Towards the Discovery of Learner Metacognition from Reflective Writing

Andrew Gibson

Queensland University of Technology, Australia

University of Technology Sydney, Australia

andrew@andrewresearch.net

Kirsty Kitto

Queensland University of Technology, Australia

Peter Bruza

Queensland University of Technology, Australia

ABSTRACT: Modern society demands renewed attention on the competencies required to best equip students for a dynamic and uncertain future. We present exploratory work based on the premise that metacognitive and reflective competencies are essential for this task. Bringing the concepts of metacognition and reflection together into a conceptual model within which we conceived of them as both a set of similar features, and as a spectrum ranging from the unconscious inner-self through to the conscious, external, social self. This model was used to guide exploratory computational analysis of 6,090 instances of reflective writing authored by undergraduate students. We found the conceptual model useful in informing the computational analysis, which in turn showed potential for automating the discovery of metacognitive activity in reflective writing, an approach that holds promise for the generation of formative feedback for students as they work towards developing core 21st century competencies.

Keywords: Metacognition, reflection, reflective writing analytics, computational analysis, natural language processing, learning analytics, 21st century competencies

1   INTRODUCTION

Continuous rapid change is a dominant feature of modern society. Technology is rapidly evolving, and this is leading to the demise of entire employment sectors, and the creation of completely new ones. Within this context, we can hardly anticipate what skills, knowledge, and experience the students of our formal education systems are going to need to know upon graduation, let alone ten years into the future. Learning how to acquire new ideas, skills and approaches for themselves is an important strategy that will help our students to succeed in this age of uncertainty. Learning to learn, and the related concepts of self-directed learning, self-regulated learning, and independent learning, are certainly not recent pedagogical ideas (Herber & Nelson Herber, 1987; Paul, 1990), but the general approach has, over time, gained greater prominence (Cornford, 2002; Takanishi, 2015; Wilson, 2015) as we have witnessed significant changes in

various domains of knowledge (CEDA, 2015). While changes in the body of knowledge have always been with us, it seems that they have never been as rapid as they are now, and this current acceleration of change appears to be increasing rather than abating. We cannot escape this context, and within it we must aim to educate our students in a way that prepares them for the future, equipping them with adaptive capabilities that can flourish in this fluid environment.

Historically, formal education has focused on knowledge of facts, and while foundational content is still seen as critical, entire domains are starting to recognize the need to shift educational focus from presenting large amounts of content, towards instilling a kind of meta-knowledge; that is, the capability for understanding what knowledge may be needed for a situation and when to apply it, as well as the ability to find the identified requisite knowledge. Development of autonomy in this respect is not possible for the learner without them drawing upon metacognitive abilities to evaluate their current cognitive position, and formulate strategies for advancing it. Despite the changes in formal education to accommodate this paradigm shift, further work is required, and intelligent tools are needed to support it.

This modern educational context has provided an underlying motivation for the work presented here. In this paper, we demonstrate potential for computational analysis of reflective writing (Reflective Writing Analytics) as a means of discovering evidence of metacognitive activity in the reflective writing of a learner. As we shall see in Section 2, the distinction between metacognition and reflection is blurred. However, the literature suggests significant relationships between them, and that the analysis of written reflection may allow for the discovery of metacognitive activity on the part of the learner. An empirical study was designed to explore this intuition.

The research was conducted in two parts, the first being the development of a conceptual model drawn from the metacognition and reflection literature (Section 2), and the second being an exploration of the model's application to the computational analysis of Reflective Writing Analytics (Section 3). We conclude with an examination of the implications of this study to the field of learning analytics (Section 4).

2   METACOGNITION AND REFLECTION

In the century since John Dewey introduced the concept of reflection in education (Dewey, 1916), the educational community has increasingly considered reflection an important part of learning. This has become more focused since Flavell (1976) defined the term metacognition, which encouraged educators to embrace the significance of thinking about thinking. In a pedagogical context, both reflection and metacognition regularly co-occur, suggesting a strong link between them (Schön, 1983; Moon, 1999). Indeed Boud, Keogh, and Walker (1985, p. 141) state that "Reflection is thus 'meta-thinking'" and Sanders (2009, p. 688) asserts that "reflection is a metacognitive process." This link holds significance for our work, since while metacognition can be hidden from view, reflection can be made explicit in students' reflective writing. Our intuition is that exploiting this link between reflection and metacognition will provide a means for discovering evidence of metacognition in the reflective writing of the learner.

In this section, we examine some of the key literature on metacognition and reflection, considering each in turn as individual concepts. We then identify key similarities and differences between them, and propose a conceptual model that encapsulates these features.

2.1 Metacognition

John Flavell originally defined the term metacognition as "the active monitoring and consequent regulation and orchestration" of information processing activities, "usually in service of some concrete goal or objective" (1976, p. 232). Flavell's (1979) definition matured into a model of cognitive monitoring that comprised metacognitive knowledge, metacognitive experiences, goals, and actions. He noted that these four components continually interact throughout the cognitive process. These components provide an indication of how metacognition may present if it were made visible in student writing.

Studies conducted subsequent to Flavell's work have identified two types of metacognition: regulatory strategies, and strategic knowledge (Quirk, 2006). Regulatory strategies align with the monitoring and regulation aspects of Flavell's definition, whereas strategic knowledge incorporates the orchestration and goal aspects. This apparently dual nature of metacognition is also identified by others, but not always along the same lines. For example, Amsel et al. (2008) examined the relationship of metacognition to scientific reasoning, explicitly taking a dual process perspective: "In dual-process theory, metacognitive skills function to regulate conflicts between analytically and experientially based responses" (Amsel et al., 2008, p. 454). Amsel et al. defines the experiential as heuristic and automatic, comparing it to the analytical, which is formal and systematic. These dimensions appear to accord somewhat with Flavell's 1979 model with metacognition experience and knowledge; however, it seems likely that Flavell would not take Amsel et al.'s view of metacognition as a moderating skill that overrides the experiential, for Amsel et al.'s definition is specific to a style of cognition that presents as scientific reasoning, and may not translate to other forms of cognition. Lehmann, Hähnlein, and Ifenthaler (2014) also promoted a dual view of metacognition within a wider model of self-regulated learning. They note that metacognition can include a structural component that is relatively stable, including task knowledge and knowledge of cognition, and a process component that is more immediate to a particular situation. This component includes planning, monitoring, and evaluation. We note the similarities between these components and Flavell's goals, actions, and regulatory strategies.

The view that metacognition is a trainable skill, and that people should be encouraged to be more metacognitive, has existed for as long as the term itself. Indeed Flavell (1979) asserted that cognitive monitoring is trainable, and that it should be promoted more widely: "I am absolutely convinced that there is, overall, far too little rather than enough or too much cognitive monitoring in this world" (p. 910). However, we should be careful to avoid the conclusion that more is better, as excessive metacognition is evident in people who are pathologically obsessive, and when non-selective, cognitive monitoring can become rumination on negative outcomes, leading to anxiety related problems (Quirk, 2006). This suggests that discovery of metacognition should be more concerned about its qualities rather than the quantity in which it occurs.

Metacognition has often been associated with intentional and consciously accessible thought; however, this does not represent a complete picture, and ignores cognitive activity that occurs without awareness. "Indeed, the conscious character of metacognition was challenged by early research on meta-memory showing that people monitor and control their cognition without being consciously aware that they do so" (Efklides, 2008, p. 281). Koriat (2000) suggested that metacognition actually possesses two faces: an implicit automatic mode and an explicit controlled mode. In fact, he proposes that metacognition, particularly metacognitive experience, mediates between two layers of consciousness stating that "they serve to interface between implicit-unconscious-automatic processes on the one hand, and explicit-conscious-controlled processes on the other" (p. 152). While education is typically focused on the explicit controlled mode of metacognition, it is the existence of the implicit automatic mode of metacognition that we suggest differentiates it from the activity of reflection.

