Reflective Writing Analytics for Actionable Feedback

Andrew Gibson[1], Adam Aitken[1], Ágnes Sándor[2], Simon Buckingham Shum[1],
Cherie Tsingos-Lucas[1], and Simon Knight[1]
[1]University of Technology Sydney, Australia
[2]Xerox Research Centre Europe, Meylan, France
[1]firstname.lastname@uts.edu.au,[2]agnes.sandor@xrce.xerox.com

ABSTRACT

Reflective writing can provide a powerful way for students to integrate professional experience and academic learning. However, writing reflectively requires high quality actionable feedback, which is time-consuming to provide at scale. This paper reports progress on the design, implementation, and validation of a Reflective Writing Analytics platform to provide actionable feedback within a tertiary authentic assessment context. The contributions are: (1) a new conceptual framework for reflective writing; (2) a computational approach to modelling reflective writing, deriving analytics, and providing feedback; (3) the pedagogical and user experience rationale for platform design decisions; and (4) a pilot in a student learning context, with preliminary data on educator and student acceptance, and the extent to which we can evidence that the software provided actionable feedback for reflective writing.

CCS Concepts

•Applied computing → Computer-assisted instruction;

Keywords

Reflective Writing Analytics, Feedback, Learning Analytics

1. INTRODUCTION

Tertiary institutions are continuously confronted with the issue of assessing students in a way that is authentic, promoting learning for the future. It is generally accepted that when used well in education settings, reflective tools enhance lifelong learning and professional practice [30, 39]. When students engage in reflection on action [34], they are self-critical, identify and analyse their responses to challenging issues, and are prompted to reflect on how this experience reveals knowledge that could be applied in future. In this way reflective processes in learning become authentic when they are formative and future-oriented [3], and when higher level, meta-cognitive thinking about experience is required for students to connect their thinking to a wider world [14].

Reflective writing is promoted by many researchers as a valid way to assess reflective practice in a tertiary context [15, 28]. Reflective writing allows students to engage with both certainty and with what they do not know about a situation [27]. This typically involves students documenting their experiences, and writing reflective essays, journals and blogs based on their personal reflections on those experiences. The challenge is in how actionable feedback can be provided at scale when large numbers of students are required to write reflectively. For any feedback to be actionable by the student, it needs to be "... of sufficient quantity; timely; it should focus on learning not marks; it should be related to assessment criteria and be understandable, attended to and actually used by students to make improvements on their work." [25, p. 337]

Along with actionable feedback teaching students to write reflectively requires clear definitions, instructional instruments and rubrics of reflection. Reflection is thinking with a purpose about complicated or unstructured ideas where there is no obvious solution. [23]. From a pedagogical perspective reflective writing is directed towards a learning goal [4], and a corrector of distortions in thinking [22].

Because reflection is not necessarily an inherent skill the art of writing reflectively can be challenging [40], in part because reflective writing is different in nature and purpose to analytic academic writing, and students can struggle to present a personal stance within an academic context.

1.1 The Project

The A3R (Authentic Assessment Analytics for Reflection) research project was established to investigate the extent to which automated Reflective Writing Analytics (RWA) might offer potential to deliver formative feedback on student reflective writing at scale, and whether such feedback might encourage action on the part of the student. RWA involves the automated analysis of reflective writing, and the feeding back of analytics to the writer [14]. The key question motivating this research was to investigate how RWA might enable a scalable provision of actionable feedback to students on their reflective writing, and thereby augment existing pedagogical approaches to authentic assessment and reflective practice.

1.2 Participation

Courses across a range of disciplines at the University of Technology Sydney (UTS) use reflective writing assessment as it has been shown to be beneficial in fields as diverse as Engineering [27], Pharmacy [40], Business [36], Medicine [32],

LAK '17 March 13–17, 2017, Vancouver BC, Canada

DOI: 10.475/123_4

Psychology [10], Nursing [44], and Teaching [24].

For this research, we worked primarily with academics from seven subjects across three disciplines (Science, Business, and Engineering). Our academic partners provided reflective writing rubrics and samples of student reflective writing, primarily in the form of journal entries and reflective essays.

The project, which extended over two semesters (March 2016 - November 2016), involved meetings with members of the university's teaching and learning unit (IML), discussions with academic language and learning (ALL) experts, and workshops with representatives from all stakeholder groups to discuss the way reflective writing is used across the curriculum. The high level of stakeholder engagement was key to establishing the pedagogical objective of actionable feedback, and grounded the research in the learning and teaching of the university.

1.3 Approach

Despite the diversity of perspectives that came as a result of working across different disciplines, the academic partners reported common problems with reflective writing assessment: (1) Large numbers of students, many of whom were unfamiliar with reflection and writing reflectively; (2) Difficulty in clarifying the key features of good reflective writing; (3) The need for students to receive timely formative feedback on their writing, but a lack of staff time to meet this need; and (4) A need to ensure that any feedback provided is actionable by students in a way that maximises the benefits of the learning task.

The A3R project redesigned and extended (§4) the AWA (Academic Writing Analytics) software platform which was developed through previous research [18, 8, 35] .[1] A focus on actionable feedback for students resulted in AWA being redesigned to accommodate multiple layers of feedback, drawn from a larger set of analytics which were derived using multiple Natural Language Processing (NLP) techniques. The approach to analysing reflective writing was also redesigned with the aim of serving actionable feedback.

We took a theory first approach to the research, synthesising a large body of previous research on reflection and reflective writing, together with the practical implementation as evidenced by our academic partners. We then implemented the resulting framework computationally in the AWA platform, and applied the software in a learning context. This approach yielded a number of contributions that we believe advance the field of Learning Analytics (LA): (1) A theoretical framework for reflective writing; (2) a method for using the framework computationally to analyse reflective writing; (3) a pedagogically driven design approach to LA; and (4) preliminary evidence supporting our approach. Our approach is detailed in the sections that follow.

