounting context in the next design iteration (Iteration 4) illustrating the generalisability of CLAD. The online writing intervention in this iteration was also rated to be useful to improve writing (M = 3.8, SD = 0.9), with 66% of students selecting 4 or 5 (highest) in relation to the perceived usefulness score. The DBR approach resulted in a stabilised learning design that can be appropriately employed by Learning Analytics applications like AcaWriter to improve pedagogical practice in authentic classrooms. In terms of perceived usefulness, the results showed a higher impact of AcaWriter feedback than no automated feedback in the last two writing interventions, which supports the use of tools like AcaWriter to provide formative feedback to students on their writing.

The second sub question **RQ1b** studied the impact of automated feedback on the actual revisions made by students in their writing. In design iteration 2, several methods were explored to study revisions to the provided text using automated and manual means, including: assessment using scores, calculation of metrics like word counts, text dissimilarity cosine distance, and graphical representations like n-gram graphs and rhetorical moves graph. Although no significant differences were found between the scores of the different feedback groups, students who received AcaWriter feedback scored higher (M = 8.87, SD = 2.52) than the other groups, and were less likely to degrade the given text than other groups, demonstrating that the use of feedback directed them to useful revisions. A similar effect was observed in iteration 3, where no significant difference between groups was found in the scores of the revised essay even though AcaWriter feedback group scored higher (M= 7.98, SD = 2.49) than the No feedback group. When the marks were banded by performance, the automated feedback group was found to have produced significantly more improved essays than the other group. In design iteration 4 where student performance on their own assignments were studied (written communication scored out of 30), students who engaged with AcaWriter feedback deeply by understanding and applying it on their writing scored significantly more (M = 23.33, SD = 3.29) than students who demonstrated shallow engagement: t (101) = 3.21, p = 0.002, with a medium observed effect (Cohen's d = -0.62, 95% CI [0.22, 1.02]). This shows a positive impact of AcaWriter feedback on student revisions in their writing if they engaged with the feedback deeply.

7.1.2 RQ2 - Writing Processes

The second research question **RQ2** aimed to study the writing processes by exploring how students engage with automated feedback using the questions below:

RQ2 - Writing Processes: How do students engage with automated writing feedback?

RQ2a. How can we study students' interactions with automated feedback?

RQ2b. How does scaffolding (using peer feedback and additional instruction) impact student engagement with automated feedback?

RQ2a investigated how we can study student interactions with automated feedback in design iteration 3, and introduced a new technique for doing so in Section 4.4.5.4 by tracking student revisions and the revision process using a revision graph. A revision graph was built using text similarity calculation and graph visualization

techniques by comparing the provided text and the changes made by the student at sentence level. Some useful metrics from the revision graph like the number of sentences unchanged in the revised text, and the number of sentences with the presence of rhetorical moves were indicative of the quality of students' revised texts. Based on automated feedback requests, most students had a moderate number of interactions with feedback, and in general their amount of interaction did not have a significant effect on the score of the revised essay. Multi-stage revision graphs were also constructed to study the internal stages/ writing processes in student interactions and their subsequent revisions. Students classified into high, low and moderate interaction categories demonstrated varied behaviours while engaging with automated feedback and revising their texts leading to no finite conclusions with respect to their scores. While this novel technique of studying students' interaction with automated feedback has provided an initial contribution to the body of knowledge, a detailed analysis in this direction could lead to more insights in the future.

In addition, **RQ2b** investigated how additional scaffolding can impact students' engagement with automated feedback from AcaWriter. Design iteration 3 explored scaffolding to automated feedback in the form of peer discussion and feedback, and iteration 4 explored prompts for self-evaluating automated feedback for reflective use of the feedback. Students engaged in discussion with peers and providing feedback to complement the automated feedback from AcaWriter in iteration 3 on the given sample writing. From transcripts of this discussion, it was observed that peer discussion facilitated students' engagement to some extent, helping students to identify ways to improve their writing, suggest improvements, share their interactions, and get clarifications. Some students further used peer discussion to critique the automated feedback, leading to deeper engagement. However, the limited time allotted for this task and a small sample size meant that these results have to be validated in future research for generalizable findings. In iteration 4, the prompts for self-evaluation made visible students' thoughts in terms of their agreement, understanding and use of AcaWriter feedback on their own writing. Students who engaged with the feedback deeply, as demonstrated in their responses to the prompts, scored significantly higher marks for written communication than students who demonstrated shallow engagement. Hence, for the automated feedback from AcaWriter to be effective to all students to improve their performance, it requires developing the feedback literacy of students to make

sense of and critically engage with the automated feedback. This literacy and its substituent components can be explored in more detail in future work.

7.1.3 RQ3 - Educator Perspectives

RQ3 explored the perspectives of educators, who were the practitioners aiding implementation of AcaWriter feedback in authentic writing practice, using semi-structured interviews based on the following questions:

RQ3 - Educator Perspectives: What are practitioner perspectives on automated writing feedback in authentic practice?

RQ3a. What factors influence adoption in authentic classrooms?

RQ3b. How do the practitioners engage in implementation across disciplines, and what are the outcomes?

RQ3a studied the factors influencing learning analytics adoption (automated writing feedback in this case) in classrooms including their motivations, challenges, and support required. Improving students' written communication and self-assessment, ability to provide direct feedback for students, saving time while working with large classes, instructor knowledge of writing and motivation to support it, and openness to the role of technology motivated them to use the learning analytics application AcaWriter as part of their teaching. The instructors identified challenges including resource costs, pressures on disciplinary-research outputs, and the need to consider wider teaching teams in implementing learning analytics. While they noted the tool was imperfect, they also highlighted its potential particularly in interventions such as this one in which students are engaged in appropriate use of feedback. They noted that the collaborative nature of the project and the agency provided to them in designing for their students supported successful implementation. To support adoption of AcaWriter, they recommended raising awareness about the use of learning analytics tools in the classroom, and the support available to academics for implementation.

RQ3b studied how the interventions were implemented, and the outcomes they achieved. The instructors traced back the design process by explaining the steps they undertook while co-designing the intervention with researchers, and the obstacles on the way. They explained measures like maintaining authenticity of the experience, making it relevant to students and incentivising students that helped them in this implementation. In terms of the outcomes, the instructors reflected that the

interventions had: enabled them to provide direct feedback to students, improved students' self-assessment and engagement with their writing, brought about a general improvement in their writing skills, developing their own knowledge and – over the long term – saving time. Their expected outcomes for the process of co-design which motivated them were met by the time the intervention was designed and implemented. The tutors who facilitated some of these interventions in class also said that it added value to their teaching and were interested in using the intervention and AcaWriter again in future semesters.

7.2 Overview of Contributions and Implications

The thesis has achieved its core aim of investigating the implementation, and impact on student writing, of an automated feedback tool in higher education teaching practice. This was done by co-designing, implementing, and evaluating the sociotechnical infrastructure for the AcaWriter automated writing feedback tool, an approach designed to be adaptable to different educational contexts and disciplines. The main contributions of the thesis cut across the fields of learning analytics, writing research, and adoption of technology in classrooms.

As discussed in the literature review in Section 2.2, Writing is a core skill for higher education students and needs better support in practice using formative feedback and strategy instruction (Graham & Perin, 2007). The thesis introduced an approach that led to significant improvements in student writing with the use of automated feedback from a learning analytics tool called AcaWriter and a pedagogical design embedding its usage. The design based research (DBR) methodology applied in the thesis conducted trials in multiple cycles of iterative design. DBR contributes to the creation of new theoretical principles that are intertwined with practice rather than basing the implementation on an existing theory from the start (Hein, 2017), and led to a stable design and set of design principles. The approach measured impact on student writing in authentic practice using AcaWriter, validating its applicability to support new writing contexts.

