Effects of instructional conditions and experience on the adoption of a learning tool

Dragan Gasevic[1,2] , Negin Mirriahi[3], Shane Dawson[4], Srecko Joksimovic[1]

[1]Moray House School of Education, University of Edinburgh, United Kingdom,
Holyrood Road, Edinburgh EH8 8AQ, United Kingdom

[2]School of Informatics, University of Edinburgh, United Kingdom,
10 Crichton Street, Edinburgh, Midlothian EH8 9LE, United Kingdom

[3]Learning and Teaching Unit, University of New South Wales,
Level 4, Mathews Building (F23), Via Gate 11, Botany Street, , Sydney, NSW 2052, Australia

[4]Teaching Innovation Unit, University of South Australia,
160 Currie St, Adelaide, SA 5000, Australia

Abstract: This paper presents the results of a natural experiment investigating the effects of instructional conditions and experience on the adoption and sustained use of a learning tool. The experiment was conducted with undergraduate students, enrolled into four performing art courses (N=77) at a research intensive university in Canada. The students used the video annotation software CLAS for course-based self-assessment on their performances. Although existing research offers insights into the factors predicting students' intentions of accepting a learning tool, much less is known about factors that affect actual adoption and sustained tool use. The study explored the use of CLAS amongst undergraduate students in four courses across two consecutive semesters. Trace data of students' tool use, graph-based measures of metacognitive monitoring, and text cohesion of video annotations were used to estimate the volume of tool use and the quality of the learning strategy and learning products created. The results confirmed that scaffolding (e.g., graded activity with instructional feedback) is required to guide students' initial tool use, although scaffolding did not have an independent significant effect on the quantity of tool use. The findings demonstrated that the use of the tool is strongly influenced by the experience an individual student gains from scaffolded conditions. That is, the students sustained their use of the learning tool in future courses even when the tool use was not graded nor was instructional feedback provided. An important implication is that students' tool use is not solely driven by motivation – rather, it is shaped by instructional conditions and experience with the tool use.

Keywords: learning technology adoption; instructional scaffolding; self-regulated learning; learning analytics

1 Introduction

Despite the many reported benefits of technology for facilitating student learning and engagement (Chen, Lambert, & Guidry, 2010; López-Pérez, Pérez-López, & Rodríguez-Ariza, 2011), studies have identified that a great majority of students (above 60%) can be classified as limited learning technology users (Lust, Elen, & Clarebout, 2013; Lust, Juarez Collazo, Elen, & Clarebout, 2012). The observed limited use is not simply a function of poor course design. The authors also noted a lack of student engagement with technologies even when learning tools are specifically embedded into course designs that follow pedagogically sound and empirically validated principles (Lust et al., 2013, 2012). Essentially, there is an educational challenge to first motivate students to accept the learning tool and second, to sustain their use of it.

Much research has been undertaken to understand the conditions for promoting student acceptance and long term adoption of learning tools (Cheung & Vogel, 2013; Edmunds, Thorpe, & Conole, 2012; Escobar-Rodriguez & Monge-Lozano, 2012). Most prominent in this area has been the technology acceptance model (TAM). The TAM was first proposed by Davis (1989) and comprises of two primary factors that are perceived to contribute to technology adoption: perceived ease of use and perceived usefulness (Sánchez & Hueros, 2010). The explanatory power of this model is further extended when additional constructs are incorporated such as self-efficacy, enjoyment, and learning goal orientation. These additional constructs provide additional explanatory power beyond that of TAM to better understand student use of technical systems (Yi & Hwang, 2003). However, while there is much to learn from these studies, the adopted constructs tend to explain factors influencing students' intentions to accept using learning tools, rather than their actual adoption (Clarebout, Elen, Collazo, Lust, & Jiang, 2013). Moreover, there is limited understanding of the conditions and pedagogical approaches that sustain students' use of educational technology, especially when the use of a certain tool is optional and not assessed.

1.1 Learning Tool Use as Self-Regulated Learning

Contemporary research investigating student use of education technologies is increasingly situated within the context of self-regulated learning (SRL) (Trevors, Duffy, & Azevedo, 2014). As originally suggested by Azevedo (2005), SRL provides a robust theoretical framework to inform the study of technology-enabled learning (in computer-based learning environments). In this paper, we have adopted SRL as a framework to understand the conditions that sustain students' use of a learning technology. In particular, we incorporated Winne and Hadwin's model of self-regulated learning (Winne, 2006; Winne & Hadwin, 1998). This model considers five elements: conditions, operations, products, evaluations and standards (COPES) – that collectively influence self-regulatory processes of learning (Winne, 1996). According to the COPES model, learners use tools (cognitive, digital or physical) to *operate* on raw information (e.g., watching video recordings of a lecture) in order to construct *products* of their learning (e.g., recall of information introduced in the video recordings). To regulate their learning process, students *evaluate* the products of their learning (e.g., quality of their recall) and the effectiveness of their learning strategies according to internal (e.g., whether video watching results in satisfying information recall within the time budgeted for learning) or external *standards* (e.g., whether they received a passing mark on a quiz that accompanied the video). Consistent with modern educational psychology, the Winne and Hadwin model (1998) deems learners as active agents in the learning process. As active and constructive participants, learners monitor their learning and choose the tools they are going to adopt and the standards they will follow to evaluate the products of their learning (Winne, 1996) as a part of their metacognitive control and monitoring. This decision making process is based on internal (e.g., experience with tools, epistemic beliefs, and prior knowledge) and external (e.g., tasks mandating the use of a tool) *conditions* (Winne, 2011; Winne & Hadwin, 1998). Thus, certain conditions are required for learners to select and regularly use a particular learning tool.

As previously posited by Winne (2006) and empirically validated in several studies conducted by Clarebout et al. (2013), there are generally four main conditions that influence learners' decisions regarding tool selection and use. First, learners need to be aware of the value of the tool and its availability in their learning environment. Second, learners need to recognize that the tool can be applied to the specific task at hand. Third, even if the learners are cognizant of the benefits of the tool for the assigned task, they need to have sufficient skills to utilise the selected tool effectively. Finally, learners need sufficient motivation to invest the time necessary to use the tool. These conditions can explain why certain tools are not always adopted by learners despite having a positive prior experience (Sarfo, Elen, Clarebout, & Louw, 2010). In this context, Clarebout et al. (2013) proposed that learners first need to have some prior experience with a tool before their conceptions of it can be used as a predictor of future use.

1.2 Instructional Conditions for the Sustained Use of a Learning Tool

In this paper, we accept and extend Clarebout et al.'s (2013) proposition to further suggest that for a tool to have sustained use, learners must first be exposed to the learning tool; and second, gain a level of proficiency in its use. In the absence of any previous experience with a tool or if a learner is only familiar with it in alternate contexts (i.e., transfer across contexts can be challenging (Perkins, 1985)), it is unlikely that learners will be able to recognize the value of the tool. That is, two of the conditions suggested by Winne (2006), value and awareness of a tool, are not met. *We posit that in order to meet these conditions and facilitate learners' ability to effectively use a tool, a level of scaffolding is required to guide learners in their initial use of the tool and how it can be applied to a particular learning task* (Azevedo & Hadwin, 2005; Beed, Hawkins, & Roller, 1991). The effects of the instructional conditions on learners' decision making and technology acceptance is well-documented in the literature (Azevedo, Moos, Greene, Winters, & Cromley, 2008; Cho & Kim, 2013; Garrison & Cleveland-Innes, 2005; McGill & Klobas, 2009; Trigwell, Prosser, & Waterhouse, 1999). Based on this literature, we suggest that *a learner's initial experience with a tool, should:*

- *have at least one task where the use of the tool is required to complete a course task and the task assessed (mandated in the course design); and*
- *be accompanied with guidance and feedback on how the student can use the tool effectively in order to complete the assigned learning tasks.*

To establish a sustained level of use of a particular tool, additional conditions need to be met. First, as recognized by the research on educational technology acceptance and illustrated by TAM, learners need to perceive the tool as easy to use and useful in order to preserve their intention to use the tool in the future (Sánchez & Hueros, 2010). This is particularly important when the use of a tool is optional. In other words, a tool must be intuitive to use without an extensive learning curve or extraneous cognitive load that could create an added layer of complexity impeding a student's ability to complete an assigned task (Devolder, van Braak, & Tondeur, 2012; Kirschner, Sweller, & Clark, 2006).