2. THEORETICAL FRAMEWORK

The A3R approach is grounded in the theory of reflection and reflective writing [11, 22] and informs a framework (see figure 1) which can be used both for human analysis of reflective writing and as a basis for developing methods of computational analysis. A key objective was to accommodate the major findings of previous research together with the main aspects of the various assessments for the subjects that we worked with. This approach ensures that it is pedagogically useful for the subjects in which it is applied.

[1]AWA aims to make visible to learners the ways in which they are using (or failing to use) language, and to do this in a way that is scalable . Designed to complement existing curricula tools AWA does not automate the grading of texts.

2.1 Reflective Writing

With a brief to synthesise key ideas and relate them to the current practice of our academic partners, we examined the dominant research on reflective writing [29, 17, 23, 37].

A key pedagogical issue is how feedback can guide students to reflect more deeply and to learn rhetorically purposeful language to achieve this [19]. A number of researchers characterise depth of reflection as a shift from a descriptive style of mere impressionistic reporting of events, through to a more critical style that focuses on integrating, analysing, and restructuring experience [30, 17], and on outcomes [4].

In terms of how feedback can guide learner's actions, the framework maps five levels of depth. Students are prompted to reflect on these by considering self-directed questions that correspond to these levels. This informed the vertical dimension of the framework (figure 1). The levels range from the lowest simple impressions through to the highest level of an intention to act:

1. **Impression:** What is happening around me? What is important to me?
2. **Interpretation:** How do I make sense of my impressions within my current situation?
3. **Internalisation:** How does this relate to me, to my knowledge, my wider context, my learning, my disposition, my emotions? How does it make me feel, and what are my reactions?
4. **Integration:** How does this fit with other knowledge, experiences, and differing perspectives? Can I learn from others?
5. **Intention:** Why am I concerned with this, and what do I intend to change with regards to myself based on what I have come to understand through being reflective?

These levels accommodate theoretical models together with pedagogical descriptions of the actions that might be expected of students when writing. These actions were drawn from a range of assessment instruments including rubrics from our academic partners, and through the analysis of a range of examples of student writing from across the disciplines.

2.2 Student Writing Analysis

For many students reflective writing in a tertiary setting is a novel genre, and tutors may be inexperienced in its assessment. While subjects provide rubrics for reflective writing assessments, these do not always provide explicit instruction on the language of reflective writing. UTS academic communications expert Rosalie Goldsmith designed a resource that helps engineering academics recognise linguistic features of reflection in student journals [8]. Goldsmith itemised frequently occurring lexico-grammatical features, e.g. thinking and feeling reporting verbs (I felt, I realised, I became aware...), statements of a challenge, a critical incident, expressions of learning intentions and more.

Rhetorical analysis methods [38] and genre theory [21] further informed our analysis of: (1) The lexico-grammatical features of student reflective writing; and (2) the discursive

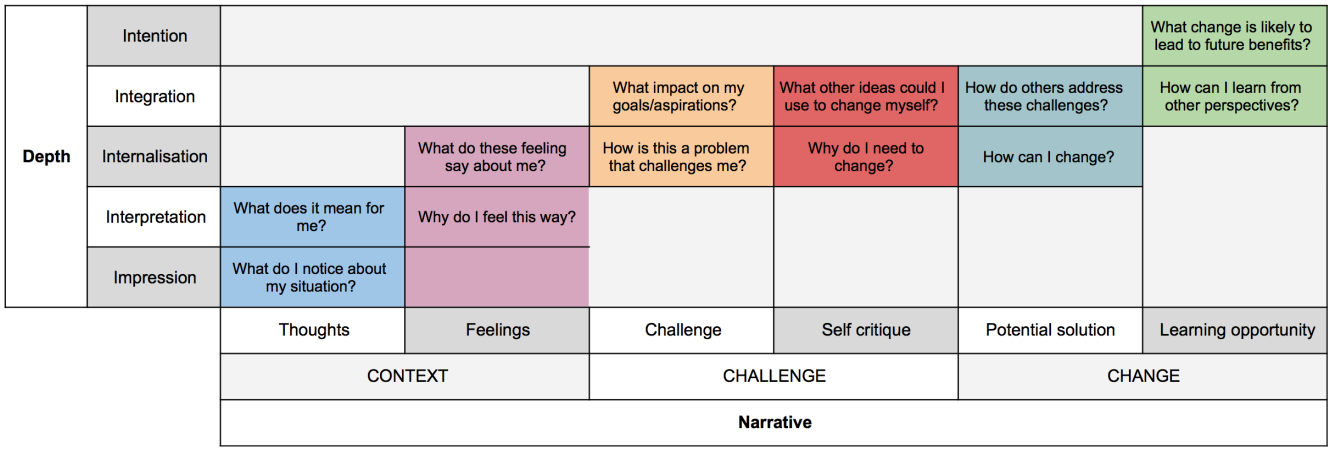

Figure 1: The Reflection Framework which synthesises theories of reflection and narrative together with discipline approaches to assessing reflection.

structure and cohesive patterns most often found in whole texts. Depth and structural dimensions are marked by rhetorical 'moves' or recurrent discourse elements that characterise written and oral genres and which perform a coherent communicative function [38]. Used with intent by writers, moves realize 'rhetorical action' [13]. To take an example of analytic writing, a scientific research paper will include a statement of a research gap in the abstract and introductory stages [38]. Structurally, the research paper obeys reader expectations of the genre, and requires a sequenced orderly progression of these moves.

In reflective writing rhetorical moves communicate personal shifts in perception, express self-critique and self-doubt, and register changes in belief. Its rhetorical, lexico-grammatical and structural features, considered together, suggest that reflective writing is a genre, by which we mean formalised writing with a goal-oriented communicative purpose, and which is structurally constrained in a series of stages suggesting reporting or narrative form [21]. Different models of moves or stages have captured types of reflection [42, 41] and also depth of reflection [23, 30, 44]. Of particular salience was Birney's [2] study of teacher trainee reflection, a fine-grained systemic-functional mapping of reflection depth with linguistic realisation. This research concludes that strong reflective writers are able to draw on up to ten linguistic features ranging from use of feeling and thinking verbs to adjectives (e.g. 'a positive impact') and reasoning adverbs (e.g. 'extremely challenging'), and future tense modality (e.g. 'I intend to bring this into my future practice').

