ABSTRACT
Reflective writing is an important learning task to help foster reflective practice, but even when assessed it is rarely analysed or critically reviewed due to its subjective and affective nature. We propose a process for capturing subjective and affective analytics based on the identification and recontextualisation of anomalous features within reflective text. We evaluate 2 human supervised trials of the process, and so demonstrate the potential for an automated Anomaly Recontextualisation process for Learning Analytics.

Categories and Subject Descriptors
K.3.1 [Computer Uses in Education]: Computer-assisted instruction; J.1 [Administrative Data Processing]; Education; H.1.2 [User/Machine Systems]: Human factors

Keywords
Reflective Text, Learning Analytics, Affective Computing

1. INTRODUCTION
Reflective writing is used by educators to help students develop the metacognitive capability required for effective reflective practice, an important dimension of continual improvement in many professions [6, 5, 13]. However, despite it’s educational value, reflective writing presents challenges. Firstly, it is difficult to assess [13], and can require a lot of reading time on the part of the educator. Secondly, the inherent lack of structure in reflective writing, its personal style, and variability in quality [6], makes it difficult to treat in the same way as other more structured written tasks, and thus an unlikely candidate for computational analysis.

Despite these challenges, the fact that when we read text we easily recognise features such as sarcasm, humour and emotional tone, suggests that features are there, and could possibly be discovered computationally. Although external factors like body language and eye contact can help with interpretation of spoken language, they are by no means always necessary as human readers have an ability to intuitively determine subtexts, and to postulate on their unstated meanings. In reviews of others’ writing we can easily understand phrases like “thinely veiled contempt”, “dripping with sarcasm” and “unbridled enthusiasm”, which remind us of the case with which we perceive the affective dimensions of an author’s perspective from their writing. When considering reflective writing we also make interpretations about the way an author feels about a topic, their degree of openness and honesty, the extent to which the author is expressing affective ideas like conflict, fear or uncertainty, or even the degree to which an author acknowledges a community or focuses on themselves.

Although computational tools for natural language processing (NLP), such as linguistic analysis [7] and topic modelling [1], have proven very effective for information retrieval, they have typically been less capable with the more subjective text characteristics such as emotion and affect. Sentiment analysis [12] techniques can be used to identify affect, however they tend to perform best with polarised data that is more uniform in sentiment. We would prefer more complete representations of these subjective features when analysing reflective text for learning analytics.

It is not easy to see how the standard computational approaches can be extended to more nuanced writing. The personal nature of reflective writing can involve the use of complex linguistic devices such as irony, which are particularly challenging for computational analysis. The statement “I’m spending my weekend marking assignments. I love it - can’t imagine doing anything else” is identified reasonably easily by a human reader as negative. This is because the human approach to anomalous information is to learn [15]. We tend to expand the repertoire of contexts which we draw upon to make sense of what we read. For example, when we contextualise a person’s words as humour, we don’t label them a liar for expressing obviously false statements.

Intriguingly, there is one exception to this typical human approach for accommodating anomalous data within different contexts. Sometimes we judge the data itself to be inaccurate, in which case the tendency is to dismiss the anomaly and keep the existing context [4]. This judgement of accuracy is much closer to what we have come to expect from computational processes, which can influence our use of them. Thus, while an anomaly in the computer world is usually a violation of specific rules or conditions, an anomaly in the human world is something out of context. This is the intuition that we seek to formalise.
2. ANOMALY RECONTEXTUALISATION

Our approach, which we call Anomaly Recontextualisation (AR), is based on accommodating rather than eliminating anomalous data. It is fundamentally a 2 step process. Step 1 requires the identification of an anomaly within a given context. Step 2 involves recontextualising the anomaly such that it is no longer anomalous within its new context.

To illustrate our approach, consider the following metaphor about a bird called Tux:

Context: Birds have a feature of flight which is associated with a feature wings.
Data 1: Tux has feature wings, and belongs to context bird.
Data 2: Tux doesn’t fly - lacks feature flight
Anomaly: Tux doesn’t fly even though Tux has wings, is a bird, and birds fly.
New data: Tux has additional feature swims
New context: Penguins have a feature swims and a feature wings, but lack feature flight.

Recontextualisation: Tux is a penguin!

Note that in the Tux illustration, we don’t modify the original context. The birds context with feature flight still contains a general truth for birds. By recontextualising anomalies, we retain a general truth which allows us to recognise that birds fly, but to also adopt exceptions for those that do not fit precisely into the feature set of our original context.