Second, learners need to be able to transfer the use of a tool to new contexts (Salomon & Perkins, 1989). As suggested by Winne (2006), learners need to be able to recognize when a tool can be appropriately applied to complete a new task. If a student's previous experience with a tool is similar to the new task, they are more likely to adopt the tool again to complete the requested task. However, if the context is significantly different, then the student's selection of the same technology is less likely. Winne (2006) describes this as a mediation deficiency. That is, a situation when learners are "unable to assemble bridging information between tools and to-be-learned information" (Winne, 2006, p. 7). The study reported in this paper focuses on the sustained tool use in similar tasks rather than on the transfer across different contexts.

1.3 Measurement of the Use of Educational Technology

Studies on the adoption and effects of educational technology on self-regulated learning have primarily been based on measures of learner *operations*, as defined in the COPES model (Winne, 1996, 2006). These measures have tended to rely on learners' self-reports of their perceptions; use, and degree of use of a particular tool or learning approach (Clarebout et al., 2013; Lust et al., 2012; Sánchez & Hueros, 2010; Yi & Hwang, 2003). While SRL studies have often relied on self-report methodologies (e.g., think aloud protocols and surveys), alternate options are rapidly emerging such as the analysis of captured trace data from learners' interactions with educational technology (Azevedo, 2015). The analysis of trace data to inform learning, teaching, and research has recently amplified due to a growing interest in the fields of learning analytics and educational data mining (Baker & Yacef, 2009; Gašević, Dawson, & Siemens, 2015; Gasevic, Mirriahi, Long, & Dawson, 2014; Siemens & Gašević, 2012). Typical measures derived from a learner's trace data include the frequency and time spent on the various operations performed with learning technologies. For example, in the context of self-regulated learning, Cho and Shen (2013) found that a student's ability to regulate social interaction with others (Cho & Jonassen, 2009) was a significant predictor of the amount of time spent online in a learning management system (LMS). Similarly, Jeske, Backhaus, and Stamov Roßnagel (2014) noted that trace-based variables, such as time spent and frequency of navigation through a sequence of resources in an online lesson, were sound proxies of motivation and self-regulation strategies that mediated the association between learning experience and test performance in an controlled experiment.

To date, there has been limited research that has investigated the effects of conditions associated with the COPES model (Winne, 1996, 2006) on technology use and acceptance. While internal conditions have been studied, such as self-efficacy, goal-orientation, and prior knowledge (Cho & Shen, 2013; Clarebout et al., 2013; Jeske et al., 2014), the effects of external conditions have received limited research attention. As posited in the COPES model (Winne, 1996, 2006), external conditions (e.g. grading of learners' self-assessments or sharing the self-assessments with peers) can have significant effects on the standards the learners use to evaluate the products of their learning and the learning strategies they chose to apply. While, the quantity of the products of learning and the adopted learning strategies may remain consistent, the quality of learning can be markedly different. For example, in an online software engineering course, Gašević, Adesope, Joksimović, and Kovanović (2015) demonstrated that the quality of a learner's discourse (operationalized as cognitive presence) significantly improved after changes to the instructional design and resources had been made to include scaffolding learner participation in a discussion forum. However, the authors noted that the quantity of the discussion remained at the same level as previous course iterations. Similarly, Kuhn (1995) suggested that learners do not increase their usage of a newly acquired learning strategy[1], but rather apply this strategy in a more effective manner. In other words, when a strategy is effectively applied, the quantity remains consistent while the quality of the learning product increases (Malmberg, Järvelä, & Kirschner, 2014). Hence, we posit that:

i) *the instructional conditions provide learners with an opportunity to experience an educational technology (or its tools). This experience influences a learner's motivation to use this technology in future similar learning contexts;*

ii) *the instructional conditions influences both the quality of learning products created and how the operations (i.e., strategy) manifest.*

1.4 Research Questions

To address the propositions outlined above, this paper reports on the results of an empirical study that aimed to address the following research questions:

RQ1. What is the effect of the instructional conditions at the course level (assessed vs. non-assessed) on students' extent of use of a learning tool in terms of count of annotations produced, quality of learning products created, and learning strategy followed?

RQ2. What is the effect of a students' prior experience with a learning tool on their future adoption in terms of counts of annotations produced, quality of learning products created, and learning strategy followed under different instructional conditions?

These questions aim to investigate the effects of different instructional conditions and experience on the three main dimensions: i) count of annotations created, ii) the quality of learning products created, and iii) learning strategy adopted in the use of video annotation software for self-assessment purposes. Therefore, both research questions are operationalized according to these three research dimensions and the results section is organized accordingly.

2 Method

2.1 Study Setting

The research design can be described as a natural experiment (Dunning, 2012). This approach was driven by situating the study within the context of the courses available for student enrolment at a research-intensive higher education institution in North America. As such, the researchers had no

[1] According to Malmberg, Järvelä, & Kirschner (2014, p. 4), a learning strategy is defined as "a coordinated set of study tactics that are directed by a learning goal, and that are aimed at acquiring a new skill or gaining understanding (Alexander, Graham, & Harris, 1998; Weinstein, 1988; Winne, 2001; Zimmerman, 1998)".

control over the experimental assignment of the study participants. Rather, the study was conducted in an ecologically valid setting whereby the assignment to the experimental conditions was performed through the participants' enrolment in the courses used in the study. The trace data logged by the video annotation software called the Collaborative Lecture Annotation System (CLAS) (Mirriahi & Dawson, 2013; Risko, Foulsham, Dawson, & Kingstone, 2013), were incorporated for further analysis. The experimental conditions were determined by the instructional conditions associated with each of the courses involved in the study.

2.2 Materials

The study participants used CLAS, a web-based application for annotating videos of student performances. The design of CLAS extends other previously established video annotation software such as Microsoft Research Annotation System (MRAS) (Bargeron, Gupta, Grudin, & Sanocki, 1999), Media Annotation Tool (MAT) (Colasante & Fenn, 2009), and Digital Video Digital University (DiViDu) (Hulsman, Harmsen, & Fabriek, 2009). Informed by the design of prior video tools and associated research in this domain resulted in the CLAS software being perceived by learners as easy to learn, easy to use, and useful for their studies (Risko et al., 2013). This validation by Risko et al. of the software's perceived ease of use, ease of learning and perceived usefulness addressed an important proposition voiced in the theoretical background for this study.

In terms of user functionality, CLAS has *two forms of annotation features*: i) *time-stamped annotations* that offer students and instructors with opportunities to create time-stamped notes that are associated to a specific part of a video; these notes can be accessed later for review (Dawson, Macfadyen, Risko, Foulsham, & Kingstone, 2012; Risko et al., 2013); ii) *general annotations* that are not associated to any specific part of a video, but allow users to post a general note or summary of the video (ibid). Both types of annotations can be either private or collaborative offering an opportunity to share annotations among peers. Furthermore, CLAS has a feature for visualizing the position of time-stamped annotations. CLAS has additional features for performing operations on the video such as pause playing, resume playing, rewind, and fast forward – as is common for contemporary video players. All these operations are recorded by CLAS in log files and thus, each operation performed in CLAS has time-stamped trace data that can be used for research (Mirriahi & Dawson, 2013). CLAS does not allow for direct downloading of the videos. Hence, students can only view and annotate the videos while using the CLAS software. This increases the accuracy of the trace data collected by CLAS, as users cannot interact with the videos used in the specific courses outside of the educational software.