In our framework a horizontal dimension represents a narrative-like sequence of rhetorical moves, while a vertical dimension models levels of depth. The intersection of depth and sequential dimensions represent the possible stages through which the text moves and the linguistic features used to realise these stages - from 'shallow' impressionistic description to 'deep' future-minded intentional statements. To prompt students to improve their reflection, the framework's prompt questions encourage them to move from 'shallow' to 'deep' levels of reflection (see figure 1).

Indeed, an analysis that we carried out on student reflective texts confirms that similar reflective writing moves occur in different disciplines. Our work also confirms other studies that show strong writers utilise a wider range of linguistic features than weaker reflective writers [19]. Highly rated journal writers attempt to make sense of the relationship between themselves and their situations [9], and question their own assumptions that underpin their actions [37]. They reflect on how to proceed in the face of uncertainty [27], and reflect on decisions [26]. We found that the rhetorical moves announce a personal response to a learning context (Context), acknowledge the challenging nature of learning (Challenge), and indicate a learning experience is self-transformative (Change) (See figure 1).

For example, one student wrote about how habitual ways of thinking lead to problems ('obstacles'), which then lead to a new or 'rapid' understanding of the situation:

> ...I rapidly understood that language was not the most important barrier. Our ways of thinking, acting, and our values became obstacles we needed to overcome.

When interpreted as a coherent rhetorically staged text, we find that the writer emphasises a rapid change in understanding (I rapidly understood) as well as the notable *challenging* element in the learning situation ('the barrier', 'obstacles'). They then note that this situation leads to a *change* in perception, and a departure from a previous habitual view (language is the main source of misunderstanding). We show how these semantic relations inform computational modelling in section 3.

Feedback from academic partners confirmed that they found the framework to be a helpful lens through which to view their reflective writing assessment, and that it would assist in giving feedback to students in a way that might encourage them to improve their reflections.

2.3 Simplification

After assessing the extent to which existing technologies available to us might enable us to operationalise it computationally, we simplified the framework, especially in the way we approached the levels of depth. While we could see the potential for using the linguistic properties of the sentences to identify key moves such as `CONTEXT`, `CHALLENGE`, and `CHANGE`, it was not obvious to us how we might detect levels

within the depth dimension. Therefore we simplified depth by distilling a dominant common characteristic feature in deeper reflection pertaining to all the three moves: that it is applied personally. We identified that this is indicated where the student links one of the key moves to themselves. We referred to this feature as `LINK2ME`.

We also noted that the framework might inform feedback to the student in ways other than by the rhetorical moves, such as: (1) Indicating affective elements and emotive expressions in the writing; (2) drawing attention to how good reflection links to existing knowledge via epistemic expressions; and (3) indicating critique, particularly self-critique, as a key feature of depth.

Thus, our final simplified framework for feedback purposes comprises three moves (`CONTEXT, CHALLENGE, CHANGE`), a modifier of these moves to indicate depth (`LINK2ME`), and three expression types (`EMOTIVE, EPISTEMIC, CRITIQUE`). This provided a much more accessible version of the reflective writing framework for students while maintaining its connection to the theory. The simplified version also assisted with the computational implementation which was tackled by two different processes: (1) The implementation of the modelling of reflective rhetorical moves. This is detailed in section 3; and (2) the aggregation of different computational analyses, and implementation of multiple layers of feedback. This is described in section 4.

3. REFLECTIVE RHETORICAL MOVES

Building on the theoretical and linguistic description of reflective writing moves in Section 2 we have developed a system to computationally detect sentences that convey the three primary rhetorical moves of reflective writing determined by the A3R framework: Context, Challenge and Change. Each of these can be augmented by the Link2Me feature. To detect sentence elements indicating the rhetorical moves we used the concept-matching rhetorical analysis framework [33]. The analysis model is implemented in the natural language processing tool Xerox Incremental Parser (XIP) [1].

3.1 The concept-matching analysis framework

The concept-matching framework models rhetorical moves as pre-defined patterns of abstract elements called constituent concepts. Constituent concepts are determined by the definitions and descriptions of the rhetorical moves. For example, the Challenge move is defined in the A3R framework as "the challenge of new surprising or unfamiliar ideas, problems or learning experiences". The constituent concepts suggested by this definition are `CONTRAST` (cf. challenge, unfamiliar, problem), `STANCE` (c.f. surprising, experiences), `ANALYSIS` (c.f. ideas, learning) and `SUBJECT` (the author). Any combination of these concepts constitutes a possible pattern for Challenge, as in the following formula:

```
CHALLENGE =
(SUBJECT+ANALYSIS) AND (CONTRAST+ANALYSIS)
```

The '+' designates grammatical coherence and 'AND' co-occurrence in the same sentence. This formula will match sentences that describe contrast in the author's thinking. The following sentences are examples that match this pattern (constituent concepts in bold):

1. I reflected on this and felt decision making was like second nature, yes **I over-thought my decisions** whether it was personal or professional **but I never thought** of the act of having to justify my decisions.
2. **I** continued to **contemplate how I** was going to tactfully **address** this questionable behaviour.

While traditional approaches to analysing rhetorical writing rely on the detection of lexical and lexico-grammatical features, the distinctive advantage of the concept-matching framework lies with the grammatical coherence constraint coupled with the pre-defined patterns, which is able to filter out noise.

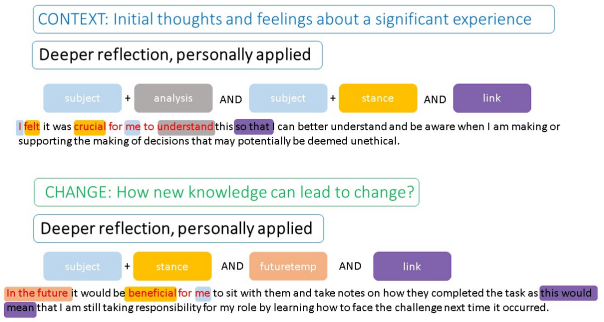

Figure 2: Example rhetorical patterns of the A3R reflective moves and illustration sentences.

