




RESEARCH ARTICLE

Open Access



Impact of combining human and analytics feedback on students' engagement with, and performance in, reflective writing tasks

Wannapon Suraworachet , Qi Zhou  and Mutlu Cukurova* 

*Correspondence:
m.cukurova@ucl.ac.uk

UCL Knowledge Lab, IOE, UCL's
Faculty of Education and Society,
University College London, 23-29
Emerald St, London, UK

Abstract

Reflective writing is part of many higher education courses across the globe. It is often considered a challenging task for students as it requires self-regulated learning skills to appropriately plan, timely engage and deeply reflect on learning experiences. Despite an advance in writing analytics and the pervasiveness of human feedback aimed to support student reflections, little is known about how to integrate feedback from humans and analytics to improve students' learning engagement and performance in reflective writing tasks. This study proposes a personalised behavioural feedback intervention based on students' writing engagement analytics utilising time-series analysis of digital traces from a ubiquitous online word processing platform. In a semester-long experimental study involving 81 postgraduate students, its impact on learning engagement and performance was studied. The results showed that the intervention cohort engaged statistically significantly more in their reflective writing task after receiving the combined feedback compared to the control cohort which only received human feedback on their reflective writing content. Further analyses revealed that the intervention cohort reflected more regularly at the weekly level, the regularity of weekly reflection led to better performance grades, and the impact on students with low self-regulated learning skills was higher. This study emphasizes the powerful benefits of implementing combined feedback approaches in which the strengths of analytics and human feedback are synthesized to improve student engagement and performance. Further research should explore the long-term sustainability of the observed effects and their validity in other contexts.

Keywords: Reflective writing, Engagement feedback, Learning analytics, Time series, Human-AI collaboration, Google Docs analytics

Introduction

Reflective writing refers to students' written journal of learning experiences over time to develop self-awareness of their own learning (Thorpe, 2004). It is often utilised in higher education to stimulate their thoughtful reflection on the incidents and hence promote transformative learning and practice (Ryan, 2013). Several studies showed that incorporating reflective writing could increase students' content comprehension (Strong et al., 2001), academic performance (Connor-Greene, 2000) and life-long

learning skills (Boutet et al., 2017; Thorpe, 2004). However, reflective writing is a complicated and demanding task that requires learners to be able to regulate their own learning (Zimmerman & Risemberg, 1997). Self-regulation skills allow a person to strategically regulate their behaviours and environment towards their goals (Zimmerman, 1989). To be more specific, reflective writers are required to set goals on the coverage contents and deploy multiple cognitive processes including planning, writing and revising to complete a writing task (Graham & Harris, 1994).

There is strong evidence that pedagogically incorporating reflective writing tasks without appropriate support is unlikely to lead to effective learning due to students' challenges in engaging with independent learning (Cukurova et al., 2018) and with critical reflective skills that can transfer into other contexts (McIntosh, 2010). Therefore, feedback is a necessary condition to help learners engage in their writing tasks and also develop key skills associated with reflective writing practice (Sadler, 2010). Written feedback provided by teachers has long been recognised as an effective method to improve students' performance (Page, 1958), especially in written assignments (Stewart & White, 1976). Yet, research on reflective writing has so far mainly focused on teachers' written feedback that emphasises cognitive development and content acquisition (Aronson et al., 2012; Thorpe, 2004), frequently overlooking other factors that also contribute to an improvement in learning, such as students' emotions, motivations and behaviours. The quality and meaningful feedback elements suggested for reflective writing tasks at the motivational and behavioural levels (Aronson et al., 2012; Dekker et al., 2013) as well as reactions to the tone of feedback (Dekker et al., 2013; Rozental et al., 2021) are understudied.

In recent years, learning analytics provide opportunities to support students with meaningful feedback on motivational and behavioural aspects of their learning using data from the digital traces of student activities. For example, dashboards containing information about behavioural learning engagement with the online learning management platforms (e.g. resource uses, time spent and performance level) (Bodily & Verbert, 2017) or those that aim to monitor and support students' motivational goals (Jivet et al., 2021) have been designed and used to provide support for students' learning. Even though many studies attempted to provide feedback for the reflective writing tasks, most focused on analysing the writing contents i.e. semantic complexity to provide feedback on the content of students' reflective writings (Shibani, 2020; Shibani et al., 2019). However, trace data available in digital writing platforms are rarely combined with human feedback in interventions. It is important to study the combined feedback approaches' impact on students' learning and engagement since analytics feedback and human feedback tend to have different strengths and weaknesses (Cukurova, 2019; Luckin, 2018). Moreover, the real-world implementations of the interventions with the analytics feedback on students' reflective behaviours and more specifically academic performance are limited. We argue that behavioural engagement feedback generated with learning analytics when combined with traditional content feedback from human educators, can lead to better engagement with, and performance in, students' reflective writing. This paper presents the results of a semester-long intervention study that investigated the combined effects of personalised behavioural analytics feedback, based on a time series analysis of digital traces

from a ubiquitous online word processing platform, and human educators' feedback on reflective writing.

Literature review: reflective writing support with learning analytics

Multiple research studies aim to unravel writing processes and the quality of reflective writing from different perspectives to support reflection. On the one hand, several studies in Writing Analytics utilise methodological advances in natural language processing (NLP) to analyse the contents of writing from a final learning artefact and explore predictive features for identifying writing performance, hence helping to automate reflective writing scoring (Buckingham et al., 2016). For example, it is found that linguistic features extracted from multi-source argumentative writing essays could determine individual differences in vocabulary score and hence help develop personalised learning systems for writing (Öncel et al., 2021). Similarly, linguistic features, such as word length, sentence length, and sentence structure, could also provide valuable evidence for essay scoring in a 30-min writing task (Bridgeman & Ramineni, 2017). However, the model's value to predict students' performance in real-world writing tasks was very limited. This is in line with the study of Kovanovic et al. (2018) which explored linguistic features in relation to model reflective elements namely observation, motive, feedback and goal from reflective writing documents. The authors highlighted the major drawback of the approach to be limited to the contexts studied. Thus, whilst shown to be useful for particular contexts, the reliability, validity, and cross-context generalisability of such models of writing analytics are often critiqued (Crossley et al., 2019; Kovanović et al., 2018; Neto et al., 2021) and still considered inappropriate for real-world implementations.

Apart from investigating writing content for scoring purposes, some studies tried to gain insight into the process of writing aiming to support effective writing behaviours in general. In this stream of work, researchers looked into writing artefacts to potentially detect different cognitive operations involved in the process. For instance, Winograd and colleagues (2021) presented an NLP approach to identify the depth of scientific reasoning in students' written work. Other studies extracted digital traces of digital learning platforms to generate a visualisation of writing processes for improving awareness and hence increasing learning performance. To illustrate, Shibani (2020) presented a technique for visualising the revision behaviours by generating Automated Revision Graphs (ARG) which could provide information about the writing process and student interactions with feedback. Another work by Turkay, Seaton and Ang (Turkay et al., 2018) developed a writing analytics tool called *Itero*, aiming at visualising temporal writing processes gathered from revision logs to promote students' self-efficacy. However, these studies only focused on the revision behaviours of students, while other writing behaviours were not considered. In addition, analytics on the writing quality is often considered limited and lacking the expected level of semantic complexity to support student reflections as a standalone solution. For instance, Gibson et al. (2017) who provided a conceptual framework for reflective writing with an automated approach to model writing and provide feedback to students, showed the results that most participants reported helpfulness and expressed their willingness to use the tool in the future. However, as the authors stressed, the generalisation of the system was a major limitation. Since the writing analytics system was developed based on the content of the writing,

the transferability of the system to other contexts and different writing contents was problematic. The issue was also highlighted in the recent work of Liu et al. (2021) as the shortcomings of the NLP-based Capability for Written Reflection (CWRef) model to capture context-dependent reflective elements and adapt for variation in specific learning design and assessment were discussed at length. Moreover, this study only proposed a model, yet hasn't incorporated it into any real-world interventions for evaluation.

Moreover, there is a lack of literature focusing on the impact of long-term analytics interventions and an abundance of short-period and one-off intervention studies. For example, Cotos et al. (2020) evaluated the impact of the Research Writing Tutor (RWT), a web-based tool for academic writing which can provide different levels of automatic feedback. By analysing students' revision logs, captured screens, and stimulated recalls, the authors argued that RWT can promote students' close and deliberate examination of their produced text through different types and forms of feedback. However, RWT was only used in one class period. As the authors stated, this limited time was likely to affect the evaluation of the tool. Longitudinal data from a longer period would provide more information about how RWT might improve students' writing performance. In another study by Shibani et al. (2019), they developed an automatic writing analytics tool, called AcaWriter. The tool analyses the rhetorical moves by using NLP and provides formative feedback. It was implemented in two different contexts and has been shown to have the potential of providing meaningful contextualised support for writing. However, given the fact that both of these two implementations focused on one-off tasks, whether students can migrate the skills to future writing to reap long-term benefits remains an open question.