2.3 Procedures

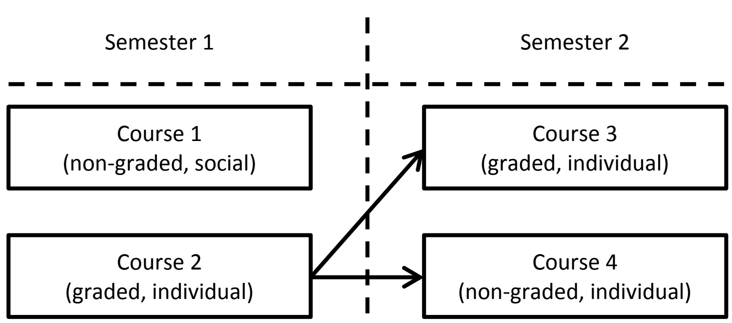
The study included four undergraduate courses in the performing arts discipline offered in the 2012/2013 academic year. The first two courses (Course 1 and Course 2) were offered in the first semester (Fall 2012), while the other two courses (Course 3 and Course 4) were offered in the second semester (Winter 2013). In all four courses, video recordings of students' own performances were available for viewing and annotating through the CLAS software. In Course 1, the videos were of group performances and there was no graded requirement for students to make an annotation or general comment. The videos in the three other courses instead focused on students' individual performances.

That is, for Course 1, all enrolled students annotated the same three videos (i.e., three group performances). Further, as this was a group-based activity, the students had an option to share their annotations with their peers. While the use of CLAS in Course 1 was optional, in Course 2 students' use of CLAS was directly assessed (i.e. a course requirement). Students had access to review four video recordings of their individual performances through the CLAS software.

Of the four courses, general annotations on three of them were graded and instructor feedback provided. For the fourth video, formative feedback on the general annotation was provided only. Overall, students made time-stamped annotations and one general annotation per video summarizing their overall reflections on their performance. Furthermore, Course 2 was a prerequisite for both Course 3 and Course 4 that were offered in the following semester. The requirements and instructional conditions for the use of CLAS in Course 3 were identical to those of Course 2 including graded general annotations. The only difference was that Course 3 included an additional video recorded performance for students to view, annotate, and submit a general reflective summary on CLAS. Course 4 was similar to Courses 2 and 3 in terms of the number of video recordings of the individual performances (i.e., four video recordings were accessible through CLAS). However, for Course 4, students did not receive a grade nor formative feedback on their general self-reflective annotations posted in the CLAS system. The course requirements and relationship between the courses and the non-graded/graded general annotations are specified in Table 1 and Figure 1.

Table 1. The numbers of student performance videos and the type of performance recorded in the four courses

Course	**Number of videos requested to annotate**	**Type of performance**
Course 1	3	Group
Course 2	4	Individual
Course 3	5	Individual
Course 4	4	Individual



Labels of the groups of the students created based on the courses they were enrolled in:

Course 1a – students who took course 1 but did not take Course 2 (n=23)
Course 1b – students who took Courses 1 and 2 (n=8)
Course 2a – students who took Course 2, but did not take course 1 (n=32)
Course 2b – students who took Courses 1 and 2 (n=8)
Course 2c – students who took Course 2, but did not take Course 3 (n=22)
Course 2d – students who took Course 2 and 3 (n=18)
Course 2e – students who took Course 2, but did not take Course 3 (n=29)
Course 2f – students who took Courses 2 and 4 (n=11)
Course 3a – students who took Course 3, but did not take Course 2 (n=10)
Course 3b – students who took Courses 3 and 2 (n=18)
Course 4a – students who took Course 4, but did not take Course 2 (n=9)
Course 4b – students who took Courses 4 and 2 (n=11)

Figure 1. The courses included in the study and the groups of students formed based on their enrolment in individual courses

Throughout the paper, we refer to the notion of instructional conditions whereby we distinguish between non-graded and graded conditions. In the graded condition, the instructors looked at the specificity of goals set by the students in their reflections. Specificity of goals is an indicator of students' recognition of the points to be improved upon in their future work. An example of the sentences of specific goal that instructors expected to see is given in this quote extracted from the students' annotations: "*I think for my next lab my goals shall be to try to make eye contact with everyone at least once.*" An example of a less specific reflection would be a simple observation of the own behavior without any specific goal set for the future work: "*I still continue to have problems with making eye contact.*" Moreover, instructors provided students with formative feedback on their performance and reflections in the graded instructional condition. For example, in cases where students missed something in their reflections, the instructor would provide feedback of this type:

> "*You are most successful when you are truly assertive in your music making. The beginning of the lab, you did not really have a clear picture of the tempo that you wanted. In conducting recits, you have to be super clear in exactly what you want and lead the ensemble. If you just beat time, it won't be successful.*
>
> *Part of this leadership comes from having a very clear picture of exactly what you want."*

In the non-graded condition, students did not receive grades nor formative feedback from their instructors.

2.4 Sample

The students enrolled in Courses 1-4 were included in the study. All the students were already enrolled in a degree program in performing arts directly linked to the courses included in the study. The sample had 77 unique students (42 female). The average age at the time of enrolment was 22.1 with standard deviation of 2.82. Each course had a different number of students enrolled: Course 1 (N=31), Course 2 (N=40), Course 3 (N=28), and Course 4 (N=20). Since the study was a natural experiment where students had the option to enroll into any course, as per the university program regulations, some students were enrolled in more than one course included in the study. The numbers of students enrolled in Courses 1-4 and their overlaps are outlined in Figure 1.

In order to account for a possible confounding effect, we report the students' grade point averages (GPAs) as they are commonly used as proxies of students' ability and predictors of future performance (Elias & MacDonald, 2007; Grove, Wasserman, & Grodner, 2006). The GPA values could also indicate the differences in skills for self-regulated learning, as typically higher academic performance is associated with higher skills for self-regulated learning (Greene & Azevedo, 2009). Table 2 reports the results of the comparison between the groups of students in the study and their grade point averages (GPAs) at the end of the academic year. Since the study was conducted as a natural experiment, the control for important confounders was not possible in the experimental assignment. The only significant difference identified was between Course 2c and Course 2d. Students with a higher GPA in Course 2 were more likely to enroll in Course 3 (Table 2). This could potentially confound the comparisons between students in groups Course 2c and Course 2d, i.e., between the students within the same experimental condition – the first experience with the graded learning tool use. However, this had no effect on the investigation of research question RQ2 where these groups are investigated. No significant difference between those students (Course 2d) and other students in Course 3 (Course 3b) were observed in the GPA values, and thus, equivalency is preserved with respect to research question RQ2 that investigated sustained tool use under the same instructional conditions in two courses (e.g., graded general annotations in the first and subsequent courses).

Table 2. The comparison of the GPA values between the identified groups in the study.

A	B	C	D	A vs. B	A vs. B
				B vs. C	C vs. D
Course 1a	Course 1b	Course 2a	Course 2b		
78.80 (75.00, 89.05)	85.33 (81.38, 88.72)	81.62 (72.35, 85.74)	85.33 (81.38, 88.72)	U=300.50, z=-.646, p=.519, r=-0.09	U=105.50, z=1.05, p=.294, r=.19
				N/A*	U=176.50, z=1.64, p=.101, r=.26
Course 2c	Course 2d	Course 3a	Course 3b		
75.35 (66.43, 84.83)	85.57 (81.96, 86.88)	80.37 (75.36, 88.88)	85.57 (81.96, 86.88)	U=142.00, z=1.30, p=.193, r=.23	U=310.00, z=3.05, p=.002, r=.48
				N/A*	U=111.00, z=1.01, p=.314, r=.19
Course 2e	Course 2f	Course 4a	Course 4b		
83.10 (72.60, 86.13)	84.70 (80.50, 87.10)	78.05 (70.60, 88.00)	84.70 (80.50, 87.10)	U=92.00, z=-.38, p=.704, r=-.06	U=178.00, z=.58, p=.565, r=.09
				N/A*	U=47.00, z=.77, p=.441, r=.18

Legend: * Comparison of the students in repeated measures could not be done, as those were the same students with a single GPA value (prior to entering to the academic year in which all the courses were offered).