2.2   Reflection

Somewhat like metacognition, reflection tends to be a concept recalcitrant to crisp definition. Moon (1999) noted that a single definition of reflection is elusive, despite attention to the topic over many years. She observed that "reflection seems to be a form of mental processing with a purpose and/or an anticipated outcome that is applied to relatively complicated or unstructured ideas for which there is not an obvious solution" (Moon, 1999, p. 98). This mental processing aspect appears to be such a common human activity that we are in danger of missing the importance of the concept altogether. Indeed, Boud et al. (1985) identified familiarity as a contributor to our lack of understanding about reflection:

> The activity of reflection is so familiar that, as teachers or trainers, we often overlook it in formal learning settings, and make assumptions about the fact that not only is it occurring, but it is occurring effectively for everyone in the group. It is easy to neglect as it is something which we cannot directly observe and which is unique to each learner. (p. 8)

Reflection can also be understood differently in different contexts. In this paper, we are particularly concerned with reflection for learning within an educational context. However it is worth keeping in mind that the term is also commonly used with respect to learning within a professional context (i.e., reflective practice), and in a personal sense (i.e., personal reflection). These different applications of reflection are not necessarily independent from each other, and share many of the same features. As a result, research directed at one application is often referred to in the context of another. For example, Schön's (1983) work, which focused on professional reflective practice, has often been used as a basis for understanding reflection within a formal education context.

Further complicating the concept, reflection as a mental activity can be easily conflated with reflective writing — a task designed to elicit mental reflective activity, or in the case of assessment, document mental reflective activity. For the purposes of this study, we considered reflection to be a type of mental activity that can be evidenced by reflective writing.

Despite the ambiguity around the term, Boud et al. (1985) identified three features of reflection that they clearly observed. First, as reflection is a personal activity, the learner is central and in control, revealing only what they wish about what they have reflected upon. Second, reflection is goal orientated and purposive. Third, reflection is a complex process with an affective dimension that interacts with other cognitive processes. These three aspects of reflection correlate well with the regulation, planning, and experiential qualities identified in metacognition.

However, unlike the literature on metacognition, there is a growing body of literature that highlights the importance of a social dimension in reflection. Indeed some would argue that reflection cannot be fully understood without examining it within a social context. Ryan (2011) proposes a multi-level model — strongly linked to the context of the author — for teaching and assessing academic reflection. This work views reflection as a socially situated activity involving a deep examination of self in relation to the social context, and the intention to take action on the basis of that examination.

Action is also a key aspect of Mezirow's (1990) work on transformative learning. Mezirow (1990) proposed a high level of reflection called critical reflection, which involves learners critiquing their pre-existing knowledge based on new information, holding "biases in abeyance, and, through a critical review of the evidence and arguments, make[ing] a determination about the justifiability of the expressed idea whose meaning is contested" (p. 10). This work certainly highlighted the importance of reflection in learning, a point that Ryan (2013) extended beyond improving one's own understanding to the very sustenance of the learning process. She argued "that students can and should be taught how to reflect in deep, critical and transformative ways to engender sustainable learning practices" (p. 145). Here we see some similarity between the application of reflection to learning and Flavell's assertion that metacognitive monitoring and control is trainable.

Also relating reflection and learning is the work of Zimmerman (2002) who applied reflection specifically to the process of creating the self-regulated learner. He stated that "these learners monitor their behaviour in terms of their goals and self-reflect on their increasing effectiveness" (p. 66). Here we see in "monitoring" and "goals" that again there is alignment between conceptions of reflection and metacognition further supporting our intuition that the two are strongly related.

Zimmerman (2002) conceptualized self-reflection in terms of self-judgment and self-reaction, with self-judgment involving the comparison of the learners' own understanding of their performance against a standard; and self-reaction involving an affective element (e.g., self-satisfaction) and an adaptive element (e.g., adjustment of strategy). This too displays similarities with metacognition, with self-judgment and self-reaction relating nicely to the judgment and affective dimensions of metacognitive experience.

In relating the various understandings of reflection evident in the above literature, we can see that reflection is usually conceived of as a deliberate process with a trigger of some sort and an underlying goal or objective. It tends to be personal in nature, engaging feelings and emotions; however, it is also socially situated and therefore not solely an internal activity. It also tends to include a level of monitoring of the

particular situation in question, and regulation of behaviour towards achievement of the desired goal. There is a sense that true reflection has depth and is a difficult process, but one that can be transformative.

2.3   A Model for Metacognition and Reflection

We have already identified some significant similarities between features of metacognition and reflection. These similar features provide a foundation for a model, but we must remain conscious of the differences between these two concepts. The most notable difference is the lack of an implicit automated version of reflection, and the absence of a socially situated type of metacognition. We found some resolution to these differences in Efklides' (2008) multilevel model of metacognition. She proposed a model that comprises three levels: a non-conscious object level, the meta level associated with personal awareness, and the meta-meta level associated with social awareness. Significantly, this approach not only relates the non-conscious and conscious aspects of metacognition, but also introduces the idea of reflection together with metacognition between the personal awareness and social levels.
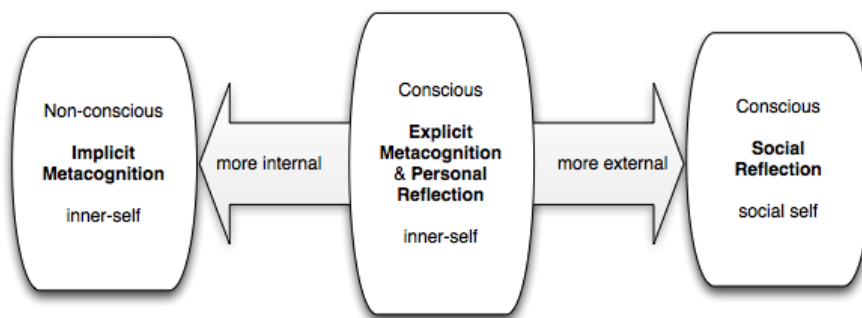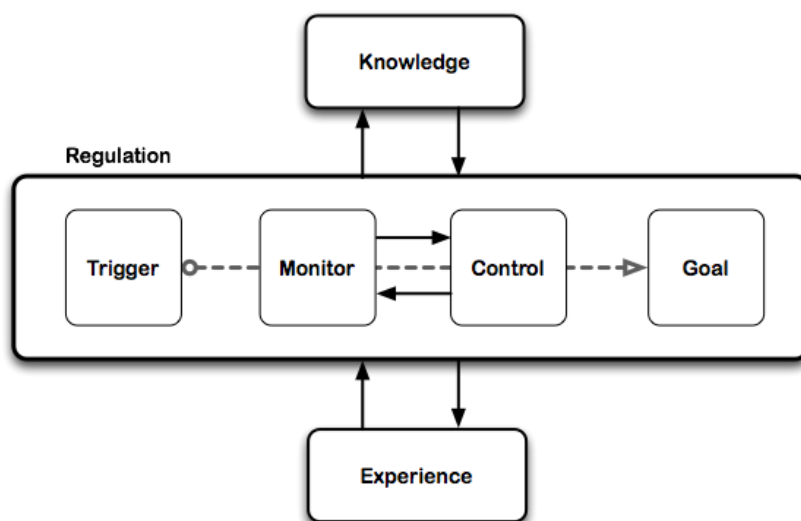

Figure 1: A spectrum view of metacognition and reflection.