In addition, even if the required technical issues are resolved with regards to generating analytics of writing content to support reflective writing, considering content feedback alone might not lead to the expected learning outcomes. For example in Wingate's (2010) intervention study on content feedback presented, while some students improved their writing quality over time based on the suggested feedback, many other students did not. Those students pointed out that the content feedback alone has lessened their motivation and self-efficacy which resulted in their disengagement behaviours with the feedback. A similar phenomenon was observed in Mitchell, McMillan and Rabbani's study (Mitchell et al., 2019) that students with low self-efficacy and high anxiety levels, experienced low capability as writers from the content feedback alone. Hence, apart from feedback on writing content, feedback on other factors such as students' behavioural engagement might be a prerequisite for effective learning outcomes. Behavioural engagement feedback appraises and supports students' commitment to their efforts toward their own learning and has been shown to support learning outcomes (Vytasek et al., 2020). For instance, previous research has shown that highly self-regulated learners develop systematic engagement patterns in reflective writing tasks which correlate with higher reflective writing performance (Suraworachet et al., 2021). Similarly, several other studies reported positive effects of engagement feedback in other contexts. For example, Plak et al. (2022) deployed behavioural feedback in a form of email nudges which could promote higher engagement in online practice exams. Iraj et al. (2020) reported a study on the feedback concerning students' online participation, directing them towards quizzes and reviewing processes and found that there was a link between timely engagement

with the feedback and success in learning. In addition, Nelson et al., (2012) claimed year-long persistent engagement of at-risk students, after implementing the continuing behavioural engagement feedback. These studies indicate that the analytics on writing content and the semantics of reflections might currently be considered limited in real-world implementations. However, the analytics on the behavioural aspects of students' writing might still bring significant value to their reflective writing performance. Based on this premise, our hypothesis is that behavioural engagement feedback on students' reflective writing, not as a replacement for writing content feedback but as a supplement to it (Cukurova et al., 2019), can create opportunities to increase students' overall engagement with their reflective writing tasks and can increase their performance. Currently, there is a lack of studies investigating such relationships between different types of reflective writing support with analytics and students' performance in long-term interventions. In this paper, we aim to fill in this gap with a semester-long real-world study investigating the impact of an intervention that combines human educator feedback on content with students' writing engagement feedback with analytics on their reflective writing performance.

Research questions and hypotheses

Behavioural feedback on students' reflective writing engagement was generated from log data of their actions in a ubiquitous online word processing platform (Google Docs). Since log data can be obtained from a pervasively accessible digital writing platform and requires no context-specific sense-making process like NLP-based content analytics, the behavioural analytics generated has the potential to be generalised into other contexts. With regards to the design of the formative feedback, we adopted Hattie and Timperley's (2007) feedback model consisting of three components for effective feedback: (1) learning goals (Where am I going?), (2) learning progress (How am I going?) and (3) activities leading to better progress (Where to next?), into our behavioural engagement feedback. We studied students' writing engagement behaviours from two authentic higher education cohorts comparing the intervention cohort that received the additional behavioural engagement feedback and the control cohort that received no behavioural engagement feedback. Students' writing behaviours were modelled to identify differences between the high-performance and highly self-regulated students, and their peers on the other ends of these spectra. More specifically, we investigated three research questions (RQ).

- RQ1: How does the intervention based on feedback about students' reflective writing behaviours affect their engagement with the reflective writing task?
- RQ2: How does the feedback intervention affect the writing task engagement of students with different self-regulated learning (SRL) competence?
- RQ3: What are the relationships between students' writing engagement behaviours, and their final grades?

Driven by the literature reviewed above, for the first research question (RQ1), we hypothesised that behavioural engagement feedback could help promote persistent or higher writing engagement in the intervention group than in the control group. In other words, there would be significant differences in terms of quantity, weekly

engagement patterns, or both, between the control and the intervention groups after the feedback intervention due to engagement encouragement from the feedback (Vytasek et al., 2020).

In relation to the second research question (RQ2), it was hypothesised that students in the intervention cohort, regardless of their SRL levels, would show persistent writing engagement after the intervention period. However, the low SRL score from the intervention cohort may particularly benefit more from the behavioural feedback by continuing to engage with the writing task after receiving the feedback compared to their engagement of the period before. Based on previous research that suggested that content-only feedback might cause high anxiety and low self-efficacy, especially on students with low SRL (Mitchell et al., 2019), it was hypothesised that the behavioural engagement feedback may help them realise the necessity to persistently work on the task; hence support engagement patterns with the reflective writing task.

Finally, it was hypothesised that there will be a significant correlation between the engagement analytics used and students' academic performance measured by their reflective scores (RQ3). More specifically, the higher the engagement of students with the reflection task, the higher their performance was expected to be.

Methodology

Educational context

The study was conducted within a postgraduate course for two consecutive years. Ethical approval was received through the institutional processes. All students were informed about the study with a clear information sheet and provided their written consent at the beginning of the module. In total, 81 students consented to participate: 40 in the control group (the former year) and 41 in the intervention group (the latter year). According to the demographic data, the two groups show no difference in their age range, $\chi^2(3, N = 81) = 4.450, p = 0.217$, gender, $\chi^2(2, N = 81) = 1.109, p = 0.574$, mode of study (full-time vs part-time), $\chi^2(1, N = 81) = 0.617, p = .432$, background of study, $\chi^2(2, N = 81) = 0.245, p = .885$, and working experiences, $\chi^2(2, N = 81) = 4.746, p = 0.093$. They were assigned into small groups of 4 or 5 students with mixed-gender and interdisciplinary backgrounds. Over a 10-week course, students were introduced to the topics of the design and use of educational technology and were asked to work collaboratively to propose a technical solution for an educational challenge they have chosen. For both years, the lectures were on Tuesdays. Each week, before the lectures, students were expected to (1) complete their weekly readings, (2) study pre-recorded videos, (3) participate in a cohort-level debate, and after the lectures, (4) write a weekly individual reflection on their learning experiences via a single Google Docs, before the next week starts.

This study focused on the individual reflective writing task. There were 9 weeks in total for this task since the writing task was optional for the first week which was an introductory week for students to get used to the format and tools used in the module. Individual reflections were taken into account as 40% of the student's final grade of the course. For both years, content-based formative feedback was manually provided by the tutors as in-text comments at mid-term (week 6), and summative feedback of the final grade was provided 4 weeks after submissions.

The intervention design

In both cohorts, the content, delivery and reflection feedback provided by the tutors were the same. However, in the intervention group, students were sent additional personalised writing engagement feedback via email apart from their content-based formative feedback on week 6 to better support them with their engagement in the writing task. As discussed in Hattie and Timperley's (2007) effective feedback model, the personalised behavioural engagement feedback in this study consists of (1) a recap of findings from the prior study on how high SRL students behaved and how this affected their performance (learning goals), (2) a description of an individual student's writing engagement extracted from the digital writing platform (learning progress) and (3) suggestions on how to improve their own engagement (prospective activities). To be more specific, the email feedback attached a graph showing an engagement comparison across the first five weeks between individual students, the control cohort, and the intervention cohort. The feedback first introduced the engagement comparison between the control and the intervention cohorts. Then, based on the individual student's pattern of writing behaviours, the feedback suggested students to reflect regularly every week and keep up with their reflective writing tasks. Meanwhile, it also stressed that a higher number of edited contents did not necessarily lead to better learning outcomes, but a systematic pattern of reflective writing did. Appendix A: Email feedback shows the template of this email feedback.

Data collection tools

Log data from Google Docs¹ were collected and exported through a modified Google Chrome plug-in called Draftback². Google stores log data as revisions and the revision number represents a unique chronological auto-incremental number of the edited document. Draftback retrieves data from Google API to generate statistical summaries and visualisations of these log data to represent students' writing activities. Data prepared by Draftback is composed of the (1) *type of activity* which can either be insertion or deletion, (2) *starting index* within the document where a particular edited activity happens, (3) *ending index* within the document where a particular edited activity ends, (4) *string* or the contents have been inserted, (5) *revision number* as generated from Google Docs, (6) *user ID* and (7) *timestamp*. We modified Draftback to be able to export log data in .csv format for further investigation of the logged data with learning analytics.

SRL instruments and clustering

In order to investigate participants' SRL competence, students from both years were asked to fill in the same standardised self-report questionnaire at the beginning of the module. The questionnaire was adapted from a meta-analysis of SRL (Sitzmann & Ely, 2011) consisting of four main dimensions, namely goal-setting (GS, 4 items), effort (E, 2 items), self-efficacy (SE, 9 items), and persistence (P, 10 items), that had the strongest effects on students' academic performance. The adapted version of the questionnaire for the studied contexts can be found in Appendix B: SRL questionnaire. Cronbach alpha

¹ <http://docs.google.com>.