2.5 Variables

2.5.1 Independent variables

Instructional conditions and experience (both binary variables) were used as independent effects, i.e., fixed effects according to the terminology of the method used in our analyses (c.f., Section 2.6). Student enrollment into Course 2 and Course 3 represents the *graded* instructional condition, whereby Courses 1 and 4 represent the *non-graded* instructional condition. Enrollment and completion of Course 2 was an indicator of experience with the tool gained in the graded condition with formative feedback. That is, the students who moved from Course 2 to either Course 3 or Course 4 had prior experience with the tool. Otherwise, all other students in all four courses were considered not to have any prior tool experience. It should be noted that some students completed Course 2 in a previous year when the video annotation software was not used. These learners (according to Figure 1, those are the students in groups Course 3a and Course 4a) were considered without experience in Courses 3 and 4. The connection between the distribution of the students according to the fixed effects (instructional conditions and experience) and the groups identified in Figure 1 is shown in Table 3.

Table 3. The connection between the distribution of the students according to the fixed effects (instructional conditions and experience) and the groups identified in Figure 1

	Instructional Conditions	
	Graded	**Non-graded**
Experience	Course 3b	Course 4b
No experience	Course 2a-e, Course 3a	Course 1a-b, Course 4a

2.5.2 Dependent variables

The following three groups of dependent variables were used in the study.

Operations on video. Counts of annotations created by the students are used as the primary dependent variable to investigate our research questions. This is due to the fact that the tool used in the study (CLAS) was designed for video annotation and the amount of its use is primarily focused on its main functionality – creating video annotations; i.e., we used the variable that represented the count of video annotations created by a student in a course. Given that the courses differed in the number of videos the students were requested to work with as shown in Table 1, we also used another variable that represented the relative count of annotations a student created per video in a course. This variable is computed by dividing the count of video annotations created by the number of videos students were required to work in a given course.

Several other secondary dependent variables were used to understand the patterns of interaction with videos under different instructional conditions and with differences in experience. These variables represented the occurrences of events recorded by CLAS in each course including the counts of pause, rewind and fast forward events. Similarly, the total time (minutes) when the play button was activated indicating time likely spent watching each video was used. The effects of these variables are reported in Appendix B.

Learning products. To assess the quality of learning products (i.e., time-stamped video annotations and general video annotations), we used the two frameworks and tools for text analysis most commonly used in educational psychology: Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2009) and Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014). From the suite of LIWC variables, we adopted word count (WC). This variable was selected, as it is commonly shown to be a good proxy for higher levels of cognitive processing (Joksimović, Gašević, Kovanović, Adesope, & Hatala, 2014; Tausczik & Pennebaker, 2009). Moreover, in addition to the counts of annotations, this variable was also an important indicator of the sustained use of the tool (especially relevant for research question 2). That is, a measure of the effort put into the creation of the annotations can be derived from the length of text (word count). In addition, we also wanted to study the length of the self-assessments in each course as a ratio of word counts per self-assessment annotation (WC/Ann). As suggested by Gašević, Mirriahi, and Dawson (2014), this ratio can provide an insight into the quality of individual annotations rather than the entire text of all annotations together.

Coh-Metrix is a well-known toolkit built on the computational linguistic techniques (Graesser, McNamara, & Kulikowich, 2011; McNamara et al., 2014). Coh-metrix is used for the analysis of characteristics of language and discourse. Numerous studies have shown that Coh-Metrix measures can be applied to identify qualitative differences in formal textual documents and discourse (McNamara et al., 2014). The Coh-Metrix offers over 100 measures that cover different dimensions of language and discourse such as genre, cohesion, and syntax. Measures of linguistic complexity, characteristics of words, and readability scores are also available in Coh-Metrix. Given the large number of the Coh-Metrix measures, a principal component analysis was applied to the 53 measures of Coh-Metrix in a study of 37,520 texts available in the Touchstone Applied Science Association corpus (Graesser et al., 2011). The principal component analysis revealed that eight principal components explained 67.3% of the variance. The top five principal components explained over 50% of the variability. The z-scores of principal components are commonly used in the literature. The components identified are well associated with multilevel theoretical frameworks of cognition and comprehension (Graesser & McNamara, 2011; Kintsch, 1998; Perfetti, 2000; Snow, 2002). This makes the components suitable for research in learning-related studies. In this study, we used the following five principal components of Coh-Metrix (Dowell, Cade, Tausczik, Pennebaker, & Graesser, 2014, pp. 126–127):

- "*Narrativity.* The extent to which the text is in the narrative genre, which conveys a story, a procedure, or a sequence of episodes of actions and events with animate beings. Informational texts on unfamiliar topics are at the opposite end of the continuum.
- *Deep Cohesion.* The extent to which the ideas in the text are cohesively connected at a deeper conceptual level that signifies causality or intentionality.
- *Referential Cohesion.* The extent to which explicit words and ideas in the text are connected with each other as the text unfolds.
- *Syntactic Simplicity.* Sentences with few words and simple, familiar syntactic structures. Polar opposite are structurally embedded sentences that require the reader to hold many words and ideas in their working memory.
- *Word Concreteness.* The extent to which content words are concrete, meaningful, and evoke mental images as opposed to abstract words."

The values of these five variables were first computed for each individual annotation. Next, we calculated a mean value for each of these five variables for every student enrolled in the course. The calculated mean values for each student in each course were used in the analyses.

Learning strategy. To evaluate students' learning strategies when using a video annotation tool for self-assessment, we created transition graphs based on the trace data of the recorded learning activities within CLAS (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Malmberg et al., 2014; Winne, Gupta, & Nesbit, 1994). Transitions graphs were created based on a contingency matrix where rows and columns accounted for all possible events. The rows represented the start and the columns represented the end points of the transition edges. Originally, all the cells in the matrix had values of zero. If an event, A, was followed by an event B, number 1 was recorded in the cell representing the intersection of row A and column B. For each future appearance of this transition, the number in the cell was incremented by 1. In this way, we created weighted and directed transition graphs.

For the purposes of the study, we distinguished between events based on the temporal parts of the videos they were associated with. This was consistent with the findings of Gašević et al. (2014) and Mu (2010) who reported that students' operations were unevenly distributed across each video. Specifically, we distinguished between time-stamped events (annotations, pause, rewind, and fast-forward) based on the quartiles of the videos they were associated with. For example, if a time-stamped annotation was associated with the 12 second mark of a 100 second long video, the event type was *annotation in quartile 1*. Similarly, if a participant rewound to second 34 of a 100 second long video, the event type was *rewind in quartile 2*. According to these rules, we created a transition graph for each student in our sample. These graphs consistent of 19 possible nodes – four event types for time-stamped annotations, pause, rewind, and forward events, one for general annotation, one for non-stop watching, and one for end of video watching.

The transition graphs, described above, were used to investigate students' self-regulated learning processes when using CLAS. In this instance, we were particularly interested in the level of students' metacognitive monitoring of their own learning. Metacognitive monitoring is the key activity for learning success, since it is commonly used for evaluation of the learning product and learning strategy (Winne, 2001; Winne & Hadwin, 1998). Azevedo et al. (2008) found that the higher levels of metacognitive monitoring were associated with an increase of feeling of knowing, judgment of learning, and monitoring of progress toward goals. Moreover, Greene and Azevedo (2009) found that monitoring activity was a "key SRL process when developing an understanding of a complex science topic using hypermedia" (p. 18). According to Hadwin et al. (2007) and Winne et al. (1994), a graph-theoretic measure of density – a ratio between the actual number of edges between nodes in a graph and all possible edges in the graph – can be used to assess metacognitive monitoring in learning strategies followed by learners. Specifically, Hadwin et al. (2007,) posit that "participants with lower overall densities have formed some distinct and regular studying patterns whereas participants with higher densities are experimenting with tactics and strategies. That is, these latter students are engaged in more metacognitive monitoring and, hence, more active SRL [self-regulated learning]" p. 114.