3.2 Implementation in XIP

XIP is a natural language analysis tool whose basic function is to provide deep syntactic analysis of sentences. XIP integrates statistical processing and grammar rules, which were developed using its dedicated rule writing mechanism. Basic XIP analysis incrementally executes a chain of treatments for each sentence of a document from segmentation through morpho-syntactic analysis, part-of-speech disambiguation and chunking to syntactic dependency extraction, i.e. the identification of syntactic functions, like subject, object, modifier, etc.

The concept-matching framework was implemented as an additional XIP module, building on the general dependency output, and using the rule writing mechanism. The implementation task consisted in detecting sentences that contain a rhetorical pattern defined in the concept-matching framework. This involved the following steps: (1) The development of a lexical database where the constituent concepts are associated with words and expressions; [2]; (2) creating XIP rules that select out of all the dependencies extracted by the general XIP syntactic analysis module those that are relevant for the rhetorical patterns; and (3) creating XIP rules that mark the sentences that contain a set of dependencies of the pre-defined rhetorical patterns. The pattern matching rules were developed using a small development corpus of ten reflective essays of various domains annotated according to the language analysis methods our reflective writing framework.

[2]The lexical resources have been acquired by various means: importing from existing rhetorical analysis modules that use the same constituent concepts, and using (morpho)-syntactic and semantic properties and part-of-speech as features, which are provided by the general XIP parser. The lexicon for the STANCE concept has been imported from a sentiment analysis module of XIP [5]

4. ACTIONABLE FEEDBACK

Our moves of Context, Challenge, Change and Link2Me were accommodated by the implementation of reflective rhetorical moves in XIP. However, much of the reflective writing literature had indicated that there were other important indicators of reflective writing, such as the three expression types (Emotive, Epistemic, and Critique) that we identified in our simplified framework. We also noted that some features might present not at the sentence level, but rather in the text as a whole, such as the general structure of the writing, or the overall sentiment expressed. There was also research that the situational context in which the reflection was based was significant, indicating that there may be specific vocabulary for different writing contexts. These factors indicated that we needed more than the sentence level reflective rhetorical move analytics. We needed to incorporate other NLP techniques that could provide us with varying levels of analysis, and a way of aggregating the resultant analytics.

Further, our objective was to provide actionable feedback to the writer, and so we needed a mechanism for creating feedback from aggregated results, and a way of evaluating the actionability of this feedback.

Responding to these challenges required significant design decisions in terms of both the architecture of the platform and the desired user experience. We outline our approach to both of these areas in the following sections.

4.1 Platform Architecture

Throughout the A3R project, AWA evolved from a web application calling remote analysis service into a platform with multiple services deployed across a mix of local and cloud infrastructure. The design was driven by Information Architecture (IA), "the process of designing, implementing and evaluating information spaces that are humanly and socially acceptable to their intended stakeholders"[12]. In our case the intended stakeholders were the students and their teachers, and the information space was AWA, but conceived as learning environment where the students could improve their reflective writing.

Importantly, this meant a pedagogically driven approach to RWA. That is, architecting the platform in a way that considered actionable feedback in the whole of the user experience: starting from the way that student writing was conceptualised; including the approaches to analysis and the delivery of the feedback; and concluding with evaluation of the feedback's efficacy.

These architectural changes were facilitated by introducing a Text Analytics Pipeline (TAP), a modular cloud based application. TAP allowed the inclusion of analysis services other than XIP to be included in the platform, facilitated the aggregation of the resultant analytics, and generated multiple levels of feedback. An overview of the key elements of TAP can be seen in figure 3.[3]

The design of TAP allowed for the aggregation of multiple sources of analysis from both external services such as XIP, third party libraries like CoreNLP [20], and internal software. A detailed description of TAP is beyond the scope of this paper, however an overview of the process of generating feedback highlights the way TAP contributes to the AWA platform.

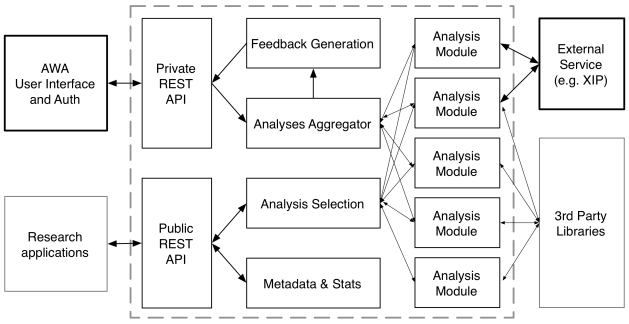

Figure 3: The core modules of the Text Analytics Pipeline (TAP) shown within dashed box, and connections to UI (top left) and XIP (top right).

Feedback is provided to the UI in the form of a JSON object composed of the differing feedback types (expression, sentence, paragraph, and document). Sentence level feedback is generated by passing the text through a module that connects to the external XIP service. The resulting analysis is then formatted similarly to the other feedback levels before being packaged up and sent back to the UI. Expression level feedback is generated using different software modules, the selection of which depend on the feedback type. 'Emotive' expressions are generated based on lexical comparisons with the Warriner [43] Corpus selecting high valence and arousal terms, whereas the 'Critique' and 'Epistemic' expressions are derived using techniques for identifying metacognition in reflective writing [14]. Paragraph and document level feedback involves complex rules applied to a range of analyses. For example, some document comments are generated based on the ratio of rhetorical moves that have been detected across the document. Other paragraph level comments combine a third party spell checking library[4] with named entity recognition from CoreNLP [20] and basic paragraph metrics to determine if a comment on spelling errors should be generated. Further detail on these algorithms can be found by examining the software.[5]

Importantly, the choice of modules, third party services, and design of the algorithms, were all driven by the pedagogical imperative to provide actionable feedback. Thus, the very architecture of the platform is shaped by the desired learning outcomes.

4.2 User Experience

The objective of actionable feedback based on our conceptual account determined that in addition to a sentence-level annotation (implemented in previous versions of AWA), feedback should also be provided at the sub-sentence (expression), and aggregate levels (both paragraph and whole document). These features were implemented, alongside user support documentation, as outlined below.