We conceptualized this difference in the relationship between metacognition and reflection as a spectrum, with the internal inner-self on one end and the external social-self on the other (Figure 1). For metacognition, the left side of the spectrum represents the implicit, automated, non-conscious mode, while the centre includes the explicit, conscious, controlled mode (Koriat, 2000). For reflection, the right side of the spectrum represents the external, socially situated dimension, and the centre includes the personal, internal, but conscious aspect. We suggest that this understanding of the relationship between metacognition and reflection not only caters to both implicit–internal and explicit–external understandings of these activities, but also provides a way of relating the explicit metacognition to internal reflection in the centre of the spectrum. In this area, the terms could be used interchangeably, while on the left and right extremes the terms have quite different meanings. We propose this model as a way of embracing the varying definitions of metacognition and reflection found in the literature while providing additional clarity on the use of each term.

However, the spectrum view provided in Figure 1 does not clearly specify the ways in which metacognition and reflection are alike. To address this, we modelled similar features as interrelated components

27

(Figure 2). Reducing a whole idea into component parts can be problematic, particularly when trying to understand how the components relate to the more general spectrum model. The common features of metacognition and reflection do not explain the changing nature of the two activities as viewed on different parts of the proposed spectrum in Figure 1, and while the spectrum highlights the general nature of both activities, it does not identify the common features. So our two views of the relationship between metacognition and reflection need to be taken together and understood as two perspectives of the one phenomena. Selecting a suitable level of granularity for the component model was important. Too few components resulted in the model lacking explanatory power, and too many components resulted in some features appearing falsely significant.

Figure 2: A common feature view of metacognition and reflection.

Our final component model (Figure 2) is the result of a number of iterations, with the most significant change being the integration of the trigger and goal components into the main component of regulation together with monitor and control. We attempted to address the granularity issue by retaining these elements as subcomponents to be mapped in the analysis, but considered them together with monitoring and control as one level of importance with respect to the overall model.

The model is designed with three primary components, each labelled according to the language of the metacognition literature. The model's core component is regulation, which at its centre involves a monitor and control loop. Because of the interrelated nature of monitoring and controlling, we tended to conceptualize these as one loop, rather than two independent components. We note that the regulation component also contains initiation and objective sub-components, respectively labelled as trigger and goal. Regulation is continually interacting with the other two components: Knowledge, which represents metacognitive knowledge (the storage of strategies and decision information), and Experience, which represents metacognitive experience (feelings and affective contribution).

Table 1: Model sub-components described in terms of metacognition and reflection.

	Metacognition	Reflection
Trigger (Regulation)	A conscious or unconscious cognitive event, particularly a problem or incongruence.	A conscious cognitive event or an external event that needs improvement or modification.
Monitor (Regulation)	Monitoring of cognitive processes, both consciously and unconsciously.	Conscious thinking about mental processes or external behaviours.
Control (Regulation)	Utilizing pre-learned strategies to control cognitive processes.	Taking action to modify mental processes, or external behaviours based on input information.
Goal (Regulation)	Resolution of the trigger problem or dissonance.	Successful modification or improvement of trigger event.
Knowledge	Memory dedicated to storing metacognitive knowledge in particular strategies and their efficacy. Used in the monitor–control loop.	Either internal memory or external recording of thoughts that can be used as necessary.
Experience	Affective impact on monitor–control loop. Assists with strategy formation and evaluation.	Emotional contribution to monitoring and control. Assists with establishing personal value and significance.

We note that the model as visualized could infer causal relationships between various components and that there is therefore a sense in which a temporal dimension to the model can be inferred. This was not our intention as the components are continually interacting with each other throughout metacognitive and reflective activity. Also, as we adopted metacognition inspired labels for the component view, the relationship between the model and reflection can be less obvious. To assist with these links, Table 1 draws attention to the applicability of the model for both reflection and metacognition.

2.4 The Use of Reflective Writing

An essential point is that the proposed model centres on the cognition of the learner, and therefore cannot be analyzed directly. An interface must be provided between this conceptual model and any learning analytics that we might hope to perform. In this study, we have used reflective writing as this interface. Reflective writing is an existing, well-accepted learning activity, enabling this type of study without imposing a non-related task on the learning process. We acknowledge, of course, that other means of analyzing metacognition exist (Koriat, 2000).

Within the modern educational context introduced in Section 1, the requirement to move towards teaching skills required for lifelong learning has resulted in increasing attention on reflective writing. Reflective writing has the potential to develop metacognitive and reflective capabilities in their own right (Reidsema, Goldsmith, & Mort, 2010), and so here we attempted to extract analytics specifically directed

towards the analysis of metacognitive activity as it arises in reflective writing. Although, this might appear a reasonable approach, we must be careful when adopting it within a formal education setting. The requirement to measure what has been learned in order to justify the importance of the learning experience can have a significant negative impact on reflective writing. "Assessment involves presenting one's best work, whereas reflection involves uncertain questioning, self-criticism, exploring, trying out ideas and acknowledging the messy nature of reality" (Wharton, 2012, p. 490). Further complicating the assessment of reflective writing, Reidsema et al. (2010) note that it allows students to "engage with their beliefs, values, uncertainties, desires and questions and to clarify what they know and more importantly, do not know about a situation" (p. 10). This can place reflective writing tasks in direct conflict with summative assessment, resulting in a conflict of paradigms between an institutional need for a positivist, product-driven perspective, as opposed to the learner's needs for a constructivist process-driven approach (Ross, 2011). This is often seen when there is a tendency to focus "on what has been learned rather than on how it has been learned, and the emphasis is on improving the reflective writing style rather than on learning about learning (metacognition)" (Mair, 2012, p. 150).

Our approach avoids these issues by focusing on the discovery of metacognitive activity through a learner's writing without requiring the use of summative assessment. This approach could potentially provide automated feedback derived from the analytics to the learner. Using this approach, metacognitive activity and reflection can be encouraged directly within the learning process, rather than via assessment.

3   REFLECTIVE WRITING ANALYTICS

We consider Reflective Writing Analytics (RWA) to be the analysis of reflective writing and reporting of resultant information about the writers and their contexts. The computational analysis that follows is RWA for the discovery of metacognitive activity on the part of learners.

3.1   Data

The research presented here involved the analysis of 6,090 student reflections captured progressively over a semester using a web application called GoingOK.1 Six hundred and fifty-seven students from three different Bachelor's Degree programs participated in the study: a third-year Environmental Engineering unit (DS-E); a first-year Science unit (DS-S); and a first-year Information Technology unit (DS-I). For each unit, the use of the software and participation in the study was optional, and the number of students who wrote reflections using the web application is summarized in Table 2. In each case, reflective writing was a requirement of the unit; however, the nature of this requirement and the extent to which it was assessed, differed markedly. All data in this study is drawn from reflections written using the web application.

---

[1] GoingOK was written by Andrew Gibson as part of a QUT Education research project that collected the reflections of early career teachers as they transitioned from student and developed their professional identities during their first year of teaching (http://GoingOK.org/)

Table 2: Reflective writing data set sizes*

	Full Data Set	**DS-E**	**DS-S**	**DS-I**
Reflections	N=6,090	188	740	5,162
Authors	657	35	145	477
Mean refs/author	9.3	5.4	5.1	10.8

* "Reflection" is a single entry in the web application, "authors" are students, and the "mean reflections per author (refs/author)" is the average number of reflections written by the students over the course of the semester.