Thus, the potential of using learning analytics for writing impact has been evidenced by this research by adopting a perspective of *'augmentation'* - "to investigate where existing good practice might be augmented by learning analytics, further strengthening that practice" (Knight, Shibani, et al., 2018). We postulate that this

approach increases adoption both of the analytics, and of the underlying practices in higher education. It addresses the general concern in technology enhanced learning by bridging the gap between expected use and the actual use of the innovations for large scale adoption (Ferguson et al., 2014; Wingate, 2012), A similar call has been made for learning analytics to capture their pedagogical intents by aligning the analytics with the learning design – as reviewed in Section 2.7.3 (Gašević et al., 2016; Lockyer et al., 2013; Mangaroska & Giannakos, 2018; Nguyen, Rienties, & Toetenel, 2017; Wise, 2014). The thesis has formalised the key concepts that emerged from the design-based research cycles, contributing a conceptual model for *Contextualizable Learning Analytics Design (CLAD)*. The model brings together elements of learning analytics (features and feedback), with learning design elements (assessment and task design) to flexibly design LA that is relevant for the context and can be scaled to new settings. It highlights the possible dynamics that can arise using the cogs metaphor since these gears can be synchronized or desynchronized, and engaged or disengaged. This model and approach can be generalized to other tools and settings for contextualizing LA applications.

The adaptation of LA technologies provides technological affordances like automated feedback, that can be contextualized to provide relevant feedback, and the implementation of the intervention informs LA research and the learning design. It used a co-design approach where the educators worked hand-in-hand with researchers across two disciplines: Law and Accounting, to effectively implement automated writing feedback. The co-design process brings context to the learning analytics application by valuing the views of the stakeholders involved and providing agency to the instructor in its design and implementation (Dollinger & Lodge, 2018; Prieto-Alvarez et al., 2018). A lack of agency leads to lack of trust in the LA application, hindering adoption, and hence, it is a top priority for LA to provide a genuine voice to educators (Buckingham Shum et al., 2019). The thesis demonstrates the value of multi-stakeholder perspectives, documenting the experiences of the students who used the tool (Chapters 4-5), and those of educators who were involved in the co-design and implementation of the LA application in an iterative fashion (Chapter 6:). The analysis of educators' experiences uncovered practical considerations, issues and motivations to implement and operationalise learning analytics research in authentic scenarios. It encourages developers of LA tools to build flexible systems that instructors can customise elements of across contexts, even if their source code is not open-access. Providing a level of control for educators also ensures transparency around the systems by giving them the

opportunity to understand and define appropriate feedback for their students, so that the models do not act as black boxes for human interpretation. There is increasing attention on the importance of ethics in machine learning and human AI interaction (Google, 2018); ensuring systems can be adapted to contexts of use is one approach to addressing some of these concerns. Increasing malleability, and reducing opacity, have implications for the architecture of software systems, the skillsets of analytics team, and how they are introduced to learning contexts.

The thesis contributes further to writing and writing analytics research through new methods and results that were obtained by evaluating the automated feedback from AcaWriter. The work is distinctive in comparison to most writing assessment research in that it combined all of the following: a) used in an authentic context, with an authentic task design, b) measured actual outcomes in writing as well as perceptions, and c) developed new methods to study writing. It introduced a novel automated method using text and graph analysis called 'Revision Graphs' prototyped in Python, enabling researchers to study the process of revision and student interactions with automated feedback (see Sections 4.2.5.5 and 4.4.5.4). Learning analytics methods to collect and analyse such process data help researchers to study writing processes in great detail. It can inform them of the drafting processes and behaviours of students during writing, which can lead to the development of tools that provide feedback on making effective revisions in the draft; the web-based ArgRewrite tool (Zhang et al., 2016) provides one example, classifying revisions as either 'surface' or 'content' based changes in students' drafts. This can provide a lens to study the writing style of different authors and find synergies among certain set of writers that attribute to better quality.

In investigating the use of automated feedback on rhetorical moves, the thesis provided empirical evidence that automated feedback is useful to improve student writing in authentic classroom settings. This builds on prior work in writing analytics that make use of automated tools to provide formative feedback on student writing (Allen et al., 2015; Cotos & Pendar, 2016; Roscoe et al., 2015); see detailed review in Section 2.4.2.2. Additionally, it has provided preliminary evidence to demonstrate relationships between students' engagement with the feedback and their final marks scored in subject assignment, which contribute to future research in writing analytics. For automated feedback on writing to be effective to all students to improve their performance, it requires development of their feedback literacy in order to make sense

of and critically engage with the automated feedback, which further contributes to improvements in their evaluative judgement (Boud et al., 2018)

Furthermore, the thesis contributes to the field's understanding of how theory and practice cross-fertilise, not only in principle, but exemplified with practical methods that achieved intended outcomes for students and practitioners. The thesis demonstrated how to transfer the use of the tool from one disciplinary context (Law) to a second (Accounting), by adjusting the 'gears' in the CLAD framework. Design abstractions like design patterns (Goodyear, 2005) and conjecture maps (Sandoval, 2014) were introduced to represent the theory and rationale behind the tasks, resources, and tools, contributing to the socio-technical infrastructure of the automated feedback tool. There are no concepts or assumptions underpinning this work that are specific to writing, or writing analytics, so we argue that the insights and lessons learned from this research can contribute more widely to other learning analytics applications, to support their adoption in authentic practice.

7.3 Limitations

All methodologies bring their own limitations, so while the selection of Design Based Research in this thesis (Section 3.2) was justified in order to conduct studies in authentic classroom settings, this traded off the ability to maintain tight control over many variables. On the one hand, the tasks, procedures, and materials supporting the writing intervention were designed to be consistent across student conditions receiving/ not receiving automated feedback, controlling for internal differences and hence validate the differences. On the other hand, real classrooms are noisy environments where experiences and consequences other than the ones we design can also occur, which can sometimes go uncaptured by the analytics and tools we use. Examples in this work included are unrecorded interactions between two students who sat next to each other during an online task, and emotions triggered by the way a particular tutor facilitated discussion in class. These external factors were not taken into account in the current study since the implementations were not controlled experiments conducted in a lab setting. Therefore, the findings should be carefully considered with regard to the context before deriving conclusive evidence from the study. Moreover, results from relatively small sample sizes from parts of the study investigating peer feedback in Section 4.4.7 and small effect sizes while evaluating the interventions in design iterations of Sections 4.2.5 and 4.4.5 should be noted before concurring/ dissenting with

them in a different context. The claims from the findings should hence be considered in context of the current study and be used with caution over generalizable tools.

Another issue is that the study did not consider demographic and personal information of students as it was not the focus of the study, and hence did not group students using such data. Data was collected from students only from the writing interventions and related marks. However, it is important to note that variables like the language proficiency of students (Native Speaker vs English as a Second Language Speaker) might play a role in how the user groups interpret the feedback and use it on their writing. Earlier studies have shown differences in the writing fluency and feedback ability comparing L1 and L2 students (Chenoweth & Hayes, 2001; Yu & Lee, 2014), so this is a factor to be considered. Hence, the results may not generalise to other populations that are very different to the current set of users. From an ethical standpoint, the requirement to cater for authentic student experience also meant that students had to be provided equal opportunities for learning so that one group was not privileged over the other. So, as earlier discussed in Section 3.7.1, the interventions were designed in a way that they provided meaningful learning experiences for all students, with or without the provision of additional support from AcaWriter. This precluded the use of between-subjects feedback conditions in a given cohort, such as providing no writing support for one group.