2.6 Analysis

Given the nested structure of our data (students within a course) and potential problem of correlated data due to grouping (Seltman, 2012), we relied on linear mixed models to address the research questions. Mixed-effects modeling provides a robust and flexible approach that allows for a wide set of correlation patterns to be modeled and is recommended method for studying similar datasets (Pinheiro & Bates, 2009; Seltman, 2012). Mixed-effects models include a combination of fixed and random effects and can be used to assess the influence of the fixed effects on dependent variables after accounting for any extraneous random effects. Fixed effects correspond to the numerical or categorical variables that are of primary interest and represent fixed, repeatable levels among which comparisons are to be made. Random effects are categorical variables that represent variability among subjects, a random selection from a larger population to which the results can be extended.

Mixed-effects modeling was used to examine the association between the two factors (instructional conditions and previous experience) and the dependent variables. In order to assess this association for each of the dependent variables above and beyond the random effects, we built three models for each of the dependent variables – *null model*, *fixed model*, and *final model* (Table 4 in in Appendix A**Error! Reference source not found.**). The *null model* initially included the random effect only (*student within a course*). However, in some cases, we were not able to fit the model with such a structure of random effects. In such cases (i.e. videos annotated, pause, rewind, word count, word count per annotation, and deep cohesion), we specified *student* as a random effect instead and were able to fit the model. Moreover, in some cases (i.e., time watched, narrativity, and referential cohesion), we could not fit the model even with the revised random effect. On the other hand, a *fixed model* included *condition* and *experience* as fixed effects, while the *final model* included *condition*, *experience*, and *interaction between condition and experience*, as fixed effects. Intraclass correlation coefficient (ICC) (Raudenbush & Bryk, 2002), second-order Akaike information criterion (AICc), and likelihood ratio test (Hastie, Tibshirani, & Friedman, 2011) were used to decide on the best fitting model (Table 4 in Appendix A). We also estimated an effect size (R^2) for each model as goodness-of-fit measure, calculating the variance explained using the method suggested by Xu (2003).

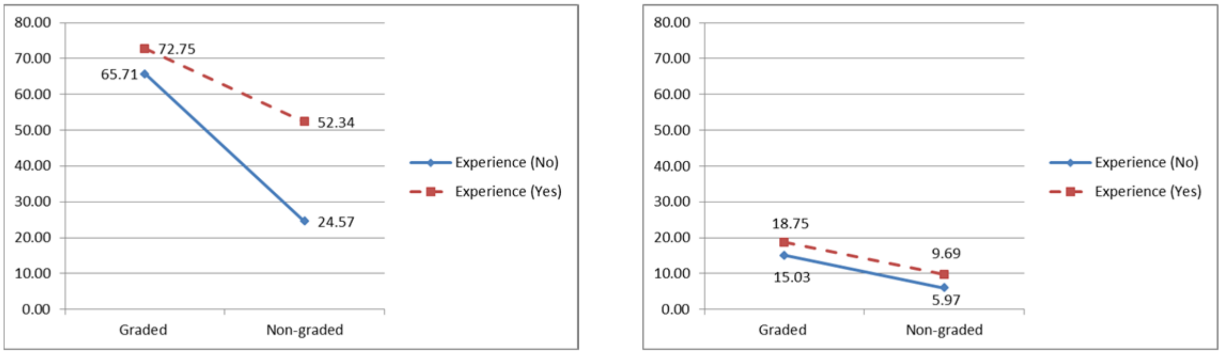
Linear mixed-effects models were conducted using R v.3.0.1 software for statistical analysis with package lme4 (Bates, Maechler, Bolker, & Walker, 2015). The hypotheses specify the direction of the effect, however two-tailed tests were used for significance testing with an alpha level of .05.

3 Results

The results in this section are organized according to the three main dimensions used to operationalize the two research questions that aimed to investigate the effects of instructional conditions (research question RQ1) and experience (research question RQ2) on the adoption of a learning tool.

3.1 Counts of annotations

The likelihood ratio test for ***counts* of *annotations*** models yielded significantly better fit of the *final* model (i.e., the model that included fixed, interaction, and random effects) than the *null* and fixed models. The model showed significant effects of previous experience $F(1, 114.94) = 11.54$, $p < .001$) and the interaction of instructional condition and experience ($F(1, 114.94) = 4.10$, $p = .045$) on the number of the counts of annotations created. The effect of instructional condition ($F(1, 2.96) = 2.52$, $p = .212$) was interestingly not significant though. The estimated mean values, calculated as a result of this model, are shown in Figure 2a. These results indicate that the students with previous experience tended to create significantly more annotations than those students who encountered the tool for the first time. There was no significant difference between the students in the counts of annotations created when they were in the graded conditions compered when they were in the non-graded condition. However, the significant interaction effect showed that the students with and without experience had different trends in annotation counts when they were in the non-graded versus graded conditions. While there was no difference in counts of annotations between the students with and without experience in the graded condition, this difference was significant between students with and without experience in the graded condition.



a) b)

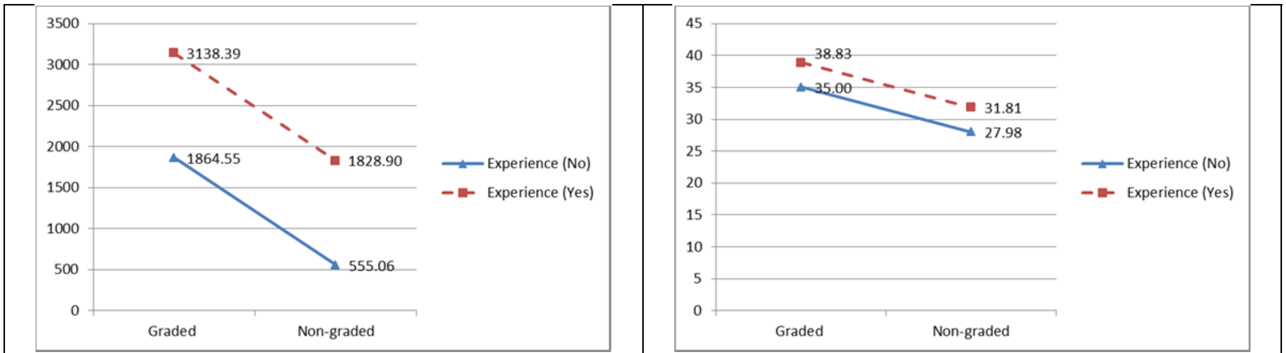
Figure 2. Effects of instructional conditions and experience with the video annotation tool on i) total counts of annotations created and b) counts of annotations per video

The likelihood ratio test for ***counts* of *annotations per video*** models yielded significantly better fit of the *final* model (i.e., the model that included fixed, interaction, and random effects) than the *null* and *fixed* models. The significant fixed effects in this model were consistent with those find in the model for the total count of annotation (discussed in the previous paragraph). That is, the effects of *experience* ($F(1, 26.17) = 25.59$, $p < 0.001$) and *interaction of experience and instructional condition* ($F(1, 26.17) = 4.85$, $p = 0.036$) were significant, while the effect of *instructional conditions* was not significant($F(1, 2.24) = 12.34$, $p = 0.061$). The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 2b. Given that the same significant effects were also found for counts of annotations, the interpretation of the results for counts of annotations per video is the same as stated for counts of annotations in the previous paragraph. Moreover, as shown in Table 4 in Appendix A and consistent with the results of the fixed effects for the count of annotations per video model, the random effect student within a course explained about 69%, while the course itself explained only 9% of the variability in the model.

The results reported in Appendix B showed similar trends with other secondary depended variables about amount of operations performed to interact with video available in the video annotation tool.