3.2   Computational Analysis

The computational analysis was governed by the conceptual model (Section 2), with particular emphasis on the key features identified in the component view. Our aim was to explore the extent to which these conceptual features may correspond to lexical and structural features in the reflective writing. Our choice of features was based on a wide range of previous research identifying relationships between text features and human factors. In particular, we drew on Pennebaker and Chung's (2011) work on pronouns, and Ryan's (2011) work on linguistic features in academic reflective writing, along with a range of other work (Wharton, 2012; Tang & John, 1999; Reidsema et al., 2010; Ullmann, Wild, & Scott, 2012).

We linked the conceptual model to the reflective text indicators via an algorithm that progressively mapped low-level grammatical features in the writing (posTags) through to higher level annotations (metaTags), which we derived from the model. The algorithm that we developed comprised four levels:

- Part of speech (POS) tagging of a sentence (posTags)
- Matching of POS tag patterns to identify key phrases (phraseTags)
- Matching of phrase patterns to identify potential for annotation (subTags)
- Filtering of matched phrase patterns to select final related model component (metaTags

Table 3: POS patterns for the selection of phraseTags

Pattern	Corresponding phraseTags
possessive pronoun followed by any adjectives and nouns	selfPossessive (e.g., my team) groupPossessive (e.g., our group) othersPossessive (e.g., her project)
pronoun followed by any verbs, adverbs, conjunctions, and prepositions	consider (e.g., we decided to go) anticipate (e.g., we needed) emotive (e.g., i am coming along really well) generalPronounVerb (e.g., we had)
preposition followed by repetition of any POS	compare (e.g., like we are) temporal (e.g., after getting) pertains (e.g., with blood donations) manner (e.g., without distractions) outcome (e.g., from our practice pitch last week) generalPreposition (e.g., at hand)
modal followed by any verb or adverb	definite (e.g., can trust) possible (e.g., would be focusing)
personal pronoun followed by any POS and ending with a personal pronoun	selfReflexive (e.g., i feel like i) and groupReflexive (e.g., we timed ourselves)

The algorithm was implemented in Scala.[2] POS Tagging was implemented using the Factorie framework (McCallum, Rohanimanesh, Wick, & Schultz, 2008) with a pre-trained parser model based on the OntoNotes English corpus. While this is not reflective writing specific, it provides state of the art accuracy in POS tagging and has proved effective for the purposes of exploring the potential of the approach.

Informed by both the literature mentioned above and exploratory trials on the data, we selected 17 phrase tag patterns that captured a range of potentially meaningful phrases. These were based on the 5 general POS patterns listed in Table 3.The final phraseTags were obtained by filtering the POS pattern matching results by a lexicon for each phraseTag.3 The filtering lexicons were derived by manually identifying phrases from amongst the results from the POS pattern matching that showed potential alignment with the model. The pronoun–verb and preposition matching also have general tags, which included all phrases not specifically filtered. To tune the algorithm, the lists of general phrases can be searched for any important phrases missed by the filter.

A second matching pattern selects subTags based on the phraseTag pattern, and then a final matching pattern combines these subTags to provide the final metaTag that links with the model. These patterns were selected using two criteria: 1) the indicative meaning of the phrase, and 2) the extent to which the

---

[2] Example code can be found at http://nlytx.io/2016/metacognition
[3] The full filtering list can be found in the code online at http://nlytx.io/2016/metacognition

phrase needed to be combined with other phrases in order to support the corresponding metaTag. The final choice of patterns is outlined in Table 4.

Table 4: Relationship between metaTags, subTags, and phraseTags

metaTag	subTag	Phrase Tag Pattern
Regulation monitorControl AND (trigger OR goal)	Trigger	outcome
	Monitor and Control	temporal OR (pertains AND consider)
	Goal	anticipate OR definite OR possible
Knowledge		selfPossessive OR compare OR manner
Experience		emotive OR selfReflexive

The progressive layering approach that we used in the algorithm was beneficial in two ways. First, it enabled us to maintain connections between the phrases in the text and the components in the model by providing clear links between each layer. Second, the links enabled us to trace inaccuracies in the analytics by looking at the underlying patterns. However, these benefits come at a cost, and evaluating each layer can be very time consuming, particularly with large amounts of data. We see this approach as a first step, allowing us to gain an indication of how well the model and the algorithm work towards the discovery of metacognitive activity from the learner's writing. An example of the application of our mapping algorithm to a sentence is illustrated in Table 5, and further examples are provided in the Appendix.

Table 5: Example derivation of metaTag tags from a sentence.

Text	On seeing	the sheer amount	of it	,	I think	we		I have	to split up the work.
posTag	IN VBG	DT JJ NN	IN PRP	,	PRP VBP	PRP		MD VB	TO VB RP DT NN.
match	start(IN) & repeat(PRP, JJ, NN, VB, MD, PDT, W, TO)				start(PRP) & repeat(RB, VB, CC, IN)			start(MD) & repeat(RB, VB)	
filter	pertains(startsWithAny(across, to, of, that, on, among, about, under, over, with, within, around,				consider(containsAny(feel, felt, seem, think, realise, being, thought, decide, know, appeared, learn, experience,			definite(startsWith Any(will, 'll, ca))	
phraseT	pertains				consider			definite	
match	temporal OR (pertains AND consider)							anticipate OR definite OR possible	
subTag	monitorControl							goal	
metaTa	regulation								

3.3 Early Decisions

A number of key decisions were made during the computational analysis. These involved questions on how we approached the raw data, and also more importantly how we would approach the granularity of the results that arose from the algorithm.

The web application used for collecting the reflective writing (GoingOK) includes temporal information in the form of timestamps. An early decision regarding the data was to ignore this temporal information. Although the timestamp data could potentially be useful in showing the development of metacognitive activity over time, we felt that utilizing it for this project would add unnecessary complexity to the analysis without providing significant value to addressing the research questions. Thus, our attention was focused on evidence of metacognitive activity rather than on development of that activity.

Initially, we had also considered combining all reflections for a single author into one document. This would have significantly reduced the number of reflections (1 per author) and for authors with very brief reflections, it would have potentially provided a higher probability for the detection of features. However it became apparent from early observations of the data, that the model we were seeking was evident in single individual reflections (even short reflections), and that we gained more information about the author's metacognitive activity by keeping the reflections separate for analysis, and then considering the resultant analysis as a whole with respect to each author, such as, for example, the frequency of metacognitive activity.

3.4 Categorization of Metacognitive Activity

As this work was exploratory in nature, we took a cautious approach with respect to the granularity of the analysis. We settled on three categories of reflective writing: 1) reflections that showed significant evidence of metacognitive activity, termed "strong" reflections; 2) reflections that showed very little or no evidence of metacognitive activity, termed "weak" reflections, and 3) reflections that were neither clearly strong nor weak, termed "undetermined." Originally, we anticipated that distinctions between these groups would be reasonably easy to make, however, the variability in the data meant that decisions regarding their categorization were more complex than we first thought.

We undertook a series of trials to determine appropriate rules for classification. To reduce complexity, these were all undertaken on the smaller DS-E data set. In what follows, we will explain how we arrived at the heuristic categorization rules (listed in Table 6) applied to the other data sets (see Table 7). This will then lead to further analysis and exploration of the results.

Initial categorization was based on the number of unique metacognitive tags assigned to the reflection, assuming that it contained a regulation tag. This requirement was applied because regulation is a core component of the model, and so a lack of regulation was expected to identify a weak reflection. If the reflection was tagged with all three tags (regulation, knowledge, and experience), then it was classified as strong. Weak reflections were identified by zero tags or a single tag that was not regulation.