In addition, the results are limited to the study of automated feedback on *rhetorical moves* from a specific tool. Results might differ when automated writing feedback based on attributes like grammar, content topic, etc. are investigated using other tools. Due attention should be provided when translating the approach and results to other writing analytics tools.

The contextualization of learning analytics using the proposed approach also has implications for resource use; co-design places different resource constraints on development compared to other approaches which prioritise technical improvements in tools. We have argued more fully elsewhere that building the sociotechnical infrastructure around the tool and tasks leads to better learning impact than just improving the technology (Knight, Gibson, & Shibani, In Submission). It should be noted that although the initial set up required for such contextualizable learning analytics designs might be slow, it works over time for relevant application of LA in many different learning contexts. In our example, the cost involved in contextualizing the

learning analytics design in the first Law context was high due to the effort required in the co-design process with the instructor. However, when the design was stabilized with resources, it was far easier to transfer to the new Accounting context. A third context also exists, in which AcaWriter was tuned to cater to *research* student writing, for the typical abstracts summarising project reports and academic publications (Abel et al., 2018). This context-sensitivity ensures that LA contributes to the learning in authentic practice for pragmatic real-life usage, in spite of the additional costs involved.

7.4 Future Work

This thesis has investigated student engagement with automated feedback from AcaWriter and its effect on writing performance over multiple sessions and cohorts, indicating – particularly with later, more developed designs – quantifiable impact on learning, and interaction with the feedback. This now opens up a range of interesting research questions. Further work should be conducted to investigate the *longitudinal* impact of automated feedback on performance. Such longitudinal research would illuminate the extent to which students learn strategies and skills to apply the feedback provided through automated feedback, to the point that those scaffolds are no longer required for their future writing, i.e., the extent to which their actioning of the feedback transfers beyond the study context. In addition, further work should investigate transfer over forms or genres of writing. Finally, each of the different scaffolding mechanisms (e.g. automated writing feedback, peer feedback, exemplars) could be explored in more detail in a stepwise process, rather than as a whole intervention, to study their individual effects.

Receiving feedback from a machine is not the same as receiving feedback from a human, but in exactly what ways is this experienced as different by students, and how do they respond? Building on the concept of *feedback literacy* (Carless & Boud, 2018; Sutton, 2012), discussed in Section 5.2.5, the concept of 'automated feedback literacy' is emerging, as a way to understand how students comprehend, make sense of, critically engage with, and are able to apply automated feedback effectively. This should be explored in more detail, for instance, by identifying further levels of student engagement with the feedback, by expanding the current classification scheme beyond 'shallow vs. deep'. Future work could include think-aloud strategies to capture students' thinking as they use the tool. By understanding students' feedback literacy, and the

types of feedback that best help students, we can design personalized support for students that adapts to their needs based on individual differences.

On the technical front, as a tool to study students' interaction with the automated feedback and how they use it for revising their writing, the Revision Graphs could be refined, and developed for user testing as a form of feedback for different stakeholder groups. The Revision Graphs are currently built for the purpose of aiding *researchers* to identify revision patterns across students and their interactions with automated feedback, and requires technical expertise to create and understand its graphical interface. Future work can extend the functionality of Revision Graphs by improving their readability, fixing the technical issues by creating a more user-friendly interface or API to work with, and evaluating the visualisations with *instructors* (and perhaps even students) for their use as a form of feedback to support learning. Revision patterns, when combined with other automated feedback literacy measures, could help explain how students interact and apply automated feedback.

Finally, future work could also explore how the tool and the approach used in this research could be adopted in other teaching and learning contexts, disciplines, and universities by researchers and practitioners. This can build on the existing socio-technical infrastructure by sharing additional resources and technical advances as a research community, contributing to Design Based Implementation Research (DBIR), earlier introduced in Section 3.2. The approach used in the thesis has thus formed the first step towards building such scalable learning analytics that are also contextualizable to encourage adoption. It can lead to sustainable impact through mainstreamed services in a university setting (Buckingham Shum & McKay, 2018), and future work could utilise DBIR processes to examine strategies for adoption and impact (Penuel et al., 2011). Work around contextualization may find further levels and elements for contextualizing LA in other authentic learning scenarios. Applicability of the Contextualizable Learning Analytics Design (CLAD) model to other types of learning analytics (e.g. learner dashboards) could test its generalisability. It could potentially transfer from the current case of student facing LA tools to other LA applications to cater to relevant contexts as well.

7.5 Conclusion

The thesis has presented pioneering work in the field of Learning Analytics (LA) by co-designing, implementing, and evaluating the sociotechnical infrastructure for an automated writing feedback tool for students to improve their writing, that can be adapted to different educational contexts. The approach led to significant improvements in students' writing using an automated feedback tool called AcaWriter, validating its applicability in higher education writing contexts. From a series of design-based research iterations across student cohorts and disciplinary contexts, it has concluded that LA should be contextualized for learning contexts, and proposes a Contextualizable Learning Analytics Design (CLAD) model to do so at scale, by aligning LA with Learning Design (LD). Subject matter experts and instructors who teach the subject to students played a major role in co-designing elements of the contextualization, so that the LA augmented existing good practices in the LD.

The results of the empirical studies showed that students found the writing intervention and automated feedback from AcaWriter useful to improve their subject-related writing skills, with evidence of impact of different types, from different data sources. By examining students' interaction patterns and their engagement, the thesis uncovered varied levels of *automated feedback literacy* among students when using the tool, a concept that lays the foundation for future research. The contextualization of different elements, particularly tuning the writing analytics tool to fit the context and the careful design of interventions, had an impact on student perceptions and uptake of the application. A design implication is that these should be designed to be configurable: we need systems that are malleable to the needs of the educator, considering the learning goals, and ensuring that such technologies are introduced to students with relevant support to help them engage effectively.

These insights reflected the perspectives of educators as well, who appreciated the agency provided to them to co-design learning innovations relevant for their classrooms. This method of involving stakeholders in the design process provides them the ability to orchestrate the innovations, and would improve adoption of learning analytics. If educators are not enthusiastic champions of learning analytics, all the technical innovations will have very limited impact in universities. Hence, this work has foregrounded educators' experiences of, and reflections on an analytics tool, to better understand the factors that accelerate or obstruct adoption, and effective use. The

insights from educators have helped us to bridge the gap between theory and practice by implementing effective learning analytics in authentic practice.

References

Abel, S., Kitto, K., Knight, S., & Buckingham Shum, S. (2018). *Designing personalised, automated feedback to develop students' research writing skills* Paper presented at the ASCILITE 2018, Geelong.

Aït-Mokhtar, S., Chanod, J.-P., & Roux, C. (2002). Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering, 8*(2-3), 121-144.

Alhadad, S. S., & Thompson, K. (2017). Understanding the mediating role of teacher inquiry when connecting learning analytics with design for learning. *Interaction, Design, & Architecture (s), 33*, 54-74.

Allal, L., Lopez, L. M., Lehraus, K., & Forget, A. (2005). Whole-class and peer interaction in an activity of writing and revision *Writing in context (s)* (pp. 69-91): Springer.

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Computer-based writing instruction. *Handbook of writing research*, 316-329.