Anomalies are missing this feature but possess the first. Hence, our objective was to outline the necessary steps that, given an anomaly in one context, allow a new context to be created in which that anomaly is resolved, without modifying the original context. This expanded AR process is as follows:

1. Identify a context based on a feature set (e.g. feathers, wings, flight, for context birds). A complementary context (e.g. not birds) can be identified from the absence of the features.
2. Classify data based on a key feature (e.g. wings) to determine each element’s membership of the context; the complementary context; or undetermined.
3. Identify anomalies through the classification using a feature strongly related to the first (e.g. flight). Anomalies are missing this feature but possess the first.
4. Classify the data based on a non key feature that is significant to the set of anomalies (e.g. swims).
5. Recontextualise successfully classified anomalies with a new context based on the associated features (e.g. wings, swims, no flight).
6. Repeat 4 and 5 until either (i) all anomalies are recontextualised, or (ii) a set limit is reached.

The final output of the AR process is a number of contexts which provide high level information about our data, along with the ability for that data to be understood in varying ways. For example, if we started the above example with a context that had features of doesn’t fly and swims then we would could include whales, athletes and penguins. However, their membership in other contexts (e.g. ocean mammals, humans, birds) would allow us to not only differentiate between them, but also to understand their interrelationships.

3. APPLICATION

We applied the AR process to the analysis of reflective texts written by first year Bachelor of IT students during a group project that spanned about 3 weeks. The cohort was offered access to GoingOK\(^1\), a web application developed by one of us (Andrew Gibson) for recording personal reflections. At the conclusion of the project, students could download their reflections and submit them for assessment together with their project deliverables. GoingOK de-identifies the user’s reflective data so that it can be used for research purposes\(^2\). 82 students signed up to use GoingOK. Of these students, 24 recorded at least 1 reflection with reflective text during the period. Recorded reflections were of varying quality, and so echoed many of the qualities that make reflective text difficult to work with computationally.

\(^1\)www.goingok.com

\(^2\)QUT Ethics Approval No.: 1400000151
chose to trial the AR process with different aspects of the indication of similarity between reflections, as well as a sense of content of reflections. An LDA topic analysis provided a measure of the extent to which words were unique to individual reflections, or more generally used across the group as a whole.

As our primary objective was to analyse the reflective text, we wanted to minimise the number of reflections that were recorded without text. However, excluding all non-text reflections would not have provided a realistic sample of the data, so we selected all reflections for users where at least one of their reflections included 2 or more words. Thus, we still captured some zero word reflections, but the majority of the reflections included a text of at least 2 words. The resultant dataset included 57 records from 24 students.

We extracted a large range of features from the data set that could be used in our computational analysis. For each individual reflection we extracted features related to: reflection point, date, word length, word and sentence counts, parts of speech, word frequencies, and topics. We also extracted features for the group as a whole: Statistical analysis of reflection point, day of week, word lengths, word and sentence counts. We also collected frequency distributions of words, counts, ratios, and various parts of speech. Other features were collected by building topic models over the full collection. Using both individual and group features, we calculated a range of comparative measures such as the deviation from group mean for individual features, or the Kullback-Leibler (KL) divergence [9] for measuring the similarity or difference between distributions. Term Frequency / Inverse Document Frequency (TF-IDF) [14] also provided a measure of the extent to which words were unique to individual reflections, or more generally used across the collection. For topic modelling, we used Latent Dirichlet Allocation (LDA) [2], which we used to identify the topical relationships between reflections and groups.

Although not part of the reflective text itself, the (numeric) reflection point was included in our analysis as it provided a quantitative measure indicating the users self-perceived general state of well-being at the time of writing their reflections. This proved a useful point of reference external to the text itself, a user annotation of their own text. Additionally, we calculated the descriptive statistics on the reflection points of the whole group, which provided a group norm against which individual reflection points could be compared.

The majority of the features used in analysis were generated from the reflective text. These included word count, sentence count, words per sentence, verb ratio, and noun ratio. As with the reflection point, we calculated descriptive statistics across the group from individual reflective text features. Some features (including parts of speech, pronouns, words, word length, and punctuation) were recorded as a frequency distribution for each reflection, and the same for the group as a whole.

We included 2 types of semantic feature distributions. A TF-IDF analysis provided a measure of uniqueness in the content of reflections. An LDA topic analysis provided an indication of similarity between reflections, as well as a sense of which topics were most significant over the group.

With only a relatively small data set to work with, we chose to trial the AR process with different aspects of the same data. We hypothesised that an individual record in the dataset would belong to a context if it contained some characteristic of a selected feature, or to the complementary context if it had the absence or opposite characteristic of the same selected feature. We trialled the process by looking at 2 aspects that were subjective in nature, relevant to LA, and pedagogically useful. These were: Progress satisfaction, and self/others focus.