We used a process called anomaly recontextualization (Gibson & Kitto, 2015) to refine the heuristics of the categorization rules. This involved running the analysis with a given rule, identifying anomalies in the results, then using additional features to resolve or recontextualize the anomalies. The first categorization resulted in a number of anomalies in both the strong and weak categories. Further examination of these reflections revealed a correlation between the reflections' sentence counts and the anomalies in the results. Although the number of sentences did not directly correlate with metacognitive tagging, all anomalous reflections that should have been identified as strong had a minimum of three sentences. We were able to use the minimum three sentence feature to resolve the anomaly and it was added to the rules.

It was found in early experimentation that the computational analysis was biased to deem reflections comprising a small number of sentences as "strong." This was due to a metric called subTagDensity, which normalized tag counts to the number of sentences by taking the total number of sub-component tags and dividing them by the number of sentences. We utilized this because of the significant variation in the length of reflections. However, this had the adverse effect of causing some reflections with a very small number of sentences (e.g., 2) and a normal number of tags (e.g., 4) to be rated higher than reflections with slightly higher sentence count, making them more likely to be categorized as strong. We addressed this issue by modifying subTagDensity to divide by the log of the sentence count (instead of just the sentence count), and by ascribing a value of 0 to subTagDensity for any reflections with fewer than 2 sentences.

As part of the initial categorization trials we had also identified reflections that were difficult to categorize and tagged them according to their likelihood to be ascribed a category. If the reflections were very close to being categorized as strong the algorithm ascribed a tag of possibleStrong, and if they were close to being categorized as weak, then they were tagged as possibleWeak. We examined the resultant reflections and noted that the potential changes identified above would resolve these boundary reflections in favour of the expected category. Consequently, we made the adjustments described and the combined changes resulted in a final categorization algorithm that utilized the heuristic rules shown in Table 6.

Table 6: Heuristic rules

Class	Rules
Strong	sentenceCount >=3 AND metaTagCount =3 OR sentenceCount >=3 AND metaTagCount =2 AND metaTags(regulation) AND subTagDensity >=4
Weak	NOT metaTags(regulation) AND metaTagCount <=1 OR subTagDensity <=3

Table 7 shows the resulting reflection quantities for each dataset after classification. It is possible that the small number of strong and weak categorizations for the DS-E data compared with the other datasets may prove problematic, as the categorization algorithm was initially tuned on the DS-E data set. We examine this potential problem further below.

Table 7: Reflection numbers after categorization.

	All	%	DS-E	%	DS-S	%	DS-I	%
Strong	837	13.8%	4	2.1%	59	8.0%	774	15.0%
Weak	4,180	68.6%	160	85.1%	564	76.2%	3,456	67.0%
Undetermined	1,073	17.6%	24	12.8%	117	15.8%	932	18.0%
Total	6,090	100.0%	188	100.0%	740	100.0%	5,162	100.0%

3.5 Sentence Level Check

At this point, we wanted to check whether the heuristic rules were working appropriately with data more generally. We were particularly interested in strong classifications as this most directly addressed the aim of the research. This check was performed by comparing the metaTags of individual sentences annotated by the algorithm against a manual coding of the sentences undertaken by one of the authors of this paper. As there were only four reflections in the DS-E dataset, we decided to code "strong" reflections from the DS-S dataset. However to keep the task manageable, we coded the smaller number of "weak" reflections from the DS-E dataset. We also limited the total number of reflections by choosing reflections by author where at least three of their reflections were categorized in the category being checked. That is, for the strong classification, we chose an author's reflections where at least three were categorized as strong.

The intuition behind this approach was that although we were checking at the sentence level, the aim was to discover metacognition on the part of the author, and therefore the sentences should ultimately be understood in this context. By including all of a selected author's reflections we retained a good selection of reflections within an author-centred context. We also expected that taking this approach would assist with mitigating against writing style variance. Example sentences can be seen in Table 8. The check is of the metaTag ascribed to the sentence that occurs prior to the strong/weak classification performed at the reflection level.

Table 8: Example sentences.

	Check	Sentence	MetaTag	SubTag	PhraseTag
DS-S Strong Authors	truePos	Compared to other groups I felt as though we are under-prepared and this is most likely due to the fact that we started well after most other groups .	regulation experience knowledge	monitorControl goal experience knowledge	anticipate emotive generalPronounVerb compare temporal pertains
	falsePos	So we started this week's meeting firstly by delegating jobs, I had printed off the template supplied on learning resources in which we could allocate a job title to each member per meeting, as I was the scribe last week we nominated Keith to do it this week.	knowledge	trigger goal knowledge	anticipate generalPronounVerb pertains manner outcome possible
	trueNeg	During this chat the whole group co-operated and each had a valued input into the task .			
	falseNeg	Next time I might just wait for someone else to change the discussion back to the topic so I'm not the only one.		goal	generalPronounVerb pertains possible
DS-E Weak Authors	truePos	All the assignments are driving me crazy , and I feel lost .	experience	experience	consider generalPronounVerb selfReflexive
	falsePos	Yesterday I submitted my Groundwater report !	knowledge	knowledge	selfPossessive generalPronounVerb
	trueNeg	Just reported the project progress to the lecturer .			
	falseNeg	It is always difficult to initiate a project , where every move seems heavy and uncertain .			generalPronounVerb

3.5.1 Check of metaTag for Strong Authors (DS-S Dataset)

Selecting from the DS-S dataset, based on authors with three or more strong reflections, yielded a total of 272 sentences for four authors. We manually coded the data for false positives, false negatives, true positives, and true negatives. Table 9 shows that, overall, 78% of the classifications were found to match, with 22% not matching, the majority of these being false negatives.

Table 9: DS-S strong author metaTag evaluation

	Positive	Negative	Total	%
true	55	156	211	78%
false	15	46	61	22%

Further analysis of the non-matching classifications was conducted to determine how these might be avoided. For the false positives, we found that nine of the 15 related to a metaTag of knowledge, and could have been classified as true negatives if the trigger subTag was used to prevent tagging as knowledge. We also found that this would result in no additional true positives. In addition, five of the false positives were based on the regulation metaTag and all of these instances were bracketed by true negative sentences. This suggests these false positives could be addressed by taking into account the classification of the sentences before and after, which aligns with how people tend to reflect, because an idea is more likely to be encapsulated in a number of sentences rather than just one.

For the false negatives, there appeared to be strong correlations with the subTags, providing a way in which many of the false negatives could be classified as true positives. In view of this, we considered that we may need to address the way regulation is arrived at from the underlying subTags. However we also noted that this could have resulted in a large increase in false positives. We found that 17 of the 46 would have been true positives if the goal subTag was used, four if the trigger subTag was used, and 14 if the pertains phraseTag was used. The numbers provide an indication that this area of the classification process requires further refinement, and that the approach will need to be more complex than purely utilizing the subTags and phraseTags.

When checking the data, we were not surprised to find that authors tended toward a particular style of reflection. We thought that these stylistic differences may have had a material impact on the classification, so we separated the results for each author (see Table 10). This could be an area for further investigation as part of a more author-focused analysis; however, for this study we found the results to be less variable than expected, and did not pursue this line of enquiry further.

Table 10: Results by author.