3.2 Quality of learning products

The likelihood ratio test for the ***count of words*** models unveiled a significantly better fit of the *fixed* model (i.e., the model that included fixed and random effects) than the *null and final* models. The model showed significant effects of previous experience ($F(1, 88.44) = 53.60$, $p < .001$) and instructional condition ($F(1, 99.82) = 59.90$, $p < .001$) on the number words per annotation. The estimated mean values of count of words written by students in their annotations, calculated as a result of this model, are shown in the diagram in Figure 3a. These findings indicate that the students with previous experience had significantly more words in their annotations than those students who encountered the tool for the first time. The students who were in the graded condition used significantly more words in their annotations compared to the students in the non-graded condition.



a) Total count of words in annotations | b) Word count per annotation

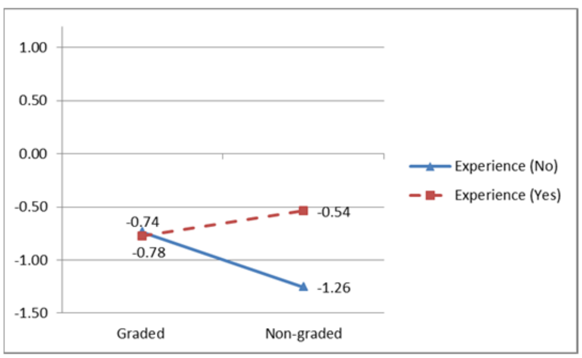
Figure 3. Effects of instructional conditions and experience with the video annotation tool on word counts in annotations and words per annotation

The likelihood ratio test for the ***words per annotation*** models revealed a significantly better fit of the *fixed* model (i.e., the model that included fixed and random effects) than the *null and final* models. The model showed significant effects of previous experience ($F(1, 39.73) = 7.02$, $p = .012$) and instructional condition ($F(1, 43.70) = 19.74$, $p < .001$) on the number of words per annotation. The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 3b. Based on these results, it could be concluded that the students with previous experience tended to use significantly more words per annotation than those students who encountered the tool for the first time. The students who were in the graded condition used significantly more words per annotation compared to the students in the non-graded condition.

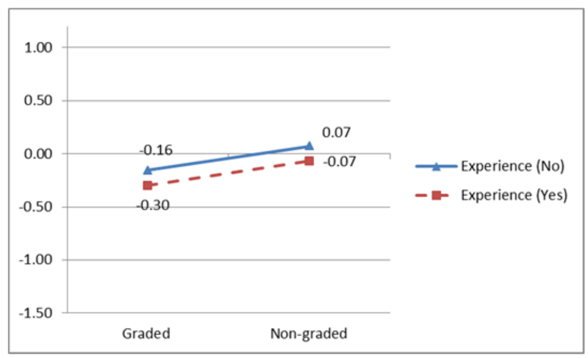
The likelihood ratio test for the ***deep cohesion*** models showed a significantly better fit of the *final* model (i.e., the model that included fixed, interaction, and random effects) than the *null and fixed* models. The model showed non-significant effect of instructional conditions ($F(1, 53.78) = 1.15$, $p = .289$), while experience ($F(1, 54.54) = 6.75$, $p = .012$) and interaction of experience and instructional condition (F(1, 111.13) = 6.11, p = 0.015) had significant effects on the scores of deep cohesion of the text in the annotations. The estimated mean values of deep cohesion in students annotations, calculated as a result of this model, are shown in the diagram in Figure 4a. Based on these results, it could be concluded that the students with previous experience tended to have annotations with the higher scores of deep cohesion than those students who encountered the tool for the first time. The students who were in the graded conditions had no different scores of deep cohesions for annotation compared to the students in the non-graded condition. The significant interaction effect indicates however that there was a different pattern of deep cohesion in annotations between students with and without experience in different (graded vs. non-graded) instructional conditions. While there was no difference in deep cohesion between students with and without experience in the graded condition, this difference between students with and without experience in the non-graded condition was significant.

The likelihood ratio test for the ***syntactic simplicity*** models produced a significantly better fit of the *fixed* model (i.e., the model that included fixed and random effects) than the *null and final* models. The model showed non-significant effects of both instructional conditions ($F(1, 2.19) = 3.51$, $p = .191$) and experience ($F(1, 21.65) = 1.54$, $p = .229$) on the scores of syntactic simplicity of the text in the annotations. The estimated mean values for syntactic simplicity of the text in the students' annotations, calculated as a result of this model, are shown in the diagram in Figure 4b. Based on these results, it could be concluded that the students with previous experience did not have different annotations with respect to syntactic simplicity than those students who encountered the tool for the first time. The students who were in the graded conditions had no different scores of syntactic simplicity for annotation compared to the students in the non-graded condition.

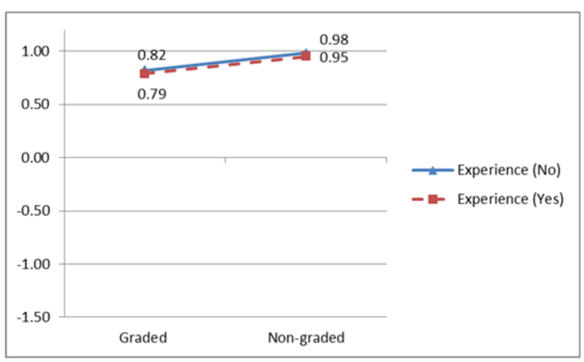
The likelihood ratio test for the ***word concreteness*** models yielded a significantly better fit of the *fixed* model (i.e., the model that included fixed and random effects) than the *null and final* models. The model showed non-significant effects of both instructional conditions ($F(1, 3.04) = 0.46$, $p = .544$) and experience ($F(1, 52.02) = 0.03$, $p = .859$) on the scores of word concreteness of the text in the annotations. The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 4c. Based on these results, it could be concluded that the students with previous experience did not have different annotations with respect to word concreteness than those students who encountered the tool for the first time. The students who were in the graded conditions had no different scores of word concreteness for annotation compared to the students in the non-graded condition.



a) Deep cohesion



b)Syntactic simplicity



c)Word concreteness

Figure 4. Effects of instructional conditions and experience with the video annotation tool on the Coh-Metrix scores of deep cohesion, syntactic simplicity and word concreteness

The likelihood ratio test for the ***referential cohesion*** *and* ***narrativity*** models did not yield significantly better fit of the *fixed* and *final* models than the *null* models.

3.3 Learning strategy

The likelihood ratio test for the ***density*** models showed a significantly better fit of the *fixed* model (i.e., the model that included fixed and random effects) than the *null and final* models. The model showed non-significant effects of both instructional conditions ($F(1, 2.97) = 6.49$, $p = .085$) and experience ($F(1, 115.95) = 0.71$, $p = .400$) on the density of their transition graphs. The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 5. Based on these results, it could be concluded that the students with previous experience did not have different annotations with respect to word concreteness than those students who encountered the tool for the first time. The students who were in the graded conditions had no different scores of word concreteness for annotation compared to the students in the non-graded condition. Interestingly though, Moreover, as shown in Table 4 in Appendix A, the random effect of course explained about 43.1% of the variability in the model. This can probably shed some light why the estimated mean values of density (Figure 5) were different between the courses with graded vs. non-graded instructional conditions.

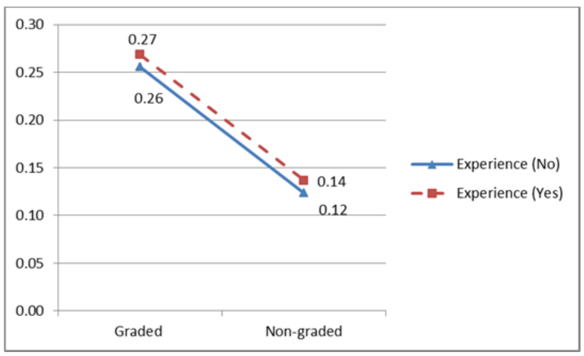


Figure 5. Effects of instructional conditions and experience with the video annotation tool on density of transition graphs

4 Discussion

4.1 Discussion of the Results in Relation to the Research Questions

The results reported in the previous section support the importance of integrating a learning tool with an assessed instructional condition when students encounter the tool for the first time (research question RQ1). This offers some evidence for our proposition that providing a scaffold is necessary (Azevedo & Hadwin, 2005; Beed et al., 1991) for guiding students towards proficient use of a tool in order to meet Winne's (2006) first three conditions for tool adoption – awareness of the tool availability, mapping to a task, and skill to use the tool. Only once these conditions are met, students will likely use the tool extensively. There was an observed increase in the counts of annotations between graded vs. non-graded condition (Figure 2). Although it was anticipated that instructional conditions (graded vs. graded) would have a significant effect on the count of annotations produced by the students (RQ1), this effect was not found to be significant in the results reported in Section 3.1. Rather and somewhat counter-intuitively, the effects of experience with the tool use and interaction of instructional conditions and experience were significant (RQ2).