Where possible we utilised computational tools to implement the various steps of the AR process (notably Factorie [10] for POS tagging and LDA, and Scala [11] for other analysis). However, due to time constraints, human supervision was involved and this is outlined below.

### 3.1 Trial 1: Progress satisfaction

The first aspect we analysed was progress satisfaction with the reflection point as the initial key feature. We expected that a positive polarity of the reflection point would indicate satisfaction with progress, and that a negative polarity of the reflection point would indicate a dissatisfaction.

A classifying function evaluated each reflection point in terms of its deviation from the mean reflection point for the group as a whole. This value was recorded in the results as a ratio to the standard deviation. The function classified all reflections of 0 or more deviation as positive, and vice versa. This yielded 29 reflections classified as positive, and 36 negative.

We took a supervised approach to the identification of anomalies, and reviewed the negatively classified results first by reading the text and assessing it for general negative sentiment related to progress. This could be implemented computationally with sentiment analysis software. We identified 16 anomalies in the negative results, which we further classified into 3 groups: one with 3 reflections which had insufficient text for a judgement, and the second had 11 reflections that were generally positive but held reflection points very close to the mean. The third group had 2 reflections that appeared to be significant anomalies. Because of the large number of anomalies with points close to the mean, we determined that the classifier function should work from a lower threshold, a slight negative deviation from the mean. We calculated that a −0.2865 deviation would catch all of these anomalies. Significantly, we noted that this new point would not erroneously classify negative reflections as positive. We re-ran the analysis with a split set at 30% of the standard deviation less than the mean. This resulted in an effective split point for the data set at 45.65. The re-run of the analysis resulted in 23 negative reflections and 42 positive reflections. This process of finding the optimum split point could be implemented computationally using a simple error minimisation algorithm.

The positive results were almost universally reflective of progress satisfaction related remarks in the text. Of the 42 reflections classified as positive, we identified 5 anomalies. Four of these were spurious: A duplicate record, one with no text, and 2 that contained irrelevant text (e.g. “lazy holidays...”), leaving only one genuine anomaly. This text was a more complex mix of negative tone, indications of progress being made, but dissatisfaction with a group member’s contribution. Despite the negative undertones, it had reflection point of 76. A portion of the text read: So we had the group meeting today. [name] didn’t make it and it’s been nearly a month since we’ve all even seen him...To be honest, this...
group meeting was pretty pointless. . . . [user: hesnav]

Of the 23 negative reflections, we identified 4 anomalies. Two with no text, one a sarcastic expression: “new place to share everything...great!” (reflection point of 33), and the other was not significantly positive, but did indicate satisfaction with progress despite having a reflection point of 27. The text read: ‘I’m happy with the app functions I created in the last assignment, but I presented in a very average fashion. I would have liked to learn more about presenting in [unit] so that I could have expressed that knowledge through my presentation. [user: cuzgus]

The initial analysis of the data resulted in 2 contexts which could be labeled: satisfied with progress and dissatisfied with progress. Of the 2 identified significant anomalies, both could be recontextualised with a label such as mixed feelings about progress. The other anomalies could be recontextualised as no relevant information. The final results for this trial are summarised in Table 1.

Table 1: Results for satisfaction with progress

<table>
<thead>
<tr>
<th>Final Context</th>
<th>Feature/s</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfied with progress</td>
<td>Positive reflection point deviation</td>
<td>37</td>
</tr>
<tr>
<td>Dissatisfied with progress</td>
<td>Negative reflection point deviation</td>
<td>19</td>
</tr>
<tr>
<td>Mixed feelings about progress</td>
<td>Sentiment, reflection point deviation</td>
<td>2</td>
</tr>
<tr>
<td>No relevant information</td>
<td>Reflection point deviation, empty or off-topic</td>
<td>7</td>
</tr>
</tbody>
</table>

3.2 Trial 2: Self-others balance

The second trial analysed for a focus on self or others, the balance between students focusing on themselves or including others in their reflection, with the hypothesis that students who focused solely on themselves would be less likely to perform well as a part of a team. Influenced by Campbell and Pennebaker’s work on Pronouns [3], we used pronoun distributions as the key feature for initial classification. We assumed that a focus on self would result in the use of less third person plural pronouns and more first person singular pronouns, and that a focus on others would reflect in greater use of third person plural and less of first person singular pronouns.

Of the total of 65 reflections, 21 of them had no pronoun distribution due to no pronouns being detected in the text. For expediency, we removed these from the data set prior to applying the classifying function, as it did not structurally alter the trial. However, in future these could be analysed for inferred pronoun usage. Reflections written in a text message style, may infer a personal pronoun without actually stating it. e.g. “Just finished assignment” as opposed to “I’ve just finished my assignment”.