² <https://chrome.google.com/webstore/detail/draftback/nmajoiemfpldioamchanognpjmocgkbg>.

values per dimension were calculated ($GS = 0.853$, $E = 0.907$, $SE = 0.881$, $P = 0.905$) to test the inter-item reliability of each dimension. To categorise students into different levels of SRL competence, the K-means clustering (MacQueen, 1967) was performed separately for each group based on their scores on these dimensions. To maximise the average centroid distance with high interpretability of the clusters, two clusters (average centroid distance (control group) = -0.972 , average centroid distance (intervention group) = -1.027) were applied: (1) high SRL cluster (control group: 25 students, intervention group: 27 students), (2) low SRL cluster (control group: 15 students, intervention group: 14 students). A Chi-square test was performed to identify whether there is a difference in the proportion of SRL levels across years.

Data analysis

In order to investigate students' engagement with their writing tasks, we have analysed their Google Docs log data using time series analysis. Time series is a sequence of time-ordered data. It is a prevalent approach to modelling complex behaviours in many disciplines and making predictions about future behaviours based on historic data including finance, engineering, and health sciences, but still uncommon in educational contexts (Shin, 2017). Time series can be decomposed into components for further inspection of certain behaviours. For example, the trend represents long-term movement in time series, and seasonality refers to a short-term periodic pattern under a fixed period. In this study, the seasonality component was used as a proxy to investigate the regularity in students' engagement with the reflective writing task. Unlike other learning activities, we asked students to reflect on the weekly contents with no obligation to perform the task at a specific time. Hence, how a particular student deliberately plans to work on a task at their own preferred time is related to their ability to regulate their learning activities.

Seasonal decomposition from the *Statsmodels* Python package³ was applied. The additive seasonal decomposition was deployed due to a static seasonal component observed from the data (Hyndman & Athanasopoulos, 2018). The additive seasonal decomposition started by extracting the trend from the time-series data using the moving average method. The detrended data was further used to extract the recurring pattern, the seasonality. A fixed period of seven days was selected both for identifying trends through a 7-day moving average and as a model parameter to consider seasonality at a weekly interval.

For comparative purposes, engagement data was divided into two periods: (1) before the study intervention (feedback email on students' engagement patterns) (Week 1–5) and (2) after the intervention until the end of the module (Week 6–10). Figure 1 represents overview of this study. In respect to the RQ1 and RQ2: how does the feedback intervention on students' reflective writing engagement affect their writing behaviours across years and across different levels of SRL, time series analysis and statistical comparison tests were applied to compare data at different levels: (1) at the cohort level (control and intervention groups) and (2) at the SRL cluster levels. Firstly, a time-indexed plot

³ <https://www.statsmodels.org>.

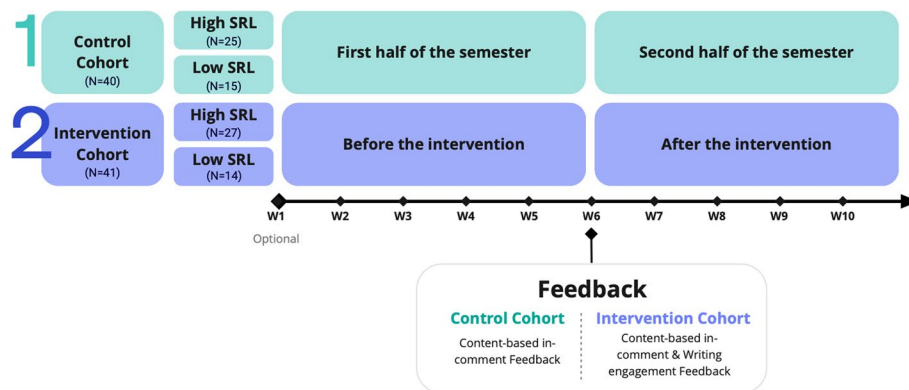


Fig. 1 Overview of the study and feedback intervention across clusters of students

was used to visually represent an overview of the time series compared before and after the feedback. Then, seasonal decomposition was deployed to extract trends and seasonality. Secondly, descriptive statistics and statistical comparison tests were used to represent and confirm statistical differences across the groups. Regarding the RQ3, Pearson’s *r* correlation was used to determine the relationship between students’ final reflection scores and their derived writing engagement behaviours. In addition, an independent sample *t*-test was deployed to measure statistical significance differences in the derived writing engagement behaviours of the intervention cohort and the control cohort.

Results

The subsequent sections represent year comparison, followed by varied SRL cluster comparisons using time series analysis and the results of the statistical comparison tests. The relationships between the derived writing engagement behaviours and the reflective scores of students are presented in the last section.

Year comparison (RQ1: How does the intervention based on feedback about students’ reflective writing behaviours affect their engagement with the reflective writing task?)

Time series analysis

Figure 2 shows the average number of edited strings per day (*AvgStrCountPerDay*) of students compared between the control group (blue) and the intervention group (red) before (Fig. 2a) and after the feedback intervention (Fig. 2b). Both cohorts followed the same patterns of engagement before the intervention feedback (Fig 2a), except in the first introductory week in which the control showed a surge in writing engagement although it was the optional week for reflection. On the other hand, whilst the control cohort exhibited similar visual patterns of engagement for the entire semester, the intervention cohort exhibited immediate and consistent engagement through higher numbers of *AvgStrCountPerDay* after the intervention (Week 6).

Seven-day seasonality was extracted to observe weekly patterns to explore how students aligned their reflection behaviours in relation to other learning activities within the module. From Fig. 3, there were observable visual differences in seasonality of *AvgStrCountPerDay* between the control cohort (Fig. 3a, blue) and the intervention cohort (lower blue) before the feedback intervention. While the control cohort developed a

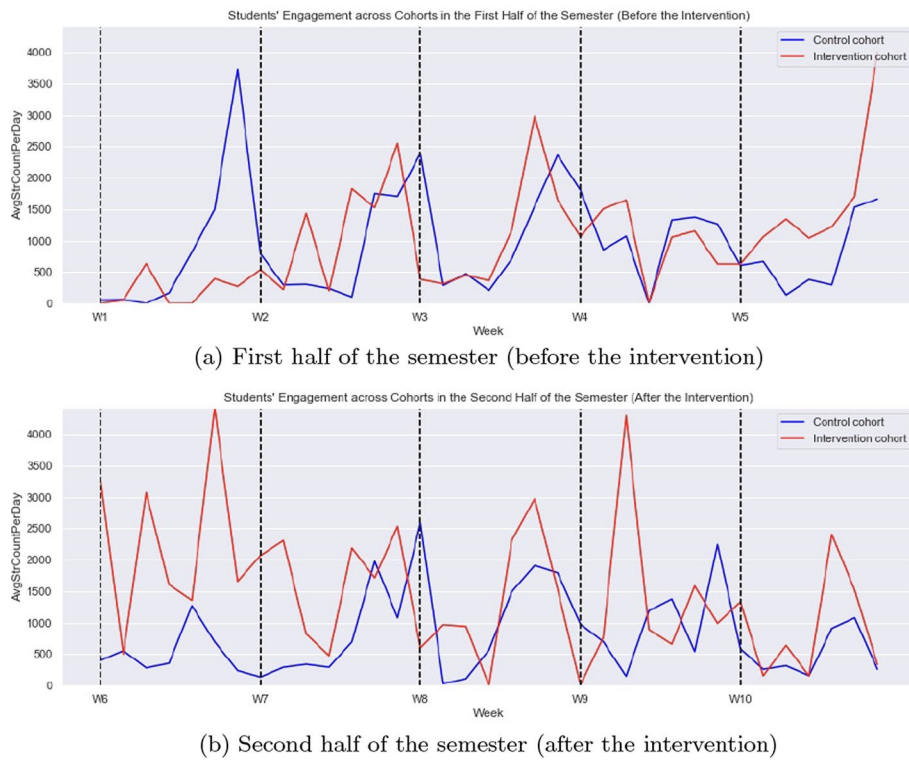


Fig. 2 AvgStrCountPerDay of both cohorts (the control group: blue line, the intervention group: red line) across semester. The vertical line denotes Monday of the week