Anders I. Mørch, I. E., Victor C. Cheng, William K. Cheung, Kelvin C. Wong. (2017). EssayCritic: Writing to Learn with a Knowledge-Based Design Critiquing System. *Educational Technology & Society, 20*(2), 213–223.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational researcher, 41*(1), 16-25.

Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *National Council on Measurement in Education (NCME), Educational Testing Service, Princeton, NJ.*

Attali, Y. (2013). 11 Validity and Reliability of Automated Essay Scoring. *Handbook of automated essay evaluation: Current applications and new directions*, 181.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., . . . Lockyer, L. (2016). *A conceptual framework linking learning design with learning analytics.* Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge.

Ball, E. (2009). A participatory action research study on handwritten annotation feedback and its impact on staff and students. *Systemic practice and action research, 22*(2), 111.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The journal of the learning sciences, 13*(1), 1-14.

Bektik, D. (2017). *Learning analytics for academic writing through automatic identification of meta-discourse.* The Open University.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education, 32*(3), 347-364.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of second language writing, 14*(3), 191-205.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*(1), 5.

Bloom, B. S. (1971). Handbook on formative and summative evaluation of student learning.

Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (2018). *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*. Milton, United Kingdom: Routledge.

Boud, D., Cohen, R., & Sampson, J. (2014). *Peer learning in higher education: Learning from and with each other*: Routledge.

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.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences, 2*(2), 141-178.

Buckingham Shum, S. (2018). Open Source Writing Analytics. Retrieved 12/09/2018, from https://utscic.edu.au/open-source-writing-analytics/

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). (Eds.) Human-Centred Learning Analytics: Design Frameworks, Tools and Techniques. *Journal of Learning Analytics (Special Issue)*.

Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). *Critical perspectives on writing analytics*. Paper presented at the Workshop at the Sixth International Conference on Learning Analytics & Knowledge.

Buckingham Shum, S., & McKay, T. (2018). Architecting for learning analytics. Innovating for sustainable impact. *EDUCAUSE review, 53*(2), 25-37.

Burstein, J., Chodorow, M., & Leacock, C. (2003). *CriterionSM Online Essay Evaluation: An Application for Automated Evaluation of Student Essays.* Paper presented at the IAAI.

Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2011). Collaborative writing support tools on the cloud. *IEEE Transactions on Learning Technologies, 4*(1), 88-97.

Caporossi, G., & Leblay, C. (2011). *Online writing data representation: a graph theory approach.* Paper presented at the International Symposium on Intelligent Data Analysis.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of second language writing, 12*(3), 267-296.

Chen, B., & Zhu, H. (2019). *Towards Value-Sensitive Learning Analytics Design.* Paper presented at the Proceedings of the 9th International Conference on Learning Analytics & Knowledge.

Chen, C.-F. E., & Cheng, W.-Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94-112.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written communication, 18*(1), 80-98.

Cheville, J. (2004). Automated scoring technologies and the rising influence of error. *The English Journal, 93*(4), 47-52.

Collins, A. (1992). Toward a design science of education *New directions in educational technology* (pp. 15-22): Springer.

Conijn, R., Loo, J. v. d., & Zaanen, M. v. (2018). *What's (not) in a Keystroke? Automatic Discovery of Students' Writing Processes Using Keystroke Logging*. Paper presented

at the 8th International Conference on Learning Analytics & Knowledge (LAK18).
Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of second language writing, 3*(3), 257-276.
Conole, G. (2008). New schemas for mapping pedagogies and technologies. *Ariadne*(56), 2.
Corrin, L., Kennedy, G., de Barba, P., Lockyer, L., Gaševic, D., Williams, D., & Bakharia, A. (2016). Completing the loop: Returning meaningful learning analytic data to teachers. *Sydney: Australian Office for Learning and Teaching*.
Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kirsty, K., Pardo, A., . . . Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper.
Cotos, E., & Pendar, N. (2016). Discourse classification into rhetorical functions for AWE feedback. *calico journal, 33*(1), 92.
Crawford, L., Lloyd, S., & Knoth, K. (2008). Analysis of student revisions on a state writing test. *Assessment for Effective Intervention, 33*(2), 108-119.
Crossley, S., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). *Pssst... textual features... there is more to automatic essay scoring than just you!* Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.
Crossley, S. A., Allen, L. K., & McNamara, D. S. (2014). A Multi-Dimensional analysis of essay writing *Multi-dimensional analysis* (Vol. 25, pp. 197-237): John Benjamin's Publishing Company.
Curry, M. J., & Hewings, A. (2003). Approaches to teaching writing. In C. Coffin, M. J. Curry, S. Goodman, A. Hewings, T. M. Lillis & J. Swann (Eds.), *Teaching Academic Writing: A Toolkit for Higher Education* (pp. 19–44). London, UK: Routledge.
Davis, J., & Nathan, L. P. (2015). Value sensitive design: Applications, adaptations, and critiques. *Handbook of ethics, values, and technological design: Sources, theory, values and application domains*, 11-40.
Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: staff and student perspectives. *Assessment & Evaluation in Higher Education, 44*(1), 25-36. doi: 10.1080/02602938.2018.1467877
Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). *Increasing the Impact of Learning Analytics*. Paper presented at the Ninth International Conference of Learning Analytics and Knowledge (LAK19), Tempe, Arizona.
Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*(1), 7-24.
Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for computer-supported collaborative learning. *Journal of computer assisted learning, 23*(1), 1-13.
Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in higher education, 24*(3), 331-350.
Dollinger, M., & Lodge, J. M. (2018). *Co-creation strategies for learning analytics.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.
Echeverria, V., Martinez-Maldonado, R., Shum, S. B., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus Explanatory Visual Learning Analytics: Driving Teachers' Attention through Educational Data Storytelling. *Journal of Learning Analytics, 5*(3), 72—97-72—97.

Erickson, F. (1985). *Qualitative methods in research on teaching*: Citeseer.
Ericsson, P. F., & Haswell, R. H. (2006). Machine scoring of student essays: Truth and consequences.
Feltrim, V., Teufel, S., Nunes, M., & Aluísio, S. (2006). Argumentative zoning applied to critiquing novices' scientific abstracts. *Computing Attitude and Affect in Text: Theory and Applications*, 233-246.
Feng, H.-H. (2015). *Designing, implementing, and evaluating an automated writing evaluation tool for improving EFL graduate students' abstract writing: a case in Taiwan.* (Graduate Theses and Dissertations), Iowa State University.
Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014). *Setting learning analytics in context: Overcoming the barriers to large-scale adoption.* Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge.
Fitzgerald, J. (1987). Research on revision in writing. *Review of educational research, 57*(4), 481-506.
Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication, 32*(4), 365-387.
Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68-84.
Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-71.
Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). Reflective writing analytics for actionable feedback.
Gibson, A., & Lang, C. (2018). *The pragmatic maxim as learning analytics research method.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.
Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304-315.
Goodyear, P. (2005). Educational design and networked learning: Patterns, pattern languages and design practice. *Australasian Journal of Educational Technology, 21*(1).
Goodyear, P., & Carvalho, L. (2014). Framing the analysis of learning network architectures *The architecture of productive learning networks* (pp. 66-88): Routledge.
Google. (2018). What If...you could inspect a machine learning model, with no coding required? , from https://pair-code.github.io/what-if-tool/
Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers, 36*(2), 193-202.
Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and writing, 26*(1), 1-15.
Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445.
Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*(3), 42–57.
Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365-379.

Hecking, T., & Hoppe, H. U. (2015). *A Network Based Approach for the Visualization and Analysis of Collaboratively Edited Texts.* Paper presented at the VISLA@ LAK.