	Author 1	%	Author 2	%	Author 3	%	Author 4	%
false positives	11	0.1%	2	0.5%	0	2.0%	2	1.2%
false negatives	23	0.9%	11	0.5%	3	0.0%	9	0.6%
true pos & neg	121	1.9%	41	2.7%	15	2.0%	34	2.6%
	35	1.0%	15	1.0%	6	20.0%	15	10.8%

3.5.2 Check of metaTag for Weak Authors (DS-E Dataset)

Initially we selected reflections from authors with three or more reflections categorized as strong. We recognized the possibility that the algorithm may show signs of success with this selection if over-fitting was occurring within the higher quality data. Therefore, we felt that we needed to undertake a similar evaluation with sentences from authors with mostly weak category reflections. That is, a selection of reflections based on authors with three or more weak reflections. We used the DS-E dataset, which yielded a total of 437 sentences for 22 authors. Like the previous check, we manually coded the data for false positives, false negatives, true positives, and true negatives (see Table 11). Overall, we were encouraged by finding very similar results to the previous study, with 80% matching, and 20% not matching (compared with 78% and 22% respectively from the strong authors). A differentiating factor with the weak author reflections was that most mismatches were false positives (rather than false negatives). This makes sense, as it suggests that our algorithm is possibly not as aggressive as it could be. A more aggressive algorithm would likely result in a reduction of false negatives in the strong reflections as well as reduce the number of false positives in the weak data.

Table 11: DS-E weak author metaTag evaluation.

	Positive	Negative	Total	%
true	53	296	0	80%
false	61	27	349	20%

Of the 61 false positives, 42 belonged to reflections classified as weak, meaning that the false tagging of these sentences did not make a material difference to the outcome. However, 40 of these sentences were tagged with knowledge, indicating that if this effect applied to all reflections, then fewer reflections may be tagged as weak. Among the other 19 false positives, 12 were part of undetermined reflections, which could possibly become weak if the classification was improved. More of an issue arose with the seven false positives among reflections classified as strong. These were all based on a regulation metaTag, and because this is a critical component of a strong classification, all seven of these reflections would have been classified as undetermined if these false positives were fixed. These issues suggested to us that the side effects of the false positives on strong classification is significant and worthy of further attention.

With the exception of two sentences, the false negatives were part of reflections classified as weak, so we needed to assess the extent of their contribution to the final reflection categorization. In our assessment, all but one of these sentences should have been classified as experience, and none of the other sentences in the reflections included a regulation tag, so there was no material change to the reflection classification. Similarly with one that could have been tagged as regulation, there were no other sentences in the same reflection tagged as knowledge or experience so as to inappropriately promote the reflection to the strong category. This suggests to us that despite the false negatives, there is a moderating effect in the algorithm that arises by taking tags from all sentences in determining the reflection classification, which appears to be a good thing.

3.5.3 Algorithm Changes

From the checks, we determined that only two changes to the algorithm would not have adversely affected the evaluation data. Both of these changes related to the strong author sentences.

The first was to consider the relationship between the trigger subTag and the knowledge metaTag. Looking at the original phrase pattern matching, we deduced that the issue might have presented due to the similarity between the underlying phraseTags of outcome (related to trigger) and compare (related to knowledge). It appeared that it would be better to have included compare together with outcome for the trigger subTag; however, we determined that further analysis is required to ascertain that this does not adversely affect other classifications. We also noted that issues with the knowledge metaTag were also present in the weak author data. This reinforces the fact that the composition of this tag needs more detailed examination.

The second change involved moderating regulation tagged sentences based on sentences before or after. That is, if the sentence before or after has no metaTag, then remove the regulation metaTag. We also determined that it would be best to not implement this change until its effects had been examined within a larger dataset to determine whether there would be any adverse side effects.

3.6   Metacognitive Analysis by Author

Our final analysis for this study was on DS-I, the largest of the three data sets. While the size of the dataset did not present any computational problems, it did make our assessment of the results problematic. The lack of pre-annotated reflective text corpora makes it difficult to undertake high-quality evaluations, and although we could choose smaller subsets of DS-E and DS-S to manually annotate, the DS-I dataset was not quite as straightforward, as it did not have a clear subset to work with. However, the larger size of DS-I does have some benefits; repeated patterns can become more obvious, as can anomalies in the data. Our approach was to examine the data as a whole for obvious patterns and anomalies.

In line with the overall research objective of identifying metacognitive activity in the learner, we decided to take an author-centric view of the data. The first task we addressed was a way to rank the authors. We wanted our ranking algorithm to rank the authors exhibiting the most metacognitive activity highest, and authors exhibiting the least activity lowest. However, we knew that authors usually had a mix of strong,

weak, and undetermined reflections, and we also knew that there was a wide variance in the length and number of reflections. We also knew that, on average, there were more than four weak reflections to every strong reflection (see Table 7). In order to address these factors, we settled on an algorithm that weighted strong reflections as +2 and weak reflections as -1. Undetermined reflections were weighted with 0. We summed the weightings of all reflections of an author and divided the result by the log of the total number of reflections.

This algorithm proved useful in ranking the authors, and on running it we were immediately able to identify two key anomalies: the highest ranked author and the lowest ranked author. Both rankings of these authors were outside the almost symmetric range of the other 475 authors (-7.44 to +7.82) with the highest ranked author at +9.232 and the lowest ranked author at -27.126. What was interesting about these two authors, however, was the significant difference in their reflections to the closest ranked authors.

The highest ranked author wrote very long reflections, the longest being 53 sentences and the average being 25.9 sentences. While some authors wrote long reflections, they tended to write very few, whereas this author wrote 15 reflections with only five of them having fewer than 20 sentences. However, the large amount of text for this author also highlighted patterns in the phraseTags that were not immediately obvious with smaller reflections. For example, there appeared to be a very large number of generalPronounVerb tags, indicating that related filtering may be too aggressive and that more words were needed in the associated lexicons to minimize phrases falling through to the general tag rather than being caught by more specific tags. The longer texts also revealed an unusually high number of pertain phraseTags, often four times greater than the next highest tag. This could mean that the filter for pertains is not aggressive enough and needs to be tightened up. Finally, we observed that goal is often the most frequent subTag and that it often co-occurs with trigger, but regulation is less frequent. This indicates that either we either need to improve monitorControl detection, or we need to relax the rule for the regulation metaTag.

At the other end of the rankings, the lowest ranked author wrote a lot of very short reflections, 145 in total, of which 113 were only one sentence long. Unlike the highest ranked author, we did not find any obvious implications for changing the algorithm using this author's reflections.

Based on these findings, we set the range of rankings from -8 to +8, which effectively excluded these two anomalous authors from the dataset. The results for authors in this range are provided in Table 12.

An interesting finding from these results is the comparison of the three classifications over the ranking range. Figure 3 shows that the undetermined classification appears to have a greater effect on the weak classification than on the strong, and that this effect strengthens in the top half of the range. This suggests that we might need to fine-tune the classification algorithm around the boundary between weak and undetermined. This is not surprising, as the algorithm is looking for the presence of patterns when classifying as strong, resulting in a relatively clear boundary condition, but it is the extent of the absence

of those patterns that results in the division between weak and undetermined. We may need some more positive indicators of lack of metacognitive activity in order to strengthen the algorithm in this area.

Table 12: DS-S analysis by ranking.