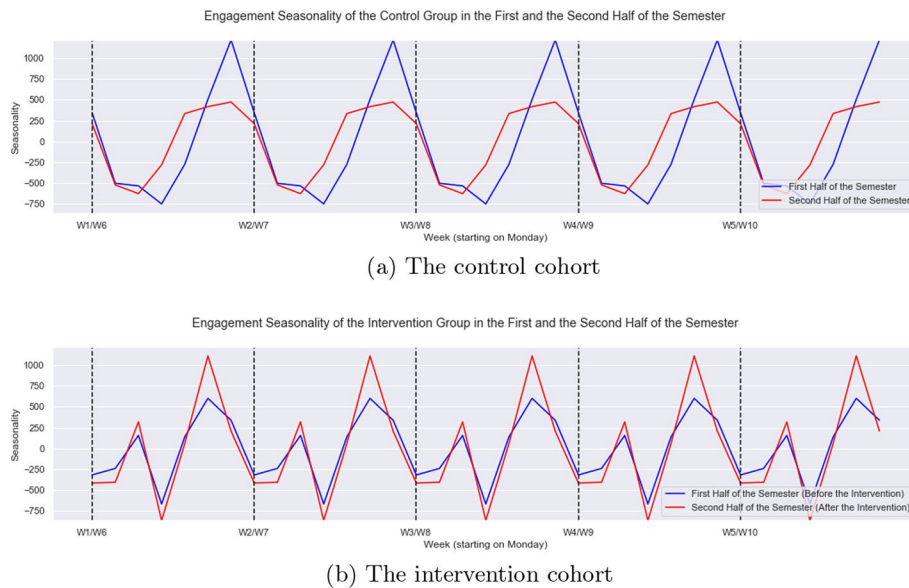


Fig. 3 Extracted seasonality of AvgStrCountPerDay of the control cohort (a) and the intervention cohort (b) compared between the first (blue) and the second half (red) of the semester (after the intervention)

1-peak weekly pattern (peaks on Sundays), the intervention cohort demonstrated a 2-peak weekly pattern (peaks on Wednesdays and Saturdays) in which the seasonality of the cohorts highly deviated from each other. They both shared their weekly minima

Table 1 The number of editing frequency per week (%) among cohorts

The number of editing frequency per week (%)	Control cohort		Intervention cohort	
	Before the feedback	After the feedback	Before the feedback	After the feedback
No edit	28.00	45.00	40.49	35.12
Edit once per week	47.50	35.50	31.21	32.68
Edit twice or more per week	24.50	19.50	28.29	32.20

Table 2 Wilcoxon signed rank test on *AvgStrCountPerDay* of the cohorts compared the first and the second half of the semester (before and after the intervention, respectively)

<i>AvgStrCountPerDay</i>	First half of the semester (Before the intervention)		Second half of the semester (After the intervention)		Z	p	Effect size
	N (Inactive days)	Md (IQR)	N (Inactive days)	Md (IQR)			
Control cohort	35 (1)	709 (1294.98)	35 (0)	564.35 (909.78)	- 0.508	0.612	0.06
Intervention cohort	35 (2)	1041.56 (1193.24)	34 (1)	1354.17 (1669.83)	- 2.162	0.031*	0.26

*p is significant at the 0.05 level (2-tailed).

IQR: interquartile range

on Thursdays. Similarly, 1-peak vs 2-peak weekly patterns among cohorts by calculating the number of editing frequencies per week among cohorts are presented in Table 1. The intervention cohort illustrated a higher percentage of editing twice or more times per week after the feedback intervention (from 28.29% before the feedback to 32.20% afterwards) whereas the control cohort showed a lower percentage of editing twice or more times per week.

After the feedback, the intervention cohort expressed a consistent 2-peak engagement pattern at a weekly level and further intensified with a higher variance of engagement patterns. These can be recognised through the extreme deviation of the seasonality around zero (Fig. 3b, the red line). In contrast, earlier fall and rise in weekly engagement were spotted from the control cohort’s seasonality on Wednesdays and Fridays respectively compared to Thursdays and Sundays in the period before the feedback (Fig. 3a, the red line). The differences in writing engagement patterns were further inspected using statistical comparison tests in the following section.

Statistical comparison tests

There were four data sets namely daily engagement of (1) the control cohort in the first half of the semester, (2) the control cohort in the second half of the semester, (3) the intervention cohort in the first half of the semester (before the intervention), and (4) the intervention cohort in the second half of the semester (after the intervention), for comparison. Due to the non-normality of datasets, related-sample Wilcoxon signed-rank tests were used to compare differences in engagement before and after the feedback intervention within the same cohort. Additionally, Mann-Whitney U-tests were applied

Table 3 Mann-Whitney U-test of *AvgStrCountPerDay* before and after the intervention compared between two cohorts

<i>AvgStrCountPerDay</i>	Control cohort		Intervention cohort		<i>U</i>	<i>p</i>	Effect size
	N (Inactive days)	Md (IQR)	N (Inactive days)	Md (IQR)			
First half of the semester (Before the intervention)	35 (1)	709 (1294.98)	35 (2)	1041.56 (1193.24)	593.00	0.819	0.027
Second half of the semester (After the intervention)	35 (0)	564.35 (909.78)	34 (1)	1354.17 (1669.83)	369.00	0.004**	0.34

***p* is significant at the 0.01 level (2-tailed).

IQR: interquartile range

to identify differences in engagement behaviours between cohorts before and after the feedback intervention.

Regarding differences within the same cohort, Wilcoxon signed-rank indicated that there was no statistically significant difference in *AvgStrCountPerDay* of the control cohort between the first and the second half of the semester, $Z = -0.508$, $p = 0.612$. In contrast, statistically significant difference in writing engagement behaviours of the intervention cohort between the first half (before the intervention) and the second half of the semester (after the intervention) was detected, $Z = -2.162$, $p = 0.031$, with a small effect size, $r = 0.26$. *AvgStrCountPerDay* significantly increased from 1041.56 characters (IQR = 1193.24, $N = 35$) before the feedback intervention to 1354.17 characters (IQR = 1669.83, $N = 35$) after the feedback intervention (Table 2).

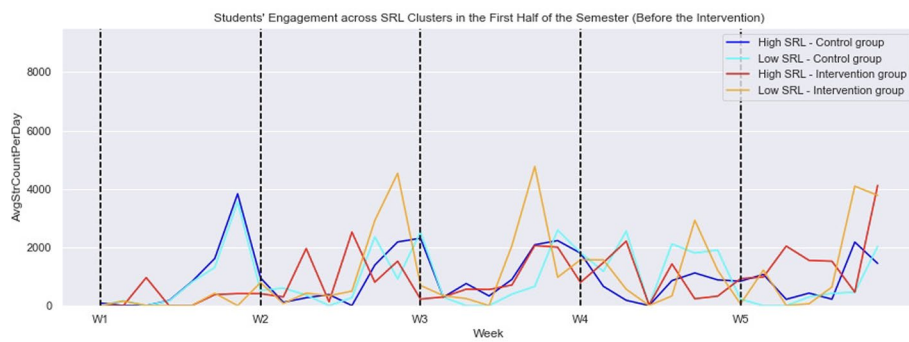
With respect to writing engagement differences across cohorts, Mann-Whitney U-test indicated that there was no statistically significant difference, $U = 593.00$, $p = 0.819$, in *AvgStrCountPerDay* between the control (Md = 709, IQR = 1294.98) and the intervention cohort (Md = 1041.56, IQR = 1193.24) in the first half of the semester (before the intervention) (Table 3). On the contrary, Mann-Whitney U-test demonstrated that there was significantly higher writing engagement in the intervention cohort (Md = 1354.17, IQR = 1669.83) compared to the control cohort (Md = 564.35, IQR = 909.78) after they received the feedback, $U = 369.00$, $p = 0.004$, with a medium effect size $r = 0.34$.

SRL Clusters’ writing engagement results (RQ2: How does the feedback intervention affect the writing task engagement of students with different self-regulated learning (SRL) competence?)

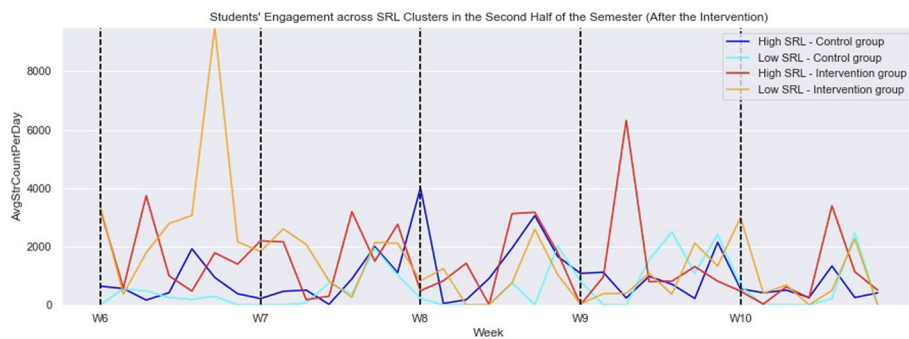
The following sections describe writing engagement results among high and low SRL groups across cohorts. It is important to note that the proportions of SRL levels did not differ by the two cohorts, $\chi^2(1, N = 81) = 0.099$, $p = 0.753$.