Hein, E. (2017). Design-Based Research. Retrieved June 6, 2019, from http://www.ethanhein.com/wp/2017/design-based-research/

Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education, 37*(2), 149-161.

Holstein, K., McLaren, B. M., & Aleven, V. (2017). *Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms.* Paper presented at the Proceedings of the seventh international learning analytics & knowledge conference.

Horstmanshof, L., & Brownie, S. (2013). A scaffolded approach to discussion board use for formative assessment of academic writing skills. *Assessment & Evaluation in Higher Education, 38*(1), 61-73.

Huang, A. (2008). *Similarity measures for text document clustering*. Paper presented at the Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand.

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: a Meta-Analysis. *Assessment & Evaluation in Higher Education, 44*(6), 863-880.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of second language writing, 7*(3), 255-286.

Hyland, K. (2005). *Metadiscourse*: Wiley Online Library.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly, 24*(4), 602-611.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). *License to evaluate: Preparing learning analytics dashboards for educational practice.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2009). *Handbook on impact evaluation: quantitative methods and practices*. Washington, D.C: The World Bank.

Kitto, K., Buckingham Shum, S., & Gibson, A. (2018). *Embracing imperfection in learning analytics.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.

Kitto, K., Lupton, M., Davis, K., & Waters, Z. (2017). Designing for student-facing learning analytics. *Australasian Journal of Educational Technology, 33*(5), 152-168.

Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019). Technological barriers and incentives to learning analytics adoption in higher education: insights from users. *Journal of Computing in Higher Education*, 1-22.

Knight, S., Allen, L., Gibson, A., McNamara, D., & Buckingham Shum, S. (2017). *Writing analytics literacy: bridging from research to practice.* Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference.

Knight, S., & Anderson, T. D. (2016). *Action-oriented, Accountable, and Inter (Active) Learning Analytics for Learners.* Paper presented at the LAL@ LAK.

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

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. (2018). Designing academic writing analytics for civil law student self-assessment. *International Journal of Artificial Intelligence in Education, 28*(1), 1-28.

Knight, S., Gibson, A., & Shibani, A. (In Submission). Implementing Learning Analytics for Learning Impact: Taking Tools to Task.

Knight, S., Leigh, A., Davila, Y. C., Martin, L. J., & Krix, D. W. (2019). Calibrating assessment literacy through benchmarking tasks. *Assessment & Evaluation in Higher Education*, 1-12.

Knight, S., Martinez-Maldonado, R., Gibson, A., & Buckingham Shum, S. (2017). *Towards mining sequences and dispersion of rhetorical moves in student written texts.* Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference.

Knight, S., Shibani, A., Abel, S., Gibson, A., Ryan, P., Sutton, N., . . . Buckingham Shum, S. (In submission). AcaWriter: A Learning Analytics Tool for Formative Feedback on Academic Writing.

Knight, S., Shibani, A., & Buckingham Shum, S. (2018). *Augmenting Formative Writing Assessment with Learning Analytics: A Design Abstraction Approach.* Paper presented at the 13th International Conference of the Learning Sciences, London, United Kingdom.

Knight, S., & Shum, S. B. (2017). Theory and learning analytics. *The handbook of learning analytics*, 17-22.

Koh, E., Shibani, A., Tan, J. P.-L., & Hong, H. (2016). *A pedagogical framework for learning analytics in collaborative inquiry tasks: An example from a teamwork competency awareness program.* Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge.

Koh, E., Tee, Y.-H., Shibani, A., Tay, K. L., Tan, J. P.-L., Hong, H., . . . Chan, H. L. (2016). *Designing a web tool to support teamwork awareness and reflection: Evaluation of two trial cycles.* Paper presented at the 24th International Conference on Computers in Education, India.

Krause, K.-L. (2001). The university essay writing experience: A pathway for academic integration during transition. *Higher Education Research & Development, 20*(2), 147-168.

Kuh, G. D. (2008). Excerpt from high-impact educational practices: What they are, who has access to them, and why they matter. *Association of American Colleges and Universities.*

Kukich, K. (2000). Beyond automated essay scoring, the debate on automated essay grading. *IEEE intelligent systems, 15*(5), 22-27.

Lea, M. R. (2004). Academic literacies: A pedagogy for course design. *Studies in higher education, 29*(6), 739-756.

Learning, V. (2003). How does IntelliMetric score essay responses. *RB-929). Newtown, PA: Author.*

Leeman-Munk, S. P., Wiebe, E. N., & Lester, J. C. (2014). *Assessing elementary students' science competency with text analytics.* Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge.

Lillis, T. (2003). Student writing as' academic literacies': Drawing on Bakhtin to move from critique to design. *Language and education, 17*(3), 192-207.

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

Liu, D. Y.-T., Bartimote-Aufflick, K., Pardo, A., & Bridgeman, A. J. (2017). Data-Driven Personalization of Student Learning Support in Higher Education *Learning Analytics: Fundaments, Applications, and Trends* (pp. 143-169): Springer.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459.

MacArthur, C. A. (2007). Best practices in teaching evaluation and revision. *Best practices in writing instruction*, 141-162.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment, 9*, 17-28.

Madnani, N., Burstein, J., Elliot, N., Klebanov, B. B., Napolitano, D., Andreyev, S., & Schwartz, M. (2018). *Writing Mentor: Self-Regulated Writing Feedback for Struggling Writers.* Paper presented at the Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations.

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*.

McDonald, J., & Moskal, A. C. M. (2017). Quantext: Analysing student responses to short-answer questions. *Facilitating social learning through learning design: A perspective of collaborative academic development*, 133.

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

Merceron, A., Blikstein, P., & Siemens, G. (2016). Learning analytics: from big data to meaningful data. *Journal of Learning Analytics, 2*(3), 4-8.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality.* Paper presented at the Advances in neural information processing systems.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series, 2003*(1).

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design. *Handbook of test development*, 61-90.

Muncie, J. (2000). Using written teacher feedback in EFL composition classes. *ELT journal, 54*(1), 47-53.

National Commission on Writing. (2003). Report of the The National Commission on Writing in America's Schools and Colleges: The neglected "R": The need for a writing revolution.

National Health Medical Research Council, A. (2007). *National statement on ethical conduct in human research*: National Health and Medical Research Council.

Nelson, B., Ketelhut, D. J., Clarke-Midura, J., Bowman, C., & Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology, 45*(1), 21.

Nguyen, Q., Rienties, B., & Toetenel, L. (2017). *Mixing and matching learning design and learning analytics.* Paper presented at the International Conference on Learning and Collaboration Technologies.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

OECD. (2000). *Literacy in the information age: final report of the International Adult Literacy Survey*: OECD.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society, 17*(4), 49-64.