	Bottom	Lower	Middle	Upper	Top	All
Rank	-8.0 to -4.7	-4.69 to -3.1	-3.09 to -0.01	0 to 2.22	2.23 to 8.0	-8.0 to 8.0
Authors	47	100	173	106	49	475
Reflections	876	1,161	1,621	785	559	5,002
Mean refs	20.33	12.99	11.6	12.7	13.59	13.85
Min refs	12	5	3	1	3	1
Max refs	33	31	26	27	25	33
Std Dev	6.06	4.62	5.11	6.12	4.6	6.12
Strong	16	24	179	238	304	761
Weak	790	991	1,043	359	137	3,320
Undeterm.	70	146	399	188	118	921
Strong %	1.8%	2.0%	11.0%	30.3%	54.4%	15.2%
Weak %	90.2%	85.4%	64.3%	45.7%	24.5%	66.4%
Undeterm. %	8.0%	12.6%	24.7%	24.0%	21.1%	18.4%
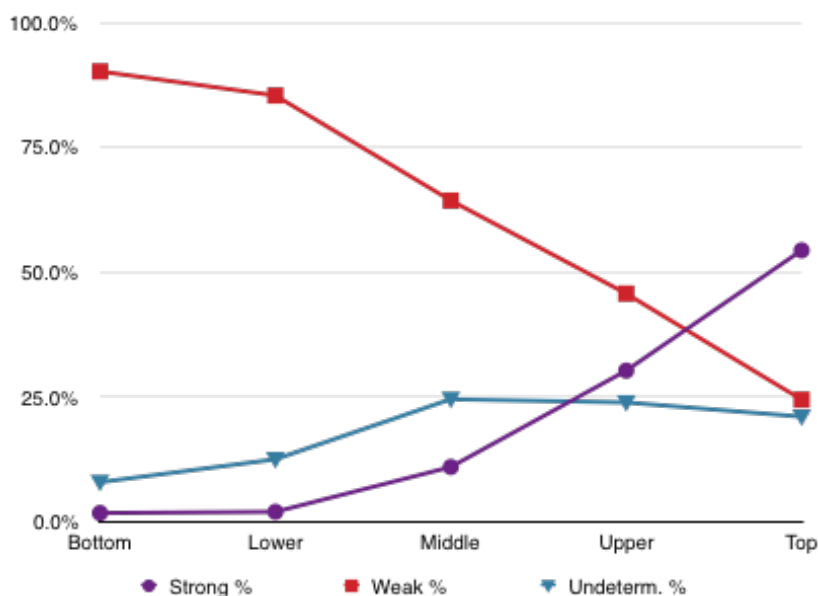
Figure 3: Percentage of metaTags for each ranking range.

Also of interest in these results were the minimum and maximum number of reflections for each ranking band. Overall, lower ranked authors wrote more reflections than those in the higher ranking. This suggests that more writing may not necessarily provide better evidence of metacognitive activity. The relationship between the number of reflections and overall metacognitive activity could benefit from further investigation, as this may have important implications for the teaching of reflective writing.

4   IMPLICATIONS FOR LEARNING ANALYTICS

We contend that our work holds a number of implications for the field of Learning Analytics that would be worthy of future pursuit.

First, a number of findings arising from the development of the conceptual model could impact the field. We found it essential to clarify the relationship between metacognition and reflection prior to the work on the computational algorithm. We see this clarification as a first step, with a lot of potential for further development of the model, particularly around the expansion of the model into the external/social area.

As our interest was fundamentally on metacognitive activity, we did not examine the various models of reflective writing, nor the relationships between them and the cognitive act of reflection. We believe that the conceptual model would benefit from further work in this area. A greater understanding of the relationships between cognition and writing may also inspire new approaches to reflective writing analytics.

The lack of an annotated corpus of reflective writing turned out to be a significant handicap in undertaking this study. We are conscious that a full and proper evaluation of the various heuristics was not undertaken, and that being able to test the algorithms against a pre-existing annotated dataset would have been a significant benefit. However, we also note that the lack of this type of resource, and the ethical issues in creating one, offers an opportunity for the development of new approaches to evaluating analytics. One such method could use automated feedback as a mechanism to collect validation from the author; that is, providing information back to authors about their reflections and requesting a response as to the extent to which it is accurate, reasonable, fair, useful, etc. This would enable an automated annotation of the corpus, which could be used to train machine learning models for computational analysis.

Related to this idea, is the automated generation of formative feedback. From the reflection data, we can easily identify key domain phrases such as "my group," "our project." We can also identify deficiencies in the type of reflection, such as a lack of expression of personal feelings. Putting these two aspects together would allow us to generate feedback along the lines of "Perhaps you could write about how you feel about your group."

Finally, as identified above with respect to the conceptual model, reflection is neither restricted to writing nor to being expressed linguistically. We consider the computational analysis of non-linguistic reflection to be a particularly interesting area for future research, with potential benefits in applying learning analytics to sports performance, musical improvisation, and other non-linguistic learning activities.

5   CONCLUSION

In this study we have examined a range of conceptualizations of metacognition and reflection and brought them together into a unified model for the purpose of providing clarity as to how the features may be analyzed computationally from reflective writing. Based on this conceptual model, an algorithm was developed that categorizes reflections based on metacognitive activity.

5.1   Limitations

Because of the exploratory nature of this study, three significant limitations should be addressed in future work.

First, we did not conduct a full empirical study of the algorithms so the results indicated in this paper are indicative of the potential of the approach rather than a confirmation of its validity. We acknowledge that in order to establish the algorithms' true potential to the field, a number of additional considerations would need to be taken into account. In particular, we would consider the following necessary: 1) the use of different data for the evaluation than that used to develop the algorithm; 2) manual annotation of the evaluation data with multiple human annotators to establish inter-rater reliability; 3) making explicit any differences between annotation criteria and algorithm criteria; and 4) where justified, the application of conventional statistical analysis to the results.

Second, the dataset used for this study was not necessarily representative of reflective writing in general. It involved the collection of mostly brief, unstructured reflections on a regular basis over a period of time. This type of reflection may be significantly different than the longer, more structured reflective writing typical of student assignments. The effect of these differences on the performance of the algorithm is unknown.

Third, the combination of structural patterns and lexical filtering necessarily means that some text will be omitted or inadvertently included, depending on the parameters and underlying lexicons. We acknowledge much room for improvement in disambiguation, pattern matching, and more comprehensive lexicons and filtering processes.

5.2   Future Work

These limitations provide some key areas for future work. We see value in a full empirical study based on the approach outlined in this paper. We see promise in both the conceptual model and the algorithm design, and believe that a more complete study with rigorous evaluation will advance both the conceptual and applied aspects of this work. We also see potential in evaluating different styles of reflective writing; in particular, investigating the differences between short, recurrent reflection and longer form, single-instance reflection. The algorithm itself presents many opportunities for further refinement, and would benefit from the incorporation of novel techniques for lexicon expansion and pattern disambiguation.

Another potential avenue of research is in the area of generating formative feedback for the author. This is a large and complex space that this study has only hinted at.

5.3   Final Remarks

Overall, the approach presented here to discovering learner metacognition through the analysis of reflective writing shows demonstrated potential, and we consider it worth pursuing in further research. While this could be used directly for the purpose of generating reflective writing analytics, we identified potential for significant impact in the field of learning analytics by developing the software further to include feedback to the learner. More broadly, this paper is an initial, yet significant, step towards a form of analytics centred upon metacognition and reflection. We have presented a way in which the detection of metacognitive activity in an author of reflective texts might be automated, and shown how this approach might be used as a formative feedback step to encourage metacognitive thinking in students. We consider this to be a key 21st century competency, and hold that it is essential that we find ways to nurture this skill in our learners. It is our ambition for this paper to open up a fruitful avenue for research in this direction of learner-centred learning analytics.

REFERENCES

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E., & Walker, R. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. Cognitive Development, 23(4), 452–471. http://dx.doi.org/10.1016/j.cogdev.2008.09.002

Boud, D., Keogh, R., & Walker, D. (Eds.), (1985). Reflection: Turning experience into learning. London: Kogan Page.

Committee for Economic Development of Australia (CEDA). (2015). Australia's future workforce? Retrieved from http://www.ceda.com.au/research-and-policy/policy-priorities/workforce

Cornford, I. R. (2002). Learning-to-learn strategies as a basis for effective lifelong learning. International Journal of Lifelong Education, 21(4), 357–368. http://dx.doi.org/10.1080/02601370210141020

Dewey, J. (1916). Democracy and education. New York: The Macmillan Company.