Time series analysis

Time series plots for the first and the second half of the semester were depicted in Fig. 4a and b, respectively. Each plot shows a comparison in the number of *AvgStrCountPerDay* across clusters: (1) the control cohort’s high SRL score students (blue line), (2) the control cohort’s low SRL score students (cyan line), (3) the intervention cohort’s high SRL score students (red line), and (4) the intervention cohort’s



(a) First half of the semester (before the intervention)

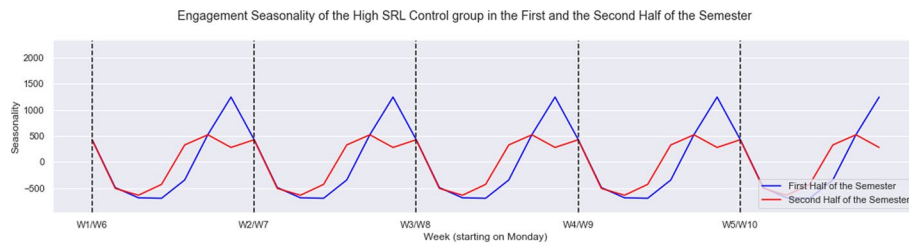


(b) Second half of the semester (after the intervention)

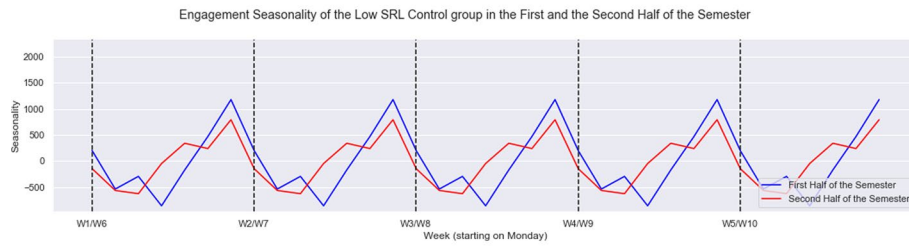
Fig. 4 *AvgStrCountPerDay* compared high SRL students from the control cohort (blue), low SRL students from the control cohort (cyan), high SRL students from the intervention cohort (red), and low SRL students from the intervention cohort (orange) across semester

low SRL score students (orange line). Overall, there was a comparable amount of writing engagement from high and low SRL groups of the control cohort in the first half of the semester (Fig. 4a, blue and cyan lines), yet higher engagement after first-half of the semester from the control cohort’s high SRL group compared to the low SRL group of the same year was observed. There were lower numbers of inactive days (the day with no engagement) among groups in the second-half of the semester compared to the first-half of the semester, except in the control cohort’s low SRL group (Table 5).

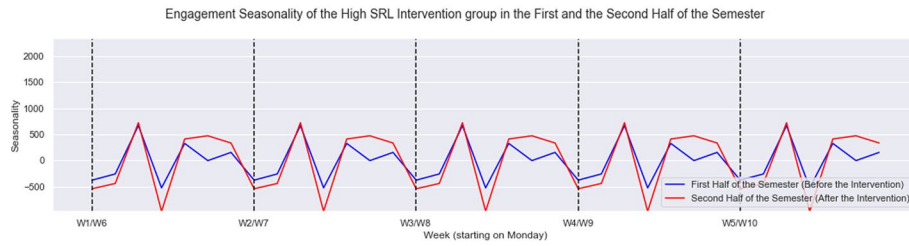
Regarding weekly patterns, seasonality compared across different SRL levels and cohorts was plotted in Fig. 4, where blue and red lines represented seasonality for the two halves of the semester, respectively. Figure 5a–d represent engagement in the order of the control cohort’s high SRL cluster, the control cohort’s low SRL cluster, the intervention cohort’s high SRL cluster, and the intervention cohort’s low SRL cluster from the top to bottom accordingly. In general, every group illustrated homologous weekly patterns for both halves of the semester with minor changes. Considering the control cohort, corresponding seasonal patterns were observed in the high and low SRL groups for both periods (Fig. 5a and b). To be specific, there was a 1-peak weekly pattern in which engagement amount was initially dropped from Mondays to reach its minimum on Thursdays and gradually increased towards the weekend, approaching its maximum amount on Sundays (blue lines) during the first half of the semester.



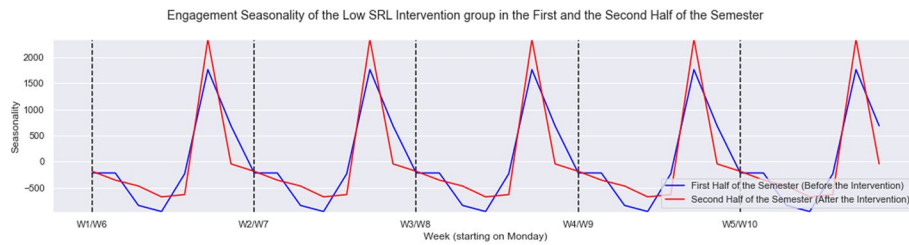
(a) The control cohort's high SRL cluster



(b) The control cohort's low SRL cluster



(c) The intervention cohort's high SRL cluster



(d) The intervention cohort's low SRL cluster

Fig. 5 Seasonal component of *AvgStrCountPerDay* compared the first-half (before the intervention) (blue line) and the second-half of the semester (after the intervention) (red line) among clusters

Table 4 The number of editing frequency per week (%) among clusters

The number of editing frequency per week (%)	Control cohort				Intervention cohort			
	High SRL		Low SRL		High SRL		Low SRL	
	H1	H2	H1	H2	H1	H2	H1	H2
No edit	31.20	42.40	22.67	49.33	38.52	37.78	44.29	30.00
Edit once per week	40.80	32.80	58.67	40.00	30.37	31.85	32.86	34.29
Edit twice or more per week	28.00	24.80	18.67	10.67	31.11	30.37	22.86	35.71

H1: First half of the semester (before the intervention)

H2: Second half of the semester (after the intervention)

Table 5 Wilcoxon signed rank test on *AvgStrCountPerDay* of SRL clusters compared before and after the intervention

<i>AvgStrCountPerDay</i>	First half of the semester (Before the intervention)		Second half of the semester (After the intervention)		Z	p	Effect size
	N (Inactive days)	Md (IQR)	N (Inactive days)	Md (IQR)			
Control cohort							
High SRL	35 (4)	822.44 (1228.36)	35 (0)	559.88 (863.36)	- 0.328	0.743	0.04
Low SRL	35 (6)	457.87 (1649.73)	35 (13)	216.60 (805.93)	- 1.820	0.069	0.22
Intervention cohort							
High SRL	35 (3)	705.81 (1225.56)	35 (2)	992.04 (1709.41)	- 2.244	0.025*	0.27
Low SRL	35 (4)	427.86 (1511.71)	35 (4)	1079.93 (1773.93)	- 0.880	0.379	0.11

*p is significant at the 0.05 level (2-tailed).

IQR = interquartile range

These similar patterns were also observed after Week 5, yet higher nuances, i.e., multi-peaks were spotted from the seasonality patterns (orange lines).

Instead, the high SRL group of the intervention cohort developed a distinctive 2-peak weekly engagement pattern for both before and after receiving the feedback intervention (Fig. 5c). The pre- and post-intervention weekly patterns showed a sharp increase in engagement from Mondays to Wednesdays, followed by a significant drop on Thursdays and a continuing increase in engagement starting from Fridays to the weekend. Even though both seasonality from the pre- and post-feedback intervention periods followed a similar pattern, engagement after the feedback revealed relatively higher fluctuation in which the minimum and the maximums were respectively lower (Thursdays) and higher (Saturdays) compared to the period before the feedback intervention. Considering the weekly engagement pattern of the intervention’s low SRL group, they demonstrated a 1-step-peak pattern for both the periods before and after receiving the feedback intervention (Fig. 5d). That is, there was a slight decrease pattern from Mondays to reaching the minimum on Thursdays, followed by a skyrocketed increase to the maximum on Saturdays before the feedback. However, after the feedback intervention, the minimum slightly shifted from Thursdays to Fridays whereas the maximum remained on the same day, yet higher variation was observed. Similar to the cohort level, the frequency of the edit per week among clusters was further analysed in Table 4. As anticipated, the high SRL group of the intervention cohort maintained their editing frequency twice or more times per week at approximately 30% before and after the feedback. The low SRL group of the intervention cohort showed an increase in their editing frequency twice or more times per week from 22.86% before the feedback to 35.71% after the feedback. The following section represents statistical test results to investigate differences in *AvgStrCountPerDay* within these four groups.