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of second language writing, 8*(3), 265-289.
Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational researcher, 40*(7), 331-337.
Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. D. (2018). Co-designing learning analytics tools with learners. *Learning Analytics in the Classroom: Translating Learning Analytics for Teachers*.
Prieto, L. P., Dimitriadis, Y., Asensio-Pérez, J. I., & Looi, C.-K. (2015). Orchestration in learning technology research: evaluation of a conceptual framework. *Research in Learning Technology, 23*.
Prieto, L. P., Holenko Dlab, M., Gutiérrez, I., Abdulwahed, M., & Balid, W. (2011). Orchestrating technology enhanced learning: a literature review and a conceptual framework. *International Journal of Technology Enhanced Learning, 3*(6), 583-598.
Prieto, L. P., Rodríguez-Triana, M. J., Martínez-Maldonado, R., Dimitriadis, Y., & Gasevic, D. (Forthcoming). Orchestrating Learning Analytics (OrLA): supporting the adoption of learning analytics at the practitioner level.
Prieto, L. P., Rodríguez Triana, M. J., Martínez Maldonado, R., Dimitriadis, Y. A., & Gašević, D. (2018). Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level.
Ramaprasad, A. (1983). On the definition of feedback. *Systems Research and Behavioral Science, 28*(1), 4-13.
Rapp, C., & Kauf, P. (2018). Scaling academic writing instruction: Evaluation of a scaffolding tool (Thesis Writer). *International Journal of Artificial Intelligence in Education, 28*(4), 590-615.
Reeves, T. (2006). Design research from a technology perspective *Educational design research* (pp. 64-78): Routledge.
Reimann, P. (2016). Connecting learning analytics with learning research: The role of design-based research. *Learning: Research and Practice, 2*(2), 130-142.
Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12-27.
Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology, 105*(4), 1010.
Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: Applications for intelligent feedback in a writing tutor. *Tech., Inst., Cognition and Learning, 10*(1), 59-79.
Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. *Computers in Human Behavior, 70*, 207-221.
Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment, 4*(4).
Ruegg, R. (2015). Differences in the uptake of peer and teacher feedback. *RELC Journal, 46*(2), 131-145.
Rust, C., Price, M., & O'donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*(2), 147-164.
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science, 18*(2), 119-144.

Sándor, Á. (2007). Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue française de linguistique appliquée, 12*(2), 97-108.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences, 23*(1), 18-36.

Sergis, S., & Sampson, D. G. (2017). Teaching and learning analytics to support teacher inquiry: A systematic literature review *Learning analytics: Fundaments, applications, and trends* (pp. 25-63): Springer.

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison: WI: Cathcart Press.

Shermis, M. D. (2015). THE ROLE OF MACHINE SCORING IN SUMMATIVE AND FORMATIVE ASSESSMENT. *The Next Generation of Testing: Common Core Standards, Smarter? Balanced, PARCC, and the Nationwide Testing Movement*, 83.

Shermis, M. D., Raymat, M. V., & Barrera, F. (2003). Assessing Writing through the Curriculum with Automated Essay Scoring. *ERIC document reproduction service no ED 477 929*.

Shibani, A. (In submission). Automated Revision Graphs: A novel visualization method to study written text as graph.

Shibani, A. (2017a). Adding CKEditor to webpages in PHP #Code. Retrieved 20 February, 2019, from http://antonetteshibani.com/code/adding-ckeditor-to-webpages/

Shibani, A. (2017b). *Combining automated and peer feedback for effective learning design in writing practices*. Paper presented at the 25th International Conference on Computers in Education, New Zealand.

Shibani, A. (2018). *AWA-Tutor: A Platform to Ground Automated Writing Feedback in Robust Learning Design*. Paper presented at the 8th International Conference on Learning Analytics & Knowledge (LAK18), Sydney.

Shibani, A., & Abel, S. (2018). Using AcaWriter to help develop student writing. Retrieved 24 March, 2019, from https://futures.uts.edu.au/events/using-acawriter-help-develop-student-writing/

Shibani, A., Abel, S., Gibson, A., & Knight, S. (2018). *Turning the TAP on Writing Analytics*. Paper presented at the Learning Analytics and Knowledge, Sydney.

Shibani, A., Knight, S., & Buckingham Shum, S. (2018a). *Understanding Revisions in Student Writing through Revision Graphs*. Paper presented at the 19th International Conference on Artificial Intelligence in Education, London, UK.

Shibani, A., Knight, S., & Buckingham Shum, S. (2018b). Understanding Students' Revisions in Writing: From Word Counts to the Revision Graph. Australia: Connected Intelligence Centre, University of Technology Sydney.

Shibani, A., Knight, S., & Buckingham Shum, S. (2019). *Contextualizable Learning Analytics Design: A Generic Model, and Writing Analytics Evaluations*. Paper presented at the The 9th International Conference on Learning Analytics and Knowledge (LAK'19), Tempe, Arizona.

Shibani, A., Knight, S., & Buckingham Shum, S. (In Submission). Educator Perspectives on Learning Analytics in Classroom Practice.

Shibani, A., Knight, S., Buckingham Shum, S., & Ryan, P. (2017). *Design and Implementation of a Pedagogic Intervention Using Writing Analytics*. Paper presented at the 25th International Conference on Computers in Education, New Zealand.

Shibani, A., Liu, M., Rapp, C., & Knight, S. (2019). *Advances in Writing Analytics: Mapping the state of the field*. Paper presented at the Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19), Tempe, Arizona.

Siemens, G. (2012). Sensemaking: Beyond analytics as a technical activity. *Presentation at the EDUCAUSE ELI.*

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

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review, 46*(5), 30.

Simsek, D., Buckingham Shum, S., Sandor, A., De Liddo, A., & Ferguson, R. (2013). XIP Dashboard: visual analytics from automated rhetorical parsing of scientific metadiscourse.

Simsek, D., Sándor, Á., Buckingham Shum, S., Ferguson, R., De Liddo, A., & Whitelock, D. (2015). *Correlations between automated rhetorical analysis and tutors' grades on student essays.* Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510-1529.

Snow, E. L., Allen, L. K., Jacovina, M. E., Crossley, S. A., Perret, C. A., & McNamara, D. S. (2016). Keys to detecting writing flexibility over time: entropy and natural language processing. *Journal of Learning Analytics, 2*(3), 40-54.

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). *You've got style: detecting writing flexibility across time.* Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). *Analysis of collaborative writing processes using revision maps and probabilistic topic models.* Paper presented at the Proceedings of the third international conference on learning analytics and knowledge.

Suthers, D., & Verbert, K. (2013). *Learning analytics as a middle space.* Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge.

Sutton, P. (2012). Conceptualizing feedback literacy: knowing, being, and acting. *Innovations in Education and Teaching International, 49*(1), 31-40.

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language learning, 62*(4), 1110-1133.

Swales, J. (2004). *Research genres: Explorations and applications*: Ernst Klett Sprachen.

Tan, J. P.-L., & Koh, E. (2017). Situating learning analytics pedagogically: Towards an ecological lens: Taylor & Francis.

Testing, T. G. o. A. a. (1987). National Curriculum: Task Group on Assessment and Testing. A Report. London: ERIC Clearinghouse.

Teufel, S. (2000). *Argumentative zoning: Information extraction from scientific text.* University of Edinburgh.

Thompson, K., Alhadad, S. S., Buckingham Shum, S., Howard, S., Knight, S., Martinez-Maldonado, R., & Pardo, A. (2018). Connecting expert knowledge in the design of classroom learning experiences *From data and analytics to the classroom: Translating learning analytics for teachers.*

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative Peer Assessment of Academic Writing Between Postgraduate Students. *Assessment & Evaluation in Higher Education, 25*(2), 149-169. doi: 10.1080/713611428

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning, 46*(2), 327-369.