Our initial classification function compared the reflections to the pronoun distribution for the whole group. We used KL divergence to assess the difference between individual reflection pronoun distributions and that of the group distribution. We assumed the group distribution would be an appropriate balance between self and others. This proved correct as the individual reflection with the lowest KL divergence (k = 0.0804) contained a good mix of self and others: “I feel like i need to start making it a habit. Otherwise I’m going to end up doing it all on one day . . . Anyway as far as the group meeting went we discussed how our observations went and we came to an agreement . . .

With supervision deciding the break point, the reflections were separated at a KL divergence of 0.3. As with the first trial, error minimisation algorithms would allow us to complete this computationally. The selected split point resulted in 12 balanced reflections, and 32 reflections that were biased to either self or to others.

Of the 12 balanced reflections, we identified 3 anomalies. As our primary objective was to obtain analytics about the learner, we decided to resolve these anomalies by taking an student focused approach, looking for the extent of balance in the student’s other reflections. If other balanced reflections were found, we aggregated the results and re-calculated the KL divergence. If not, we added the reflection to the unbalanced context. Only 1 reflection was by someone who had written another balanced reflection, so we recalculated the KL divergence for the aggregate and confirmed that was just over the split point (k = 3.1161). We classified the 3 remaining reflections as unbalanced requiring further resolution. The final number of balanced reflections was 9. Despite the manual work in this stage of the process, much if not all of it could be implemented computationally through an iterative process of expanding the net of analysed reflections based on the writer.

Initially the unbalanced reflections numbered 32, but the 3 balanced anomalies that were reclassified as unbalanced resulted in a total of 35. We treated them all as anomalies as we wanted to know more about their bias. To do this, we applied another classifying function to classify them as self focused or others focused. To determine self focus, we utilised a feature based on the first person singular value of the distribution. Given that the imbalance in the 3 anomalies separated from the balanced context was primarily related to the first person singular value, we determined that the split point for this function should be the minimum value of these 3 reflections. This would position them into the self focused group by default. Running this function resulted in 25 reflections classified as self focus, 9 as others focus, and 1 duplicate.

Of the 25 self focused reflections, no further anomalies were detected so these were recontextualised as self focused. A reflection typical of this context is: I just finished a presentation, so I feel great! [user: cobkes] Of the 9 others focused reflections, 1 reflection had no self or others focus and was manually recontextualised as non focus, leaving 8 that were recontextualised as others focused. A reflection typical of this context included text like: ... We are running behind on creating a roleplay... Also we haven’t heard from two group members in around two weeks, which is re-
ally hindering us in terms of how much we can do. [user: rutkod].

The final contexts after all recontextualisation had concluded are summarised in Table 2.

### 3.3 Findings

Our 2 trials demonstrated that the AR process has some benefits over analysis techniques which take a more binary classification approach with subjective data. The information found through the AR process could be useful to Learning analytics due to its affective nature and learner focus. Rather than software that indicates that a student is generally positive or negative, we have shown the potential for finding students with mixed feelings about their progress, or students that may not be working well in a group. And this has been done with the same sparse data source which was written with neither of the analysis objectives known.

Throughout our documentation of the trials we identified where human supervision was utilised and made suggestions for the way that these steps could be automated. However, a fully automated implementation of AR would be a challenging endeavour. One of the most significant challenges would be to develop a suite of good context-feature models that can be implemented using the AR process. Such a system is likely to require the use of machine learning techniques that learn which models to apply in which circumstances based on historical data that has been evaluated by a human expert. We anticipate that it is likely that human intervention is always likely to be required to some extent in both a data-wrangling and sensemaking capacity.

An automated AR process would also require better data cleaning prior to initial classification. Although duplicated records and records without text did not significantly impact this study, these types of issues could have been a problem with a larger data set, and would need to be mitigated against for an automated approach. Other pre-processing like the pronoun inference mentioned above would also need to be implemented.

During our application of AR, we also noted that changes to a learner’s reflections over time is likely to be useful. Unfortunately, the temporal dimension of this study was very limited (around 3 weeks) and only a few students completed a series of reflections. However, an automated system analysing data over semesters, or even years, would benefit from models which included temporal features.

Unsurprisingly, the learner is central to the AR process and we see a great deal of promise in including user modelling in future work. In particular, there are opportunities to explore the interactions between users in a group scenario, which could make a significant contribution to LA in the area of group learning.

### 4. CONCLUSION

The subjective and affective features of reflective text can provide insights on students and their progress that are unique to this type of writing. Our proposal of the AR process as a way of extracting these features from reflective texts showed potential when applied in the context of our study. We believe that Anomaly Recontextualisation holds promise for the enrichment of Learning Analytics and is worthy of future development.

### 5. REFERENCES