Statistical tests

Similar to the cohort level, *AvgStrCountPerDay* per cluster exhibited non-normal distribution; therefore, related-sample Wilcoxon signed-rank tests were applied for each cluster to identify within-sample differences in *AvgStrCountPerDay* for each halves of the semester (Table 5). Only one test appeared to be significant, confirming statistical differences between *AvgStrCountPerDay* of the intervention cohort’s high SRL group

Table 6 Pearson correlations between reflective scores and the seven features

	<i>Total Rev</i>	<i>AvgStr Count PerDay</i>	<i>AvgRev PerDay</i>	<i>Total Active Day</i>	<i>AvgStr Count Per Week</i>	<i>AvgRev Per Week</i>	<i>Total Active Week</i>
Reflective score	0.293**	0.186	0.210	0.353**	0.183	0.152	0.448**
<i>p</i>	0.008	0.097	0.060	0.001	0.101	0.174	< 0.001

**Correlation is significant at the 0.01 level (2-tailed)

Table 7 Independent samples tests of writing engagement features between the first and the second half of the semester (before and after the intervention) compared between two cohorts

Reflective writing features	Control cohort		Intervention cohort		<i>t</i> (74)	<i>p</i>
	M	SD	M	SD		
First Half of the Semester (Before the intervention)						
TotalRev	5880.97	8155.11	7446.71	10860.03	- 0.711	0.480
TotalActiveDay	5.45	2.39	5.79	3.70	- 0.479	0.633
TotalActiveWeek	3.79	1.23	3.21	1.32	1.977	0.052
Second Half of the Semester (After the intervention)						
TotalRev	4438.81	7828.83	8826.87	13024.04	- 1.791	0.078
TotalActiveDay	4.51	2.86	6.05	3.72	- 2.012	0.048*
TotalActiveWeek	2.97	1.26	3.41	1.29	- 1.494	0.139

**p* is significant at the 0.05 level (2-tailed)

before and after the feedback intervention, $Z = - 2.244$, $p = 0.025$, with small effect size, $r = 0.27$. In other words, there was a statistically significantly higher amount of *Avg-StrCountPerDay* before the feedback (Md = 705.81, IQR = 1225.56) compared to after the feedback (Md = 992.04, IQR = 1709.41). Apart from this group, no statistically significant difference was found in the amount of writing engagement within groups (High SRL students of the control cohort, $Z = - 0.328$, $p = 0.743$; Low SRL students of the control cohort, $Z = - 1.820$, $p = 0.069$; Low SRL students of the intervention cohort, $Z = - 0.880$, $p = 0.379$). Despite observable visual variation in time series patterns before and after the feedback, only the high SRL group of the intervention cohort showed a statistically significant increase in their reflective writing engagement after receiving the feedback intervention. It is worth noting that increases in the writing engagement median were only observed in the intervention cohort’s clusters.

Relationship between reflective writing engagement and academic performance (RQ3:

What are the relationships between students’ writing engagement behaviours, and their final grades?)

In this section, we present Pearson’s *r* correlation results to identify the relationship between the engagement features and reflective scores, followed by independent sample *t*-test results comparing engagement features between control and intervention cohorts’ first and second halves of the semester.

Correlation between writing engagement features and reflective score

From Table 6, reflective score was found to be moderately positively correlated with *TotalActiveWeek* ($r(81) = 0.448, p < 0.001$) and weakly positively correlated with *AvgStrCountPerDay* ($r(81) = 0.353, p = 0.01$) and *TotalRevision* ($r(81) = 0.293, p = 0.008$). There were no significant correlations found between reflective score and other engagement features we derived namely *AvgStrCountPerDay*, *AvgRevPerDay*, *AvgStrCountPerWeek* and *AvgRevPerWeek*.

Statistical tests on differences in writing engagement features before and after the feedback

In the first half of the semester (Before the feedback period), there were no significant differences in the numbers of *TotalRev*, $t(74) = -0.711, p = 0.480$, *TotalActiveDay*, $t(74) = -0.479, p = 0.633$, and *TotalActiveWeek* ($t(74) = 1.977, p = 0.052$) between 2 years (Table 7). However, after the feedback, the intervention cohort demonstrated statistically significantly higher number of *TotalActiveDay* ($M = 6.05, SD = 3.72$) than the control cohort ($M = 4.51, SD = 2.86$), $t(74) = -2.012, p = 0.048$, whereas no other statistically significant differences were found in *TotalRev*, $t(74) = -1.791, p = 0.078$, and *TotalActiveWeek*, $t(74) = -1.494, p = 0.139$ across cohorts.

Discussion

Although many previous studies attempted to provide analytics support on student writings, the value of content-specific analytics still appears to be limited in real-world settings, however, analytics of engagement with writing has the potential to provide value to students' reflective writing performance. In this study, the effects of a feedback intervention that combines human educators' feedback on students' reflective writing content with the analytics of students' writing behaviours (intervention cohort), compared to human educators' content-only feedback (control cohort), were investigated. More specifically, the number of edited contents per day was selected as a proxy to represent students' engagement with a reflective writing task. Two cohorts' numbers of daily edited content data were examined using time series analysis to visually observe any potential pattern differences. Based on these observations, hypothesised differences between the intervention and control cohorts were tested using statistical comparison tests. Apart from cohort comparisons, the impact of the feedback interventions on students with varying degrees of SRL competence was further investigated.

With respect to RQ1: *How does the intervention based on the feedback about students' reflective writing behaviours affect their engagement with the reflective writing task?* In line with our hypothesis, the results show that students who were provided with feedback on their writing behaviours engaged with the writing task significantly more after the combined feedback compared to the control cohort students. Despite the visually observed variation in weekly engagement patterns, both cohorts had similar amounts of engagement prior to the feedback intervention with no statistically significant difference. This suggests that presenting students with analytics of their writing engagement behaviours, in addition to educators' feedback on their writing content, can serve as an opportunity to encourage students' persistent engagement with their writing tasks. This result is aligned with previous research on the effects of behavioural feedback on engagement in other studies (Nelson et al., 2012). In addition, time series patterns demonstrated a

surge in engagement immediately after the feedback intervention. This might indicate an increase in the responsiveness and timely reactions of students to the intervention provided. Such characteristics of students are shown to correlate with a higher ability to regulate learning and higher academic performance (Suraworachet et al., 2021). Moreover, the seasonality of the two cohorts revealed anticipated engagement patterns in which the lowest interaction was spotted on Thursdays (the day when there was demand from another module in the same programme) and the highest engagement was reached during the weekend. However, these weekly patterns varied between the control and intervention cohorts in which the intervention group exhibited a bimodal weekly engagement pattern compared to the unimodal pattern in the control group. A bimodal seasonality coupling with an increase in weekly engagement frequency may refer to a higher frequency of reflection per week. It may represent students' strategy to divide the tasks into smaller units for better coping with tasks and/or an engagement pattern beyond fulfilling the minimum of the module requirement (Fredricks, 2004) which are associated with better learning outcomes (Yip, 2012). However, these interpretations require further qualitative in-depth investigations of students' experiences to be confirmed.

Regarding the RQ2: *How does the feedback intervention affect the writing task engagement of students with different SRL competence?* The low and high SRL groups of the intervention cohort showed increasing engagement with the writing tasks after receiving the engagement feedback as evident through their daily writing behaviours and the similar number of active days before and after the intervention. Conversely, the low and high SRL groups of the control cohort exhibited lower engagement in the second half of the module (Week 6–10) in which the control cohort's low SRL group distinctively showed a higher number of inactive days afterwards. These withdrawal effects after the sole content-based feedback in low SRL's control cohort are aligned with Mitchell, McMillan and Rabbani's (2019) study which showed that low self-efficacy students reported higher negative feelings or anxiety emerging from the content feedback alone. Moreover, the low SRL group of the intervention cohort was the only group that showed higher percentages of engagement with the writing task with twice or more times engagement within a week after the feedback intervention provided. These results indicate the positive impact of the writing analytics feedback intervention on students' engagement with the writing task mainly occurs through changing the routine behaviours of low SRL students while not influencing the engagement regularity of high SRL students significantly. These results further support the previous research evidence (Nelson et al., 2012) that low-performing students might benefit even more from timely feedback on their engagement behaviours. In terms of engagement quantity, despite variations in engagement patterns compared before and after the intervention, the intervention cohort's low SRL group did not show statistically significant differences in their daily engagement quantity. On the contrary, the intervention cohort's high SRL group was the only group that showed statistically significant higher engagement quantity after the intervention compared to the period before. Thus, these results indicate that the analytics feedback intervention is likely to help students with low SRL competence recognise the necessity of regular reflections, maintain their engagement with the reflective writing tasks and help them improve their academic performance while showing no detrimental effects on high SRL students.

Going back to our third research question: *What are the relationships between students' reflective writing behaviours, and their academic performance?*, significant correlations between students' final grades on their reflective writing tasks and writing engagement behaviours were observed. These indicated that the quantitative features of reflective writing behaviours concerning the quantity of students' writing (both at the daily and the weekly level) were not correlated to their final grades. Instead, there were weak to moderate size positive correlations between the students' final reflective writing scores and the regularity of their engagement with the writing tasks. This confirms previous research that the quality of students' individual writing is more related to the regularity of their reflections rather than the amount of reflective writing they produce in crammed sessions (Suraworachet et al., 2021). The value of spaced practice (Rohrer & Taylor, 2006; Sobel et al., 2011) and interleaving (Rohrer, 2012) for students have also been shown in multiple other studies from the learning sciences literature (Carpenter, 2014). We further investigated the extent to which the writing engagement behaviours changed in the feedback intervention cohort. The results showed that there was a higher regularity (measured through the observation of total active days in the reflective writing days) in students' engagement with the individual reflective writing task at a daily level in the intervention cohort after the feedback compared to the control cohort.