4.2.1 User Interface Layout

Actionable feedback requires a UI that prompts users to turn feedback into feed-forward[16]. To prompt re-drafting,

[3]TAP is written in Scala to run on the JVM. It is designed in a modular reactive style with Akka to facilitate flexibility and scalability as required.

[4]https://github.com/languagetool-org/

[5]The software can be found on the CIC GitHub Page (https://github.com/uts-cic)

students are alerted to what is needed in terms of reflective expression. Visual feedback has been shown to be more effective than verbal feedback which often appears complex and multidimensional [31].

Less certain is how and if certain UI features contribute in creating feed forward. A key design change concerned the accommodation of feedback to the user other than just sentence labelling and highlighting. There was a need to accommodate whole of document textual feedback, feedback associated with paragraphs, and feedback associated with groups of words or expressions. With the addition of at least three more types of feedback, a significant consideration was to ensure that the interface did not become too cluttered. For feedback to be actionable, it first needs to be comprehended. We anticipated that too much visual information would result in the user being less certain about which information should be acted up.

Because of this, our aim was a clean UI that maintained the prominence of the original text, but annotated with both sentence and expression level feedback. Paragraph feedback would be provided in a right hand margin aligned with the relevant paragraph, and whole document feedback would be provided above the text as it was intended to provide overall commentary on the writing. Taking this approach also had the advantage of being similar to what students may expect if they were to receive feedback on a printed document, i.e. parts of the text itself underlined and circled with comments in the margins and an overall summary comment. We hoped that this familiar format might assist with the overall usability of the interface. This layout can be seen in figure 4.

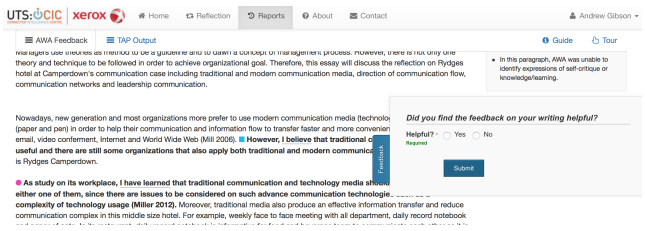

Figure 4: The AWA User Interface with a pane (lower right) for collecting a feedback response from the user.

To assist the user with understanding sentence and expression level annotation, a pop-out pane was provided that defaults to being visible when the analysis is displayed, but which can be clicked to reduce to a side tab after use (figure 5). To assist with evaluation, we also required a way to elicit a response from the user as to whether they thought the feedback that was provided on their text was useful (see bottom right of figure 4). The intention was to compare student responses via this mechanism with evidence of changes in subsequent drafts.

4.2.2 User Interface Elements

A key factor in the UI design was how the theoretical foundations and their analytics expressions were to be made visible to the user, and the extent to which they might be helpful and useable for actionable feedback.

Sentence annotation was impacted to the greatest extent by changes to the underlying analytics. The framework (§2), was simplified to ensure that it was easy to comprehend, and

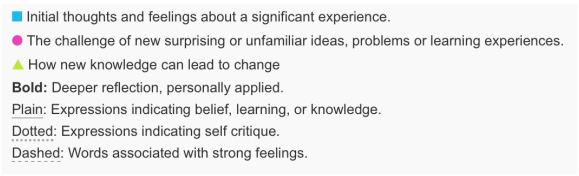

Figure 5: The UI elements used by sentence and expression level feedback

the labels derived in the process of developing the analytics (Context, Challenge, and Change) proved to be very ambiguous when presented to users. The previous version of AWA used named labels in the sentences, and our original intention was to maintain this style. However, the ambiguity caused significant problems in terms of our different users who might perceive the feedback and take action on it. After much deliberation, we settled on removing the names altogether from the UI, and settled on single sentence descriptions represented by the blue square, pink circle, and green triangle, as shown in figure 5. The final colour and shapes of the icons were chosen to allow for simplicity in the UI while accommodating users with colour blindness.

5. STUDENT USE

Student use of AWA in multiple subjects is an ongoing process with each subject having different due dates for their reflective writing assignments. As a result, for a preliminary examination of the data we have selected only those students from Pharmacy who have used the software at the time of writing this paper (from late August to mid October 2016). Although preliminary, this data has provided some insights on whether the feedback is actionable for these early users of the software. Other subjects involved in the research are expected to contribute further to the data during the remainder of the spring semester (through to late November).

5.1 The Pharmacy Learning Context

Preparing pharmacy students for the complexities and diversity of clinical practice is a consideration with pharmacy educators. A possible solution to bridging the theory/practice gap is through reflective practice [39]. One of the tools utilised to enhance reflective practice is through reflective writing. Reflective writing skills are paramount for students to think about incidents, their outcomes and how this affects the health of a patient. In 2016, First year Master of Pharmacy students (n=59) were offered the use of AWA to assist with improving their reflective writing skills. The Masters of Pharmacy Course is a 2-year intensive program which embeds reflective activities in the 520 hours of clinical placements. One aspect of the weekly reflective activities involve students writing reflective statements and uploading these to their e-portfolios. This process is integrated into the course to prepare students to make better informed decisions and clinical judgements for future practice. For example, students are required to: (1) reflect back on a weekly incident; (2) view the incident from different perspectives (eg from the perspective of the patient, carer, pharmacist, and/or other health professional); (3) write about what was learned, what challenges they undertook, what strategies they utilised to overcome these challenges; (4) recognise the strengths and skills they have, how these could further be developed, how

their behaviour or approach could be changed to enhance a future similar event; and (5) identify shifts in their beliefs and attitudes which have resulted.

5.2 Preliminary results

Data was collected from the student use of AWA in Pharmacy, and their responses represent a snapshot at the time of writing the paper.

Total students:	59	
AWA users:	30	50.8%
Total posts:	120	
Provided response:	63	52.5%
Found AWA feedback helpful	54	85.7%

Table 1: Usage and feedback data

Early discussions between the students and their teacher about their use of AWA anecdotally indicated that it "allowed them to use the tool in their own time" . . . "as often as they want to further critique their reflective writing, learn from the tags and change their statements prior to submission", and "has the potential to familiarise us who are new to the area of reflection and reflective writing skill development."