Tsai, Y.-S., & Gasevic, D. (2017). *Learning analytics in higher education-challenges and policies: a review of eight learning analytics policies.* Paper presented at the

Proceedings of the seventh international learning analytics & knowledge conference.
Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education, 28*(1), 79-105.
van Leeuwen, A., van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: a case study. *Learning: Research and Practice, 3*(1), 42-58.
Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction, 20*(4), 316-327.
Van Waes, L., Leijten, M., Wengelin, A., & Lindgren, E. (2012). Logging tools to study digital writing processes. *Past, present, and future contributions of cognitive writing research to cognitive psychology*, 507-533.
Villalón, J., Kearney, P., Calvo, R. A., & Reimann, P. (2008). *Glosser: Enhanced feedback for student writing tasks.* Paper presented at the Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on.
Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development, 53*(4), 5-23.
Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal, 3*(1), 22-36.
Watson, J. M. (2012). Box plots in the Australian curriculum. *The Australian Mathematics Teacher, 68*(3), 3-11.
Whitelock, D., Field, D., Pulman, S., Richardson, J. T., & Van Labeke, N. (2014). *Designing and testing visual representations of draft essays for higher education students.* Paper presented at the The 2nd International Workshop on Discourse-Centric Learning Analytics, 4th Conference on Learning Analytics and Knowledge.
Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and evaluation in higher education, 17*(1), 45-58.
Wingate, U. (2012). Using Academic Literacies and genre-based models for academic writing instruction: A 'literacy'journey. *Journal of English for Academic Purposes, 11*(1), 26-37.
Wininger, M. (2014). Measuring the evolution of a revised document. *Journal of Writing Research, 6*(1).
Wise, A. F. (2014). *Designing pedagogical interventions to support student use of learning analytics.* Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge.
Wise, A. F., & Cui, Y. (2018). *Envisioning a Learning Analytics for the Learning Sciences.* Paper presented at the 13th International Conference of the Learning Sciences, London, United Kingdom.
Woods, B., Adamson, D., Miel, S., & Mayfield, E. (2017). *Formative Essay Feedback Using Predictive Scoring Models.* Paper presented at the Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education, 45*(4), 477-501.
Yu, S., & Lee, I. (2014). An analysis of Chinese EFL students' use of first and second language in peer feedback of L2 writing. *System, 47*, 28-38.
Zhang, F., Hwa, R., Litman, D. J., & Hashemi, H. B. (2016). *ArgRewrite: A Web-based Revision Assistant for Argumentative Writings.* Paper presented at the NAACL-HLT 2016 (Demonstrations), San Diego, California.

Zukerman, I., & Litman, D. (2001). Natural language processing and user modeling: Synergies and limitations. *User modeling and user-adapted interaction, 11*(1-2), 129-158.

Appendices

Appendix A: Hand out on rhetorical moves provided to Law students

Preparation for Benchmarking and AWA Activities[18]

Essay structure

Good academic writing follows a particular path:

A Introduction I - Identify the issue (establishing territory)

B Introduction II -What is the particular point to be argued (identifying a niche)

C Content I - Present existing case law or scholarship (presenting current work)

D Content II – Supporting assertions and with evidence

D Content III - Evaluate current work and justify your argument

E Conclusion - Announce principal outcomes

What are rhetorical moves?

- Rhetorical moves are explicit indications by an author to a reader as to the purpose of content in a paragraph or section of text.
- Rhetorical moves indicate to the reader where on the essay's path they are up to.
- Rhetorical moves are made with the use of discourse markers.

Discourse markers

- Discourse markers are the words or strings or words that indicate to the reader that a rhetorical move is being made by the author.
- Discourse markers can operate in 4 key ways:

  1) previews (e.g., There are four stages...);

[18] The guide was prepared by Dr. Philippa Ryan, UTS, who co-designed the Law interventions described in this thesis.

2) summarisers (e.g., To sum up so far...);

3) emphasis markers (e.g., This is the key), and

4) logical connectives (for example, However, ...)

- Logical connectives provide links between sections of text, revealing the author's attitude to the text
- Discourse markers are procedural and they connect text that is conceptual.
- Without discourse markers, essays are rendered a mere shopping list of concepts

Examples of discourse markers and their function

Discourse marker	**Function**
This essay will argue This essay will conclude In summary It has been concluded Ultimately	Summarising
It is important to note Significantly Importantly	Expressing importance
Further What is more	Adding something to a previous concept
On the other hand In contrast However	Contrasting two separate things
Therefore Consequently Thus	Saying what the result of something is
In spite of this Nevertheless	Making an unexpected concession or contrast
Provided that So long as	Expressing a condition
In response to this current trend An emerging pattern Recent developments suggest	Indicating a trend
The preferred approach is A better view might be It is argued that	Indicating a position

Appendix B: Self-Evaluation Exercise (SEE) prompts for rhetorical moves using AcaWriter[19]

1) Upload your draft report to **AcaWriter** (we encourage you to upload a version that you have revised/worked on since Step 2)
2) Download and print the PDF showing the AcaWriter feedback on your report
3) Print and review the **Acawriter feedback. Do you agree with the feedback given? Why/ Why not?**

_____

_____

_____

4) On the printed copy of your report (that shows the AcaWriter Feedback), use highlighters and comments to add to the feedback by showing where your report use the following **rhetorical moves**. You should submit this report.

NOTE: AcaWriter will be able to identify most of these rhetorical moves, but not always! It is important to use your (honest) human judgement too.

Organisational analysis Where does your report provide contextual information about the organisation's objectives, strategy, structure and activities? **Defining performance** Where does your report provide your **perspective [P]** about how to define performance or success for the organisation?

[19] The prompts are part of a 4-step self-evaluation exercise prepared by Dr. Nicole Sutton and Raechel Wight, UTS, who co-designed the Accounting intervention described in this thesis.

Where does your report **provide emphasis [E]** to highlight the most important aspects of performance for the organisation?

Justification of your definition of performance

Where does your report provide convincing, persuasive justifications for your definition of performance by proposing **novel [N] or critical insights, contrasting ideas or tension [C]**?

Where does your report justify your definition of performance with reference to **background information** or **previous work [B]**?

Written communication

Where in your report do you use appropriate **summary statements [S]** to signal the content, sequence and goals of the report?

5) Spend some time working on the soft copy of your report in the AcaWriter tool, adjusting your report and re-run the AcaWriter analysis. What effect did your changes have in on the feedback from AcaWriter?

_____

_____

_____

6) After using Acawriter what changes did you/will you make to your report?

_____

_____

_____

Appendix C: Instructor Interview Guide

The purpose of this interview is to understand the views of the instructors who have used AcaWriter to help their students improve academic writing. This will help researchers to improve the feedback from AcaWriter and to better support academics in using it for their subjects. The interview will be fairly unstructured but guided by some sample questions given below. The interview will be recorded and transcribed for research purposes.

Motivation, experience, and expectations

We want to understand a little about the learning context, and your involvement in the project, so we have a few questions about that…

- What was the writing context you wanted to use AcaWriter for?
- What is your prior experience in the use of technological innovations in your classroom?
- What was the motivation for you to be involved in the project?
    - What did you expect to use AcaWriter for? What did you think it would be helpful for? What were your expected outcomes?
    - What were the new affordances you thought AcaWriter would add to your existing teaching plan?

Implementation and Usage

To introduce AcaWriter or AWA in your subject, you worked with researchers in CIC to implement a learning design, so we have a few questions about the design and implementation…

- Can you tell us a bit about how you designed and implemented the use of AcaWriter?
    - For the Law instructor: Can you tell us a bit about how you designed the tasks in the writing intervention for your context? How did they evolve in the many iterations over time?
    - For the Accounting instructors: In your subject, we provided the example intervention in Law, with various steps. Can you tell us a bit about how you adopted and adapted the learning design from that context to yours?

- How did you find the process of working with researchers?
- Did you face any problems or constraints during the implementation? How easy was this process of working with researchers and implementing the innovative approach in your class?
    - How much time did you spend preparing for it?
    - How much effort did it require to implement the intervention?
- Did you feel like you had enough agency/power in how the intervention was designed for your class?