The results showed that experience consistently had a significant effect on the quality of learning products and indicators of deep learning. The significant effects of both instructional conditions and experience were found on three measures of the quality of learning products (Section 3.2), namely, counts of words, words per annotation, and deep cohesion. The effects of these findings are illustrated in the diagrams shown in Figure 3 and Figure 4. It is interesting to observe that the students in the graded instructional condition with no-prior experience produced almost the same count of words in their annotations as their counterparts with experience in the non-graded condition (see Figure 3a). A similar pattern is observed for words per annotation as shown in Figure 3b. This finding is important, since word count has already been reported as a proxy of deep engagement into knowledge construction. For example, Joksimovic et al. (2014) found that higher levels of cognitive presence were significantly associated with higher counts of words in messages exchanged in asynchronous online discussions. This finding is consistent with essay grading research that demonstrates that the word length of an essay is the strongest predictor of the final essay grade (Page & Petersen, 1995). Somewhat surprisingly, the findings for deep cohesion did not reveal any significant effects of instructional condition, rather the significant effects were found for experience and interaction of experience and instructional conditions. Namely, the scores of deep cohesion for the students with experience in the non-graded conditions were even higher than those of the students without experience in the graded condition. An opposite pattern is found for students without experience in the non-graded condition and who had the lowest scores of deep condition. Scores of deep cohesion for students in the graded condition regardless of their experience remained on the same level. It should be noted that deep cohesion is an indicator of deeper conceptual level that signifies causality or intentionality (Graesser et al., 2011). Moreover, low scores of deep cohesion are also a sign of a low degree of goal-oriented connectives of the text in annotations with students' self-assessments (Graesser et al., 2011). Therefore, the results of this study suggest that prior experience with the tool use is a critical factor for engagement into deep learning levels. This is also consistent with the previously referenced work by Malmberg et al. (2014) on learning strategy adoption which indicates that although the quantity of the strategy use remains consistent with experience, the quality of the learning product increases.

There was an evident, though statistically non-significant, decline (Section 3.3) in the measures of metacognitive monitoring when the assessed instructional condition was removed (see Figure 5). According to our theoretical background, the decrease in metacognitive monitoring in this study was a result of the changed standards driven by the instructional conditions. The lower metacognitive monitoring can be explained by the reduced need to pay attention to the details observed in the videos in order to accurately describe the observations in annotations as shown by the drop of pause, rewind, and fast forward events in Course 4. This may have led to the decrease in the density of the graphs used for modelling the learning strategy. However, in spite of the similar numbers of annotations created and quantity of text, this lessening of metacognitive monitoring could reduce the level of students' understanding of the study topics as shown by Greene and Azevedo (2009). Therefore, additional scaffolding and instructional strategies are required in order to maintain the level of metacognitive monitoring. Research on externally-facilitated regulated learning offers some guidance for how to address this issue. For example, feedback on students' annotations or rubrics for self, peer, or instructor-assessment can help guide the quality of annotations. Moreover, sharing the annotations with peers seems to be another promising instructional strategy (Hulsman & van der Vloodt, 2015).

In contrast to the existing literature, the current study showed that grading of students use of a technology is not only a significant factor influencing uptake but also stimulates longer term adoption and approaches to studying. The prevalent assumption that the technology use is driven by grading is also consistent with the literature that emphasizes assessment as the strongest prompt for learning (Boud, 1995; Eisner, 1993). That is, what will be assessed and how it will be assessed guides students' learning and motivation in formal education. For example, Wormald and Schoeman (2009) showed that increasing the assessment weight of the anatomy course in a medical school had a significant positive impact on students' motivation to learn the subject of anatomy. In the current study, we showed that such assessment prompts (i.e., graded condition in our study) can also be used to scaffold students' approach to studying and that such initial scaffolds have a long term effect. That is, students maintained the use of the tool to aid their learning even though the instructional conditions changed (graded to non-graded).

4.2 Limitations

Future studies should collect further demographic data about students (e.g., disciplinary background, ethnic background, and language proficiency) and study whether and if so, to what extent these variables confound the findings reported in this study. Future studies should also account for the effects of individual differences – e.g., motivation to use technology, self-efficacy about the subject matter and/or technology, achievement goal orientation (Elliot, Murayama, & Pekrun, 2011), approaches to learning(Biggs, Kember, & Leung, 2001), and metacognitive awareness (Duncan & McKeachie, 2005) – that are found to be significant covariate of the adoption of learning tools (Clarebout et al., 2013).

An avenue of investigation for future studies would be to attempt to replicate the extent the findings of this study apply to other tools and technologies. Future studies could examine the extent to which the specific study tactic that was supported by the use of the tool can have an effect on future adoption. In the current study, we looked at how a technology can support student self-assessment. It could be the case that technology-task fit plays a critical role (McGill & Klobas, 2009) and the findings reported in this paper are only generalizable to the extent to which a tool effectively supports a study tactic of high value for the completion of specific learning tasks.

4.3 Implications for Research

This study provides further insight and evidence into existing body of research on learning tool use as a self-regulated learning process. The theoretical model adopted for the study (Lust et al., 2013; Winne, 2006) explains the rationale for the decisions that students made when using the tool under differing instructional conditions. Moreover, the study showed the importance of having more advanced measures of learning processes that can account for important factors affecting self-regulation of students' tool use. In particular, the use of the COPES model was found to be highly beneficial for informing the definition of the types of measures used in the study. The COPE model allowed for the theoretically-grounded interpretation of the results and the relationships observed between individual variables. With these two theoretical groundings, self-regulated learning and COPES, future studies concerning tool use should focus more on the learning aspects of the tools rather than just the various usability factors that are common in research on technology acceptance (Davis, 1989). Of course, these factors are well-established in technology acceptance research (e.g., perceived ease of use, learning, and usefulness) and are potentially important internal conditions – as per the COPES model. However, their role can and should be more closely investigated according to the model theorized in this paper. Likewise, the effects of other individual differences (e.g., self-efficacy to use a tool, achievement goal-orientation or epistemic beliefs) under alternate instructional conditions are another important avenue for future research.

To advance our understanding of learning tool use as a self-regulated learning process, it is important to develop measures that can allow for the study of learning products and learning strategies as well as their association with conditions and standards associated with specific learning situations. Text analysis (e.g., Coh-Metrix) and the analysis of temporal associations between events of different learning operations are highly relevant (e.g., transition graphs used in this paper or process and sequence mining suggested by Reimann, Markauskaite, & Bannert (2014) and Winne (2014). Moreover, the investigation of the more dynamic measures of individual differences such as the use of trace data to track achievement goal orientation (Zhou & Winne, 2012) affords more granular insights into the internal conditions that may explain why, when, and how students use particular learning tools.

The importance of longitudinal studies to address the question whether and when certain scaffolds can gradually be removed is stressed by the differences revealed in the results related to research question RQ2 (Course 2 vs. Course 3 and Course 2 vs. Course 4). As common for research on instructional scaffolding, the problem of finding a point in time when a scaffold can be faded out is essential (Brown, Collins, & Duguid, 1989). If a scaffold remains despite students' improved skills in using a learning tool, an extraneous cognitive load can be created and an expertise reversal effect triggered (Kalyuga, 2007). Therefore, future research needs to address this issue by investigating students' use of a learning tool across several courses where different types of scaffolds are gradually faded out. This is relevant in the context of this study, as students' scores of deep cohesion were the highest in the non-graded instructional condition after gaining experience with the tool use in the previous course. Adaptive scaffolding should also be investigated in the future in order to accommodate the needs of students with different levels of experience with and skills to use the tool, since that can be encountered commonly in course enrollments, as shown in the present study.