More formal responses were collected via the AWA UI. The feedback tab in AWA poses the question: "Did you find the feedback on your writing helpful?" Students using the software at this early stage have generally been positive about its helpfulness with 85.7% of those giving feedback selecting 'yes' - the feedback on their writing was helpful (see table 1). Of those who gave more detailed feedback ratings (see figure 6), about half provided a neutral rating (some of which could be explained by the fact that this was the default in the UI). Of those that made a positive or negative choice, the negatives were all just less than neutral (3), whereas the positives were an even combination of just positive (5) and strongly positive (7).

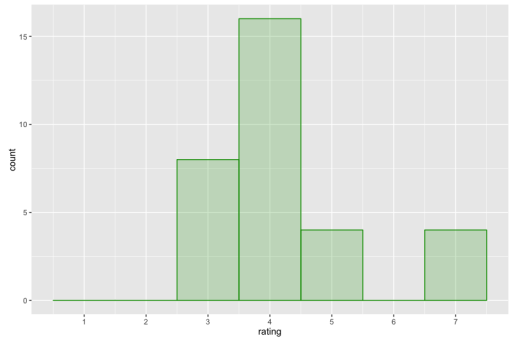

Figure 6: Pharmacy feedback ratings of AWA. 1 is not helpful and 7 is helpful. 4 is the default.

While we acknowledge that perceptions of helpfulness don't necessarily translate into action, we believe that feedback perceived as not helpful would be highly unlikely to result in any action. Therefore, we take these results as an early promising sign.

5.3 Student comments

Our cautious optimism was also supported by student comments that suggested that it encouraged them to think about action:

> *I was fascinated by how it works and can see its implication in future, to determine which phrases need more work/ which can be improved.* (Student A)

> *It details where I've made reflective statements and shows where I can improve as well as add to and fill in aspects to which I have not confirmed.* (Student B)

> *Prompted me to follow through with the reflection to the last step of the process - i had written about my thoughts and feelings, discussed challenges, but had not followed through with reflecting on how this can lead to change. . . . The reports also direct me to write more personally, using language that evokes emotion, and less descriptively* (Student C)

A comment by one student indicated that they saw the potential for AWA to assist them in improving their grade, and even suggested that it should provide a grade:

> *This system has allowed me to identify the strengths and weaknesses of my reflection, highlighting on what criteria I have addressed and which ones I haven't. I wish there was feedback on how I could improve to get full marks and wish this reflection gave a mark at the end.* (Student D)

No Pharmacy students left clearly negative comments. However, some early use of the software by students from other subjects highlighted a theme in their more negative comments: criticism either at what the software did not do, or towards a lack of clarity in what needed to change. For example:

> *Doesn't elaborate on the features that are lacking and often they appear there but are not recognised. Good way of highlighting other points.* (Student E)

> *its not clear what needs improving.* (Student F) and *comments are not clear enough* (Student G)

> *I don't understand what AWA reproach to my work. It is said that there isn't a good balance but In the text I can't see how... not clear* (Student H)

This lack of clarity for the student goes to the heart of the actionability of the feedback, suggesting that the students want to use the feedback to take action, but that it is not clear enough for them to do so. However, it is not clear if more data might reveal this as a significant issue or whether this was more indicative of the immaturity of the platform at the time.

5.4 Evidence of Action

Of the 30 users who posted text to AWA, 18 users (60%) posted more than once. We were interested in whether there was evidence of 'action from feedback' in the writing of these users. We postulated that if students were provoked to take action based on the feedback, then they would post a new draft to AWA to check if it addressed the feedback.

From those posting more than once, 5 users (27.8%) showed evidence of modifying drafts, and 2 of these users drafted

more than once on the same text. We view this as a positive indicator that our objective of actionable feedback can be attained. While most of the users (13 users, 72.2%) wrote different reflections for their multiple posts, this was not unexpected as the subject requires them to write a new reflection each week. What is difficult to assess with these users is whether feedback on one reflection was taken into account with subsequent reflections.

With the exception of addition of new information, most draft modifications appeared to improve the quality of the reflection. For example:

> *I made sure to be understanding and not force the customer to purchase just for the sake of receiving a sale (ethics, social responsibility).*

was changed to:

> *Initially I was confused as to why this would be an issue like isn't it exactly the same thing? But for good pharmacy practice, I decided to be understanding and not force the customer to purchase a product just for the sake of receiving a sale (ethics, social responsibility).*

The first of these had no sentence tagging, whereas the second was tagged with a pink circle (Challenge). Students also changed their writing to introduce how they felt about a situation. For example:

> *This situation didn't sit well with me.*

was changed to:

> *This situation didn't sit well with me, I felt as it these patients didn't receive the best care possible.*

6. DISCUSSION

6.1 Actionable Feedback

Our preliminary results from the use by Pharmacy students suggested that the feedback provided by AWA was helpful and actionable. However, we had also received some input from other subjects that suggested that the feedback lacked clarity. Although this is likely to be resolved through the collection of more data from multiple subjects, it raises questions as to whether it is the AWA platform that makes the feedback actionable, or whether there are other factors outside of the system that contribute to this. For example, the inclusion of teaching on reflection, and/or the requirement to write regular weekly reflections may assist the students in recognising the value of the feedback, and help them understand how to use it. A subject that has little teaching input and only requires a single piece of reflective writing may provide the students with less understanding on how to work with the software. Regardless of how a subject approaches reflective writing, we believe that alignment between the analytics and pedagogy is critical. The extent to which we can determine if this alignment contributes significantly to actionable feedback is yet to be found.