Findings, value added and future usage

The next few questions are focused on the impact of the intervention, and how effective the AcaWriter tool is...

- What impact do you think the intervention has had on learning?
    - Do you think students learned/ engaged more in their writing?
    - Did AcaWriter help your students become more aware of and reflect on their writing?
- What value did you think it added to your teaching?
    - Did the tool and/or intervention encourage you to reflect on your previous teaching of academic writing?
    - Do you think AcaWriter improved the efficiency of your teaching?
    - Did you learn anything new from this intervention?
- How do you see the role of student data and analysis in academic writing?
    - Is it clear how and why the writing data is being analysed?
    - Does the AcaWriter tool help make student learning (e.g. key features of academic writing) visible? (to them and/or you?)
    - Do you have any concerns e.g. about whether the AcaWriter tool makes it clear what data is being collected
- Can you see any improvements that could be made to the tool or/and intervention?
    - What changes would you make in the future?
    - Will you use it again in future semesters?
    - What could we do to support other academics to adopt the tool?

Appendix D: Tutor Feedback Questionnaire

To trial AcaWriter and study if it helps students learn to reflect more on their writing, we worked closely with *[subject co-ordinator names]* to create meaningful writing tasks with AcaWriter in *[subject name]*. We are researching the effectiveness of these interventions, and would like to receive your feedback (both positive and negative). As tutors who facilitate these activities in class, you play a key role in bringing these technologies to students for effective use. We'd like to hear your thoughts on how these interventions can positively influence student writing, as well as limitations or challenges you've noticed, and any support you feel would help tutors in the future. Data is collected as part of the Academic Essay Self Evaluation and Revision project ([Ethics number], approved by UTS Ethics Review), email antonette.shibani@uts.edu.au for any questions. By completing this form, you consent to use this data for research. We'd appreciate if you can provide detailed responses to the questions below.

[Section 1 for Law:] AcaWriter and Writing Intervention: To help improve students' written communication in Law essays, a writing intervention was designed for students with online activities including self-assessment and revision with exemplar essays in a tutorial session. Students also used the AcaWriter software providing automated feedback on rhetorical moves. Below are a few questions about that:

[Section 1 for Accounting:] AcaWriter and Writing Intervention: To help improve students' written communication in business reports, a writing intervention was designed for students with homework activities in Week 1, in-class discussion in Week 2, and a self-evaluation exercise for assignment, using the AcaWriter software providing automated feedback on rhetorical moves. Below are a few questions about that:

1. How familiar are you with the use of Writing analytics software (E.g. Turnitin, Grammarly, AcaWriter) in general?

1 (Not at all familiar) to 3 (Very familiar)

2. To what level do you understand the role of AcaWriter and the writing intervention in developing students' written communication?

○ I fully understand

○ I somewhat understand

○ I do not understand at all

3. Would you require more training to facilitate Acawriter feedback discussion in class?

○ Yes

○ No

Any other comments?

[Section 2] Time and Effort: We would like to know about the time and effort required to facilitate the intervention in class. So here are some questions about that:

1. How easy was it to implement the writing intervention in your tutorial? (Was it a lot of effort?)

1 (Not at all easy) to 3 (Very easy)

2. How much time did it require for you to understand and implement the writing intervention with AcaWriter feedback?

○ Less than an hour

○ About an hour

○ More than an hour

Any other comments?

[Section 3] Outcomes: We would like to know if there were noticeable outcomes in your students that you observed while in class and/or in marking their assignments. The below questions focus on that:

1. How useful was the writing intervention with AcaWriter for students to improve their writing?

1 (Not at all useful) to 5 (Very useful)

2. Did you notice improvements in students' writing skills? Why/ why not?

3. Did you notice changes in students' engagement with writing (E.g. Did they learn to self-assess/ reflect more on their writing/ include more rhetorical moves?) Why/ why not?

4. What value did you think it added to your teaching? Did you learn anything new?

[Section 4] Future Directions: We have some final questions about possible future interventions to improve student writing:

1. Would you be interested in using the intervention and AcaWriter again in future semesters?

○ Yes

○ No

2. What changes would you make in the future?

Is there anything else you would like to know more about?

Appendix E

Highlights of findings for the research questions

Research Question (RQ)	**Iteration**	**Key findings**
<u>RQ1. Writing Products:</u> What is the impact of automated feedback on rhetorical moves in student writing?		
RQ1a. What are students' perceptions of the writing task with/ without automated feedback?	2, pilot for 3, 3, 4	-Automated feedback from AcaWriter showed higher impact than no feedback in the perceived usefulness rating of students on the writing intervention in stable design iterations. -Actionable automated feedback that is contextualized for the subject was perceived most useful to improve student writing.
RQ1b. What is the impact of automated feedback on student revisions?	2, 3, 4	-Manual and automated methods of studying revisions including: assessment using scores, calculation of metrics like word counts, text dissimilarity cosine distance, and graphical representations like n-gram graphs and rhetorical moves graph were exemplified. - Automated feedback group produced significantly more improved essays than the No feedback group. - Students who deeply engaged with AcaWriter feedback by understanding and applying it on their writing scored higher marks in their assignment than those with shallow engagement.

RQ2. Writing Processes: How do students engage with automated writing feedback?		
RQ2a. How can we study students' interactions with automated feedback?	3	-Introduced a novel revision graph using text similarity and graph visualization techniques to study student interaction with feedback and revisions they made to the text at sentence level. - Most students had a moderate number of automated feedback requests, and the number of requests did not relate to their performance. - Multi-stage revision graphs uncovered high, low and moderate interaction behaviours of students engaging with automated feedback from AcaWriter.
RQ2b. How does scaffolding (using peer feedback and additional instruction) impact student engagement with automated feedback?	3, 4	- Peer discussion and feedback scaffolded students' engagement with AcaWriter to some extent: helping students identify new ways to improve their writing, suggest improvements, share their interactions, get clarifications, and critique the automated feedback, leading to deeper engagement. -Prompts for self-evaluation made visible students' thinking when using AcaWriter feedback. -Students who engaged with AcaWriter feedback in a deep way by understanding and applying it on their writing scored higher marks for written communication than those with shallow engagement.

<u>RQ3. Educator Perspectives:</u> What are practitioner perspectives on automated writing feedback in authentic practice?		
RQ3a. What factors influence adoption in authentic classrooms?	End of all iterations	-Improving students' written communication and self-assessment, ability to provide direct feedback for students, saving time while working with large classes, instructor knowledge of writing and motivation to support it, and openness to the role of technology motivated instructors to use AcaWriter as part of their teaching. -The instructors identified challenges including resource costs, working with imperfect tools, pressures on disciplinary-research outputs, and the need to consider wider teaching teams in implementing learning analytics. - The collaborative nature of the project and the agency provided to them supported successful implementations.
RQ3b. How do the practitioners engage in implementation across disciplines, and what are the outcomes?	End of all iterations	-The design process was elaborated by instructors. They explained measures like maintaining authenticity of the experience, making it relevant to students in this design. - Instructors' outcomes included providing direct feedback to students more frequently, improved students' self-assessment and engagement with their writing, a general improvement in students' writing skills, developing their own knowledge and – over the long term – saving time.

[i] Quotes from interviews are given verbatim, although are not exhaustive of related content. Where words have been removed for brevity this is indicated […]; in all such cases we do not believe that the removal has changed the meaning of the quotation. Detail is added in some places using square brackets, for example, to expand the acronyms that academics use in their everyday discourse.