Limitation and future research

Finally, several limitations should be noted. First, this is not a randomised controlled trial study in which the true effects of the feedback intervention can be claimed in causal arguments. Due to the ethical and practical concerns of studying real-world teaching and learning contexts, true randomisation of the intervention group and the control group was not possible. Therefore, we opted in for a quasi-experimental set-up in which two different cohorts were used as intervention and control groups. Although the engagement behaviours of two cohorts were observed before the intervention and no statistically significant differences were detected, the contextual variations of two different cohorts might influence the results presented in this study. For this reason, and also due to the limited sample sizes in both cohorts, it is difficult to generalise the results into other contexts without further investigations. Hence, this study encourages similar future work to be conducted in the field to expand understanding of the proposed feedback intervention's effect on students' writing engagement behaviours in other contexts. Moreover, it is worth highlighting that although we used a particular group of writing behaviours as proxies to predict students' writing engagement, there are other possible proxies which might be equivalently worth exploring (i.e. average revisions, time spent, the number of writing sessions, etc.). In addition to the current methods of analysis, an additional study on students' opinions on the feedback intervention and an extended analysis of the content presented could help provide a more comprehensive picture of students' understanding of the feedback provided and its intentional impact on their engagement with the writing task at behavioural, cognitive and emotional levels.

Conclusion

This quasi-experimental study demonstrates a robust investigation of the real-world effects of an original human educator and analytics combined feedback intervention on students' writing engagement behaviours in an authentic semester-long graduate module. The intervention group consisted of students who received personalised engagement feedback with analytics during the mid-term in addition to their content feedback from educators and exhibited higher engagement during the second half of the semester after receiving the analytics feedback. Additionally, they also demonstrated higher regularity in engaging with the reflective writing task both at a daily and weekly level which significantly positively correlated with their academic performance. The combined feedback intervention was found to be more effective, especially for students with low SRL competence. This result contributes to the broader research in reflective writing support to consider coupling feedback from both cognitive and behavioural aspects to match learners' SRL levels. It is particularly significant given the context independency of the behavioural feature we engineered and the ubiquitous use of the Google Docs platform for generating the analytics feedback. However, further investigations on the longevity of the impacts, as well as their cross-context validity, should be undertaken.

Appendix

Appendix A: Email feedback

Sample email of the designed engagement feedback

Dear ,

First of all, thank you for continuing to engage with the individual reflection. You must have already received feedback on the content of your reflections as comments in the shared document. This email is a formative evaluation of your individual reflection behaviours. We hope this can help you to better regulate your reflective writing behaviours.

According to previous research on students' reflective writing behaviours, high self-regulated learners tend to 1) show systematic patterns by allocating a specific time slot for completing the reflection task, 2) react promptly to the feedback and 3) avoid postponing the work until the deadline. More information could be found here: *anonymised*

The graph attached here shows a comparison of the number of edited strings per day across five weeks between you (red), last year's cohort (blue) and this year's cohort (green). Compared to last year's cohort, your cohort are more active in the first 5 weeks. Well done for all your hard work!

In the first 5 weeks, you have been active for 3 weeks. Although there was a delay during the first few weeks, we are glad to see you became more active by mid-term. According to the graph, you have completed most of your writing in Week 5. It would be better to allocate specific time to complete the individual reflection. Writing regularly after reading or discussing may help with your individual reflection.

It is worth highlighting that the higher the number of edited contents itself might not always lead to better learning outcomes but the systematic pattern of reflective writing behaviours can.

The analytic data presented here are for reflective purposes, not part of your summative assessment. We look forward to reading more of your individual reflections.

Please let us know if you have any questions or concerns.

Best regards

Appendix B: SRL questionnaire

Questions and references			
No.	Category	Question	References
1	Goal Setting	I set standards for my assignments in a class/subject/module.	OSLQ (Barnard et al., 2009)
2	Goal Setting	I set short-term (daily or weekly) goals as well as long-term goals (monthly or for the semester).	OSLQ (Barnard et al., 2009)
3	Goal Setting	I keep a high standard for my learning in a class/subject/module.	OSLQ (Barnard et al., 2009)
4	Goal Setting	I set goals to help me manage study time for a class/subject/module.	OSLQ (Barnard et al., 2009)
5	Persistence	Regardless of whether or not I like materials in a class/subject/module, I work my hardest to learn it.	Persistence (Elliot et al., 1999)
6	Persistence	When something that I am studying gets difficult, I spend extra time and effort trying to understand it.	Persistence (Elliot et al., 1999)
7	Persistence	I try to learn all of the testable material "inside and out," even if it is boring.	Persistence (Elliot et al., 1999)
8	Persistence	I work hard to do well in a class/subject/module even if I don't like what we are doing.	Effort regulation (Pintrich et al., 1991)
9	Persistence	Even when class/subject/module materials are dull and uninteresting, I manage to keep working until I finish.	Effort regulation (Pintrich et al., 1991)
10	Persistence	When I was feeling bored, I forced myself to pay attention.	Motivation control (Warr & Downing, 2000)
11	Persistence	When my mind began to wander during a learning session, I made a special effort to keep concentrating.	Motivation control (Warr & Downing, 2000)
12	Persistence	I increased my effort when the material did not really interest me.	Motivation control (Warr & Downing, 2000)
13	Persistence	I pushed myself even harder when I began to lose interest.	Motivation control (Warr & Downing, 2000)
14	Persistence	Whenever I lost interest in my work, I made a special effort to pay attention.	Motivation control (Warr & Downing, 2000)
15	Effort	I usually spent more time than the requirements of my class/subject/module.	Effort (Adapted from Fisher & Ford, 1998)
16	Effort	I usually provide extra effort in my class/subject/module.	Time on task (Adapted from Brown, 2001)
17	Self-efficacy	I'm certain I can understand the basic concepts in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
18	Self-efficacy	I believe I will receive an excellent grade in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
19	Self-efficacy	I'm certain I can understand the most difficult material presented in the readings for any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
20	Self-efficacy	I'm confident I can learn the basic concepts taught in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
21	Self-efficacy	I'm confident I can understand the most complex material presented by the instructor in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
22	Self-efficacy	I'm confident I can do an excellent job on the assignments in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
23	Self-efficacy	I expect to do well in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)

Questions and references

No.	Category	Question	References
24	Self-efficacy	I'm certain I can master the skills being taught in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)
25	Self-efficacy	Considering the difficulty of this module, the teacher, and my skills, I think I will do well in any class/subject/module.	Self-efficacy for learning and performance (Pintrich et al., 1991)

Abbreviations

SRL	Self-regulated learning
NLP	Natural language processing
RQ	Research question
GS	Goal-setting
E	Effort
SE	Self-efficacy
P	Persistence

Acknowledgements

We would like to thank DUTE 2020/2021 and 2021/2022 students for granting permissions to collect data for this study and also thank Prof. Yannis Dimitriadis for his valuable comments on an earlier version of the manuscript.

Author contributions

All authors contributed in constructing the materials, intervention design for the study, data analysis and paper writing. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

Ethical approval was received from the institutional ethics review committee.

Competing interests

The authors declare that they have no competing interests.

Received: 3 June 2022 Accepted: 13 October 2022

Published online: 03 January 2023

References

- Aronson, L., Niehaus, B., Hill-Sakurai, L., Lai, C., & O'Sullivan, P. S. (2012). A comparison of two methods of teaching reflective ability in Year 3 medical students: Comparison of teaching methods for reflection. *Medical Education*, *46*(8), 807–814. <https://doi.org/10.1111/j.1365-2923.2012.04299.x>.
- Bodily, R., Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. Vancouver British Columbia Canada: ACM, pp. 309–318. <https://doi.org/10.1145/3027385.3027403>
- Boutet, I., Vandette, M. P., & Valiquette-Tessier, S.-C. (2017). Evaluating the implementation and effectiveness of reflection writing. *The Canadian Journal for the Scholarship of Teaching and Learning*. <https://doi.org/10.5206/cjsotl-rceace.2017.1.8>.
- Bridgeman, B., & Ramineni, C. (2017). Design and evaluation of automated writing evaluation models: Relationships with writing in naturalistic settings. *Assessing Writing*, *34*, 62–71. <https://doi.org/10.1016/j.asw.2017.10.001>.
- Carpenter, S.K. (2014). Spacing and interleaving of study and practice. In: Applying science of learning in education: Infusing psychological science into the curriculum. Washington, DC, US: Society for the Teaching of Psychology, pp. 131–141.
- Connor-Greene, P. A. (2000). Making connections: evaluating the effectiveness of journal writing in enhancing student learning. *Teaching of Psychology*, *27*(1), 44–46. https://doi.org/10.1207/S15328023TOP2701_10.
- Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the research writing tutor during revision. *Journal of Writing Research*, *12*(1), 187–232. <https://doi.org/10.17239/jowr-2020.12.01.07>.
- Crossley, S. A., Kim, M., Allen, L., & McNamara, D. (2019). Automated summarization evaluation (ASE) using natural language processing tools. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (pp. 84–95). Springer International Publishing.