6.2 Contextual Feedback

Through the discussions with academic partners and the trial of AWA with students, we saw an interesting tension emerge between general and specific feedback. Early in the project, many stakeholders suggested that feedback should be as detailed and specific as possible. While this notion aligns with the educational literature on feedback [16], human feedback is always contextualised to some extent. We found that the lack of reference to context resulted in the rejection of detailed feedback. For example, when the same paragraph feedback was provided for many paragraphs in the text, stakeholders were less likely to appreciate the relevance of the feedback and more likely to criticise the repetitive nature of it. A contextual approach would mean that comments for one paragraph would be made 'in the context of' what is stated with other paragraphs.

Similarly, some paragraph feedback is more important in certain parts of the text than in others. For example, setting the context for the reflection can tend to be more descriptive and therefore emotive expressions are less important in these sections. An improved system of generating feedback would have an 'awareness' of these contextual interrelationships, and modify the feedback accordingly.

6.3 Modelling Disciplinary Differences

The current AWA platform is missing another contextual feature, and that is the ability to recognise and work with language and reflection style differences across disciplines. While we had hoped to include some modelling of disciplinary specific features in the platform, that has not been implemented to date. The lack of this ability resulted in some significant disagreement between project stakeholders with regards to structure and relating to knowledge.

In particular, the Engineering subject required reflective writing that was highly structured, and a business subject required the use of citations to show how the students were relating their reflection to the fields of existing knowledge. In both of these cases, the requirements were considered essential for actionable feedback in the subjects concerned, but feedback on these features were problematic for other subjects. While providing discipline specific versions of AWA might be a short term solution to this issue, in the long term it is intractable as there would need to be as many versions as there are subject. Instead, we believe that the ability to model differences within the subject, and allow AWA to make decisions according to combinations of models, might allow a future version of AWA to accommodate these needs. What we are certain of, is that these pedagogical requirements are important for learning, and that allowing them to drive the development of the tool holds much greater value for learning than changing the learning to fit the tool.

6.4 Algorithmic Accountability and Integrity

The 'transparency' and 'accountability' of algorithms are important qualities if we are to achieve an acceptable level of 'analytical system integrity' [6] in our educational tools. In the context of this project, this required us to understand the relationship between the computational representation of a text, and the human experience of giving and receiving feedback on writing. The computational representation concerns the fine-grained textual features, and the reasoning behind the algorithms for combining, and acting on, different patterns (§3). The default educator and student experience deploys a much simpler language, provided by the set of constructs in the document-centric display in which feedback wrapped around, and overlaid onto, the student's writing (§4). Our assumption is that educators should demand a

level of accountability that is less rigorous than a researcher might demand, but more rigorous than a student may demand. That being said, we should also expect and encourage students to demand accountability to whatever level of detail they require, and technically minded students might indeed be more demanding than their tutors as their data literacy grows, and they are encouraged to reflect critically how their activity is tracked in all spheres of their lives [7]. In the previous design iteration of this project [8] we noted that a key ingredient of any discussion of 'accountability' among design stakeholders is trust: "Trust is built through reciprocity, which in learning analytics design means ultimately, that you feel you can influence the code." We have continued that co-design process in which stakeholders (the UTS academics and Academic Literacies expert, and the Xerox computational linguist) shaped the performance of AWA until they felt that it was good enough to pilot with students as an experimental application. The fact that the design team involves academics who trust the tool, because they had the chance to test and give feedback on its performance with their own students' writing, should be a source of assurance for students (and indeed other academics considering trialling AWA). Ultimately, however, it comes down to the student experience of the tool, since many emergent phenomena may arise in authentic use. The preliminary feedback from students reported here, coupled with results from another deployment [18] provide encouraging early evidence that students valued AWA, but there remains much more to improve.

7. CONCLUSIONS AND FUTURE WORK

Through the A3R project we have created a new conceptual framework for reflective writing that synthesises the dominant theory. We have used a simplified version of this framework to develop computational approaches to reflective writing analysis and in doing so have created RWA that can be used for feedback to the writer. A significant aspect of the research was allowing the pedagogy to drive the design, not just of the framework, but of the platform architecture as well. A pilot of the software in the subject of pharmacy has shown this pedagogical design approach to be successful, with a early signs of the feedback being actionable by students.

Strengths of the project include the strong theoretical foundations of the work, and the maturity of the genre/narrative analysis approach and its instantiation in XIP. We believe that the research has also shown significant promise and the approach to incorporating multiple analytics for the generation of feedback as implemented in TAP. The centrality of actionable feedback, throughout the process has also been a significant aspect of the work.

However, we also uncovered a number of weaknesses. We found it difficult to meet the needs of all stakeholders, and the lack of data from all participating subjects means that at this point we are unable to ascertain the extent to which this has impacted the students. Nevertheless, early feedback from students suggested that some feedback lacked clarity, and our own analysis showed that in order to improve the quality of the feedback, we need methods for using contextual information.

These weaknesses set an agenda for future work. Firstly, we plan to develop over coming months a subject specific version of the platform for Engineering. We anticipate that this will test some of our hypotheses about the need for more specific feedback and the use of contextual information in the generation of that feedback. Secondly, we need to collect much more data from a greater variety of disciplines. This would allow us to substantiate our preliminary findings. Finally, while we have shown the potential of designing LA for actionable feedback, there is much more work to be done in this space. We expect that ongoing research in RWA will improve our ability to provide actionable feedback to students.

8. ACKNOWLEDGEMENTS

We acknowledge the input of Natalia Nikolova, Walter Jarvis, Alan Parr, Andy Leigh, Peter Jones, Jo McKenzie, Rosalie Goldsmith, Susan Hoadley, Isabelle Bennett, Sarab Mansoor, Keenan Wilson, and Jeff Browitt. We also acknowledge Xiaolong (Shawn) Wang for the UI software development, and Xerox Research Centre Europe (XRCE) for the Xerox Incremental Parser (XIP).

9. REFERENCES

[1] S. Aït-Mokhtar, J.-P. . Chanod, and C. Roux. Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering*, 8(2-3), 6 2002.

[2] R. Birney. *Reflective Writing: Quantitative Assessment and Identification of Linguistic Features*. PhD thesis, Waterford Institute of Technology, 2012.

[3] D. Boud and N. Falchikov. Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4):399–413, 8 2006.