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. European Psychologist, 13(4), 277–287. http://dx.doi.org/10.1027/1016-9040.13.4.277

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), The nature of intelligence (pp. 231–236). Hillsdale, NJ: Erlbaum.

Flavell, J. H. (1978). Metacognitive development. In J. M. Scandura & C. J. Brainerd (Eds.), Structural/process theories of complex human behavior. Alphen a. d. Rijn, The Netherlands: Sijthoff & Noordhoff.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34(10), 906–911. http://dx.doi.org/10.1037/0003-066X.34.10.906

33

Gibson, A., & Kitto, K. (2015). Analysing reflective text for learning analytics: An approach using anomaly recontextualisation. Proceedings of the 5th International Conference on Learning Analytics and Knowledge (LAK '15), 275–279. http://dx.doi.org/10.1145/2723576.2723635

Herber, H. L., & Nelson-Herber, J. (1987). Developing independent learners. Journal of Reading, 30(7), 584–588.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. Consciousness and Cognition, 9, 147–171. http://dx.doi.org/10.1006/ccog.2000.0433

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. Computers in Human Behavior, 32(C), 313–323. http://dx.doi.org/10.1016/j.chb.2013.07.051

Mair, C. (2012). Using technology for enhancing reflective writing, metacognition and learning. Journal of Further and Higher Education, 36(2), 147–167. http://dx.doi.org/10.1080/0309877X.2011.590583

McCallum, A., Rohanimanesh, K., Wick, M., & Schultz, K. (2008). FACTORIE: Efficient probabilistic programming for relational factor graphs via imperative declarations of structure, inference and learning. Paper presented at the NIPS*2008 Worskshop on Probabilistic Programming, Whistler, Canada. Retrieved from https://people.cs.umass.edu/~sameer/files/factorie-nipsws08.pdf

Mezirow, J. (1990). Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning. San Francisco, CA: Jossey-Bass.

Moon, J. A. (1999). Reflection in Learning and Professional Development: Theory and practice. New York: Psychology Press.

Paul, R. (1990). Towards a new measure of success: Developing independent learners. Open Learning: The Journal of Open, Distance and e-Learning, 5(1), 31–38. http://dx.doi.org/10.1080/0268051900050106

Pennebaker, J. W., & Chung, C. K. (2011). Expressive writing: Connections to physical and mental health. In H. S. Friedman (Ed.), Oxford handbook of health psychology. New York: Oxford University Press.

Quirk, M. E. (2006). Intuition and metacognition in medical education. New York: Springer.

Reidsema, C., Goldsmith, R., & Mort, P. (2010). Writing to learn: Reflective practice in engineering design. Paper presented at the 9th Annual American Society for Engineering Education (ASEE) Global Colloquium on Engineering Education, 18–21 October 2010, Singapore. Retrieved from https://www.academia.edu/501660/Writing_to_Learn_Reflective_Practice_in_Engineering_Design

Ross, J. (2011). Traces of self: Online reflective practices and performances in higher education. Teaching in Higher Education, 16(1), 113–126, http://dx.doi.org/10.1080/13562517.2011.530753

Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. Teaching in Higher Education, 16(1), 99–111. http://dx.doi.org/10.1080/13562517.2010.507311

Ryan, M. (2013). The pedagogical balancing act: Teaching reflection in higher education. Teaching in Higher Education, 18(2), 144–155. http://dx.doi.org/10.1080/13562517.2012.694104

Sanders, J. (2009). The use of reflection in medical education: AMEE Guide No. 44. Medical Teacher, 31, 685–695. http://dx.doi.org/10.1080/01421590903050374

Schön, D. A. (1983). The reflective practitioner: How professionals think in action. New York: Basic Books.

Takanishi, R. (2015). Do our education systems do enough to enable learners to flourish as independent, autonomous and well-balanced individuals? European Journal of Education, 50(2), 123–127. http://dx.doi.org/10.1111/ejed.12113

Tang, R., & John, S. (1999, December). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. English for Specific Purposes, 18, Supplement 1, S23–S39. http://dx.doi.org/10.1016/S0889-4906(99)00009-5

Ullmann, T. D., Wild, F., & Scott, P. (2012). Comparing automatically detected reflective texts with human judgements. In A. Moore, V. Pammer, L. Pannese, M. Prilla, K. Rajagopal, W. Reinhardt, T.D. Ullmann, C. Voigt (Eds.), Proceedings of the 2nd Workshop on Awareness and Reflection in Technology-Enhanced Learning (AR-TEL '12), Saarbrucken, Germany (pp. 101–116). Retrieved from http://ceur-ws.org/Vol-931/paper8.pdf

Wharton, S. (2012). Presenting a united front: Assessed reflective writing on a group experience, Reflective Practice: International and Multidisciplinary Perspectives, 13(4), 489–501. http://dx.doi.org/10.1080/14623943.2012.670622

Wilson, M. D. (2015). Pracademia: The future of the lifelong learner. About Campus, 20(2), 28–31. http://dx.doi.org/10.1002/abc.21189

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into Practice, 41(2), 64–70. http://dx.doi.org/10.1207/s15430421tip4102_2

APPENDIX

A live demonstration version of the software created as part of this study, as well as links to the source code and various technical documents, can be found at: http://nlytx.io/2016/metacognition
Additional examples for the process involved in deriving the metaTags follow:

Example 1

Text	I know I	can count	on [name] to do her bit and	I know she	will do	it well as the recovery part is not news	to us , Enviro Science chicks .
posTag	PRP VB PRP	MD VB	IN NN TO VB PRP NN CC	PRP VB PRP	MD VB	PRP RB IN DT NN NN VBZ RB NN	TO PRP NNP NNP NNS
match	start(PRP) & repeat(RB, VB, CC, IN) start(PRP) & any & end(PRP)	start(MD) & repeat(RB, VB)	start(IN,TO) & repeat(PRP, JJ, NN, VB, MD, PDT, W, TO)	start(PRP) & repeat(RB, VB, CC, IN)	start(MD) & repeat(RB, VB)	start(PRP) & repeat(RB, VB, CC, IN)	start(IN, TO) repeat(PRP, JJ, NN, VB, MD, PDT, W, TO)
filter	*containsAny (know)*	*startsWith Any* (can)	*startsWithAny* (on)	*containsAny (know)*	*startsWith Any* (will)	containsAny (well)	*startsWithAny* (to)
phraseTag	Consider selfReflexive	definite	pertains others-Possessive	consider	definite	emotive	pertains
match	temporal OR (pertains AND consider) anticipate OR definite OR possible emotive OR selfReflexive			temporal OR (pertains AND consider) anticipate OR definite OR possible emotive OR selfReflexive			
subTag	monitorControl, goal, experience			monitorControl, goal, experience			
metaTag	regulation, experience						

Example 2

Text	They all seemed	like lovely people	and are interested	in participating to achieve	the best possible mark .
posTag	PRP DT VBD	IN JJ NNS	CC VBP JJ	IN VBG TO VB	DT JJS JJ NN
match		start(IN, TO) repeat(PRP, JJ, NN, VB, MD, PDT, W, TO)		start(IN, TO) repeat(PRP, JJ, NN, VB, MD, PDT, W, TO)	
filter	startsWithAny (like)			*startsWithAny* (in)	
phraseTag	compare			pertains	
match	selfPossessive OR compare OR manner				
subTag	knowledge				
metaTag	knowledge				