- Cukurova, M. (2019). Learning analytics as AI extenders in education: Multimodal machine learning versus multimodal learning analytics. In: Artificial intelligence and adaptive education. Vol. 2019. AIAED.
- Cukurova, M., Bennett, J., & Abrahams, I. (2018). Students' knowledge acquisition and ability to apply knowledge into different science contexts in two different independent learning settings. *Research in Science & Technological Education*, 36(1), 17–34. <https://doi.org/10.1080/02635143.2017.1336709>.
- Cukurova, M., Kent, C., & Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, 50(6), 3032–3046.
- Dekker, H., Schönrock-Adema, J., Snoek, J. W., van der Molen, T., & Cohen-Schotanus, J. (2013). Which characteristics of written feedback are perceived as stimulating students' reflective competence: an exploratory study. *BMC Medical Education*, 13(1), 94. <https://doi.org/10.1186/1472-6920-13-94>.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>.
- Gibson, A., Aitken, A., Sándor, A., Shum, S.B., Tsingos-Lucas, C., Knight, S. (2017). Reflective writing analytics for actionable feedback. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. Vancouver British Columbia Canada: ACM, pp. 153–162. <https://doi.org/10.1145/3027385.3027436>
- Graham, S., & Harris, K. R. (1994). The role and development of self-regulation in the writing process. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (Vol. 1, pp. 203–228). Lawrence Erlbaum Associates Inc.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Hyndman, R.J., Athanasopoulos, G. (2018). Forecasting: Principles and practice. 2nd. OTexts. OTexts.com/fpp2.
- Iraj, H., Fudge, A., Faulkner, M., Pardo, A., Kovanović, V. (2020). Understanding students' engagement with personalised feedback messages". In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. Frankfurt Germany: ACM, pp. 438–447. <https://doi.org/10.1145/3375462.3375527>
- Jivet, I., Wong, J., Scheffel, M., Torre, M.V., Specht, M., Drachslar, H. (2021). Quantum of Choice: How learners' feedback monitoring decisions, goals and self-regulated learning skills are related. In: LAK21: 11th international learning analytics and knowledge conference. pp. 416–427.
- Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., Dawson, S. (2018). Understand students' self-reflections through learning analytics. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. Sydney New South Wales Australia: ACM, pp. 389–398. <https://doi.org/10.1145/3170358.3170374>
- Liu, M., Kitto, K., & Shum, S. B. (2021). Combining factor analysis with writing analytics for the formative assessment of written reflection. *Computers in Human Behavior*, 120, 106733. <https://doi.org/10.1016/j.chb.2021.106733>.
- Luckin, R. (2018). Machine Learning and Human Intelligence: The future of education for the 21st century. ERIC.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1: 281–197.
- McIntosh, P. (2010). *Action research and reflective practice: Creative and visual methods to facilitate reflection and learning*. Routledge.
- Mitchell, K. M., McMillan, D. E., & Rabbani, R. (2019). An exploration of writing self-efficacy and writing self-regulatory behaviours in undergraduate writing. *The Canadian Journal for the Scholarship of Teaching and Learning*. <https://doi.org/10.5206/cjsotl-rcacea.2019.2.8175>.
- Nelson, K. J., Quinn, C., Marrington, A., & Clarke, J. A. (2012). Good practice for enhancing the engagement and success of commencing students. *Higher Education*, 63(1), 83–96. <https://doi.org/10.1007/s10734-011-9426-y>.
- Neto, V., Rolim, V., Pinheiro, A., Lins, R. D., Gašević, D., & Mello, R. F. (2021). Automatic content analysis of online discussions for cognitive presence: A study of the generalizability across educational contexts. *IEEE Transactions on Learning Technologies*, 14(3), 299–312. <https://doi.org/10.1109/TLT.2021.3083178>.
- Öncel, P., Flynn, L.E., Sonia, A.N., Barker, K.E., Lindsay, G.C., McClure, C.M., McNamara, D.S., Allen, L.K. (2021). Automatic student writing evaluation: investigating the impact of individual differences on source-based writing. In: LAK21: 11th International Learning Analytics and Knowledge Conference. ACM, pp. 620–625. <https://doi.org/10.1145/3448139.3448207>
- Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. *Journal of Educational Psychology*, 49(4), 173–181. <https://doi.org/10.1037/h0041940>.
- Plak, S., van Klaveren, C., Cornelisz, I. (2022). Raising student engagement using digital nudges tailored to students' motivation and perceived ability levels. *British Journal of Educational Technology*. in press.
- Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, 24(3), 355–367.
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20(9), 1209–1224. <https://doi.org/10.1002/acp.1266>.
- Rozental, L., Meitar, D., & Karnieli-Miller, O. (2021). Medical students' experiences and needs from written reflective journal feedback. *Medical Education*, 55(4), 505–517. <https://doi.org/10.1111/medu.14406>.
- Ryan, M. (2013). The pedagogical balancing act: Teaching reflection in higher education. *Teaching in Higher Education*, 18(2), 144–155. <https://doi.org/10.1080/13562517.2012.694104>.
- Royce Sadler, D. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550. <https://doi.org/10.1080/02602930903541015>.
- Shibani, A. (2020). Constructing automated revision graphs: A novel visualization technique to study student writing. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (pp. 285–290). Springer International Publishing.
- Shibani, A., Knight, S., Shum, S. B. (2019). Contextualizable learning analytics design: A generic model and writing analytics evaluations. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge. ACM, pp. 210–219. <https://doi.org/10.1145/3303772.3303785>
- Shin, Y. (2017). *Time series analysis in the social sciences: The fundamentals*. University of California Press. <https://doi.org/10.1525/9780520966383>.

- Shum, S. B., Knight, S., McNamara, D., Allen, L., Bektik, D., Crossley, S. (2016). Critical perspectives on writing analytics. In: LAK16: 6th International Learning Analytics and Knowledge Conference. ACM Press, pp. 481–483. <https://doi.org/10.1145/2883851.2883854>
- Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, 137(3), 421–442. <https://doi.org/10.1037/a0022777>.
- Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, 25(5), 763–767. <https://doi.org/10.1002/acp.1747>.
- Stewart, L. G., & White, M. A. (1976). Teacher comments, letter grades, and student performance: What do we really know? *Journal of Educational Psychology*, 68(4), 488–500. <https://doi.org/10.1037/0022-0663.68.4.488>.
- Strong, R. W., Silver, H. F., & Perini, M. J. (2001). Making students as important as standards. *ASCD Educational Leadership*, 59(3), 56–61.
- Suraworachet, W., Villa-Torrano, C., Zhou, Q., Asensio-Pérez, J. I., Dimitriadis, Y., & Cukurova, M. (2021). Examining the relationship between reflective writing behaviour and self-regulated Learning competence: A time-series analysis. In T. De Laet, R. Klemke, C. Alario-Hoyos, I. Hilliger, & A. Ortega-Aranz (Eds.), *Technology-Enhanced Learning for a Free, Safe, and Sustainable World* (Vol. 12884, pp. 163–177). Springer International Publishing. https://doi.org/10.1007/978-3-030-86436-1_13.
- Thorpe, K. (2004). Reflective learning journals: From concept to practice. *Reflective Practice*, 5(3), 327–343. <https://doi.org/10.1080/1462394042000270655>.
- Türkay, S., Seaton, D., Ang, A. M. (2018). Itero: A revision history analytics tool for exploring writing behavior and reflection. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, pp. 1–6. <https://doi.org/10.1145/3170427.3188474>.
- Vytasek, J. M., Patzak, A., & Winne, P. H. (2020). Analytics for student engagement. In M. Virvou, E. Alepis, G. A. Tsihrintzis, & L. C. Jain (Eds.), *Machine Learning Paradigms Advances in Learning Analytics* (pp. 23–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-13743-4_3.
- Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment & Evaluation in Higher Education*, 35(5), 519–533. <https://doi.org/10.1080/02602930903512909>.
- Winograd, B. A., Dood, A. J., Moeller, R., Moon, A., Gere, A., Shultz, G. (2021). Detecting high orders of cognitive complexity in students' reasoning in argumentative writing about ocean acidification. In: LAK21: 11th International Learning Analytics and Knowledge Conference. ACM, pp. 586–591. <https://doi.org/10.1145/3448139.3448202>.
- Yip, M. C. W. (2012). Learning strategies and self-efficacy as predictors of academic performance: A preliminary study. *Quality in Higher Education*, 18(1), 23–34. <https://doi.org/10.1080/13538322.2012.667263>.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81, 329–339. <https://doi.org/10.1037/0022-0663.81.3.329>.
- Zimmerman, Barry J., & Risemberg, Rafael. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22(1), 73–101. <https://doi.org/10.1006/ceps.1997.0919>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