[4] D. Boud, R. Keogh, and D. Walker. *What is Reflection in Learning?* Routledge, 1985.

[5] C. Brun. Detecting opinions using deep syntactic analysis. In *RANLP*, pages 392–398, 2011.

[6] S. Buckingham Shum. Algorithmic accountability for learning analytics. In *Invited Talk, University College London, Institute of Education & London Knowledge Lab. http://bit.ly/aala2016*, 2016.

[7] S. Buckingham Shum. Envisioning learning analytics for 21st century competencies. In *Keynote Address, LASI-Asia 2016: Learning Analytics Summer Institute, Seoul*, 2016.

[8] S. Buckingham Shum, Á. Sándor, R. Goldsmith, X. Wang, R. Bass, and M. McWilliams. Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 213–222. ACM, 2016.

[9] A. Caetano. Defining personal reflexivity: A critical reading of archer's approach. *European Journal of Social Theory*, 18(1):60–75, 9 2014.

[10] R. Dallos and J. Stedmon. 1 flying over the swampy lowlands: Reflective and reflexive practice. *Reflective practice in psychotherapy and counselling*, pages 1–22, 2009.

[11] J. Dewey. *Democracy and education*. Courier Corporation, 1916.

[12] A. Dillon. Information architecture in jasist: Just where did we come from? *Journal of the American society for information science and technology*, 2002.

[13] J. Flowerdew. John swales's approach to pedagogy in genre analysis: A perspective from 25 years on. *Journal of English for Academic Purposes*, 19:102–112, 2015.

[14] A. Gibson, K. Kitto, and P. Bruza. Towards the discovery of learner metacognition from reflective writing. *Journal of Learning Analytics*, 3(2):22–36, 2016.

[15] J. E. Griffin Jr, G. F. Lorenz, and D. Mitchell. A study of outcomes-oriented student reflection during internship: The integrated, coordinated, and reflection based model of learning and experiential education. 2010.

[16] J. Hattie and H. Timperley. The power of feedback. *Review of Educational Research*, 77(1):81–112, 3 2007.

[17] N. Hatton and D. Smith. Reflection in teacher education: Towards definition and implementation. *Teaching and teacher education*, 11(1):33–49, 1995.

[18] S. Knight, S. Buckingham Shum, P. Ryan, Á. Sándor, and X. Wang. Academic writing analytics for civil law: Participatory design through academic and student engagement. *International Journal of Artificial Intelligence in Education*.

[19] J. Luk. Assessing teaching practicum reflections: Distinguishing discourse features of the "high" and "low" grade reports. *System*, 36(4):624–641, 12 2008.

[20] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.

[21] J. R. Martin and D. Rose. *Working with discourse: Meaning beyond the clause*. Bloomsbury Publishing, 2003.

[22] J. Mezirow. How critical reflection triggers transformative learning. *Fostering critical reflection in adulthood*, 1:20, 1990.

[23] J. A. Moon. *Reflection in learning and professional development: Theory and practice*. Routledge, 2013.

[24] C. Morrison, J. Willis, L. Crosswell, and A. Gibson. Turning points in narratives of research design: Research innovation stimulating unique responses to existing challenges for beginning rural teachers. *The Journal of Educational Enquiry*, 13(1), 2014.

[25] D. Nicol. Assessment for learner self-regulation: enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education*, 34(3):335–352, 6 2009.

[26] J. M. Paxton, L. Ungar, and J. D. Greene. Reflection and reasoning in moral judgment. *Cogn Sci*, 36(1):163–77, 2012.

[27] C. Reidsema, R. Goldsmith, and P. Mort. Enabling the reflective practitioner in engineering design courses. In *2nd International Conference on Design Education (ConnectED2010)*. The University of New South Wales, 2010.

[28] M. Ryan. The teachable moment: The washington center internship program. *New Directions for Teaching and Learning*, 1988(35):39–47, 1988.

[29] M. Ryan. Reflexive writers: Re-thinking writing development and assessment in schools. *Assessing Writing*, 22:60–74, 10 2014.

[30] M. Ryan and M. Ryan. Theorising a model for teaching and assessing reflective learning in higher education. *Higher Education Research & Development*, 32(2):244–257, 2013.

[31] D. R. Sadler. Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144, 1989.

[32] J. Sandars. The use of reflection in medical education: Amee guide no. 44. *Medical Teacher*, 31(8):685–695, 1 2009.

[33] Á. Sándor. Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue française de linguistique appliquée*, 12(2):97–108, 2007.

[34] A. Schön Donald. The reflective practitioner: How professionals think in action, 1983.

[35] D. Simsek, S. Buckingham Shum, A. Sandor, A. De Liddo, and R. Ferguson. Xip dashboard: visual analytics from automated rhetorical parsing of scientific metadiscourse. 2013.

[36] P. A. Smith. Action learning and reflective practice in project environments that are related to leadership development. *Management Learning*, 32(1):31–48, 2001.

[37] F. Steier. Reflexivity and methodology: An ecological constructionism. In *Research and reflexivity*, page 257. London, 1991.

[38] J. Swales. *Genre analysis: English in academic and research settings*. Cambridge University Press, 1990.

[39] C. Tsingos, S. Bosnic-Anticevich, J. M. Lonie, and L. Smith. A model for assessing reflective practices in pharmacy education. *Am J Pharm Educ*, 79(8):124, 10 2015.

[40] C. Tsingos-Lucas, S. Bosnic-Anticevich, and L. Smith. A retrospective study on students' and teachers' perceptions of the reflective ability clinical assessment. *Am J Pharm Educ*, 80(6):101, 8 2016.

[41] L. Valli. Listening to other voices: A description of teacher reflection in the united states. *Peabody Journal of Education*, 72(1):67–88, 1 1997.

[42] M. van Manen. On the epistemology of reflective practice. *Teachers and Teaching*, 1(1):33–50, 3 1995.

[43] A. B. Warriner, V. Kuperman, and M. Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.

[44] F. K. Wong, D. Kember, L. Y. Chung, and L. Y. CertEd. Assessing the level of student reflection from reflective journals. *Journal of advanced nursing*, 22(1):48–57, 1995.