4.4 Implications for Practice

The major implication for practice from this study's findings is that tool use and sustained adoption is not only driven by student motivation. The study results indicate that when students are introduced to a new learning tool, the tool use should initially be scaffolded and integrated into the course design with assessment and instructor feedback about the learning products created with the tool. In subsequent courses, the tool use should continue to be scaffolded and be accompanied with either instructor or peer feedback, especially in situations where the use of the tool is not summatively assessed. This practice will help sustain higher levels of metacognitive monitoring, which are critical components for learning success.

5 References

Alexander, P. A., Graham, S., & Harris, K. R. (1998). A Perspective on Strategy Research: Progress and Prospects. *Educational Psychology Review*, *10*(2), 129–154. http://doi.org/10.1023/A:1022185502996

Azevedo, R. (2005). Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning? The Role of Self-Regulated Learning. *Educational Psychologist*, *40*(4), 199–209. http://doi.org/10.1207/s15326985ep4004_2

Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist*, *50*(1), 84–94. http://doi.org/10.1080/00461520.2015.1004069

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding Self-regulated Learning and Metacognition – Implications for the Design of Computer-based Scaffolds. *Instructional Science*, *33*(5-6), 367–379. http://doi.org/10.1007/s11251-005-1272-9

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, *56*(1), 45–72. http://doi.org/10.1007/s11423-007-9067-0

Baker, R., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, *1*(1), 3–17.

Bargeron, D., Gupta, A., Grudin, J., & Sanocki, E. (1999). Annotations for streaming video on the Web: system design and usage studies. *Computer Networks*, *31*(11), 1139–1153.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. Retrieved from http://CRAN.R-project.org/package=lme4

Beed, P. L., Hawkins, E. M., & Roller, C. M. (1991). Moving Learners toward Independence: The Power of Scaffolded Instruction. *The Reading Teacher*, *44*(9), 648–655.

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133–149. http://doi.org/10.1348/000709901158433

Boud, D. (1995). Assessment and learning: contradictory or complementary. In *Assessment for learning in higher education* (pp. 35–48). London, UK: Kogan.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, *18*(1), 32–42. http://doi.org/10.3102/0013189X018001032

Chen, P.-S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. *Computers & Education*, *54*(4), 1222–1232. http://doi.org/10.1016/j.compedu.2009.11.008

Cheung, R., & Vogel, D. (2013). Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning. *Computers & Education*, *63*, 160–175. http://doi.org/10.1016/j.compedu.2012.12.003

Cho, M. H., & Jonassen, D. (2009). Development of the human interaction dimension of the Self-Regulated Learning Questionnaire in asynchronous online learning environments. *Educational Psychology*, *29*(1), 117–138. http://doi.org/10.1080/01443410802516934

Cho, M. H., & Kim, B. J. (2013). Students' self-regulation for interaction with others in online learning environments. *The Internet and Higher Education*, *17*, 69–75. http://doi.org/10.1016/j.iheduc.2012.11.001

Cho, M. H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education*, *34*(3), 290–301. http://doi.org/10.1080/01587919.2013.835770

Clarebout, G., Elen, J., Collazo, N. A. J., Lust, G., & Jiang, L. (2013). Metacognition and the Use of Tools. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 187–195). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5546-3_13

Colasante, M., & Fenn, J. (2009). "mat": A New Media Annotation Tool with an Interactive Learning Cycle for Application in Tertiary Education. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications* (Vol. 2009, pp. 3546–3551). Retrieved from http://www.editlib.org/pv/31992

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.*, *13*(3), 319–340. http://doi.org/10.2307/249008

Dawson, S., Macfadyen, L., Risko, E. F., Foulsham, T., & Kingstone. (2012). Using technology to encourage self-directed learning: The Collaborative Lecture Annotation System (CLAS). In *Proceedings of the Australian Society for Computers in Learning in Tertiary Education Annual Conference (ascilite 2012) Conference* (pp. 246–255). Wellington, New Zealand.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, *28*(6), 557–573. http://doi.org/10.1111/j.1365-2729.2011.00476.x

Dowell, N. M., Cade, W. L., Tausczik, Y., Pennebaker, J., & Graesser, A. C. (2014). What Works: Creating Adaptive and Intelligent Systems for Collaborative Learning Support. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (pp. 124–133). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-07221-0_15

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128. http://doi.org/10.1207/s15326985ep4002_6

Dunning, T. (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press.

Edmunds, R., Thorpe, M., & Conole, G. (2012). Student attitudes towards and use of ICT in course study, work and social activity: A technology acceptance model approach: Exploring student perceptions of ICT in three contexts. *British Journal of Educational Technology*, *43*(1), 71–84. http://doi.org/10.1111/j.1467-8535.2010.01142.x

Eisner, E. W. (1993). Reshaping assessment in education: Some criteria in search of practice. *Journal of Curriculum Studies*, *25*(3), 219–233.

Elias, S. M., & MacDonald, S. (2007). Using Past Performance, Proxy Efficacy, and Academic Self-Efficacy to Predict College Performance. *Journal of Applied Social Psychology*, *37*(11), 2518–2531. http://doi.org/10.1111/j.1559-1816.2007.00268.x

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3×2 achievement goal model. *Journal of Educational Psychology*, *103*(3), 632.

Escobar-Rodriguez, T., & Monge-Lozano, P. (2012). The acceptance of Moodle technology by business administration students. *Computers & Education*, *58*(4), 1085–1093. http://doi.org/10.1016/j.compedu.2011.11.012

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating Cognitive Presence in Online Learning: Interaction Is Not Enough. *American Journal of Distance Education*, *19*(3), 133–148. http://doi.org/10.1207/s15389286ajde1903_2

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *24*, 53–65. http://doi.org/10.1016/j.iheduc.2014.09.006

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71.

Gašević, D., Mirriahi, N., & Dawson, S. (2014). Analytics of the Effects of Video Use and Instruction to Support Reflective Learning. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 123–132). New York, NY, USA: ACM. http://doi.org/10.1145/2567574.2567590

Gasevic, D., Mirriahi, N., Long, P. D., & Dawson, S. (2014). Editorial – Inaugural Issue of the. *Journal of Learning Analytics*, *1*(1), 1–2.

Graesser, A. C., & McNamara, D. S. (2011). Computational Analyses of Multilevel Discourse Comprehension. *Topics in Cognitive Science*, *3*(2), 371–398. http://doi.org/10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, *40*(5), 223–234. http://doi.org/10.3102/0013189X11413260

Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, *34*(1), 18–29. http://doi.org/10.1016/j.cedpsych.2008.05.006

Grove, W. A., Wasserman, T., & Grodner, A. (2006). Choosing a Proxy for Academic Aptitude. *The Journal of Economic Education*, *37*(2), 131–147. http://doi.org/10.3200/JECE.37.2.131-147

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2-3), 107–124. http://doi.org/10.1007/s11409-007-9016-7

Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (2nd ed. 2009. Corr. 7th printing 2013 edition). New York, NY: Springer.

Hulsman, R. L., Harmsen, A. B., & Fabriek, M. (2009). Reflective teaching of medical communication skills with DiViDU: Assessing the level of student reflection on recorded consultations with simulated patients. *Patient Education and Counseling*, *74*(2), 142–149. http://doi.org/10.1016/j.pec.2008.10.009

Hulsman, R. L., & van der Vloodt, J. (2015). Self-evaluation and peer-feedback of medical students' communication skills using a web-based video annotation system. Exploring content and specificity. *Patient Education and Counseling*, *98*(3), 356–363. http://doi.org/10.1016/j.pec.2014.11.007

Jeske, D., Backhaus, J., & Stamov Roßnagel, C. (2014). Self-regulation during e-learning: using behavioural evidence from navigation log files. *Journal of Computer Assisted Learning*, *30*(3), 272–284. http://doi.org/10.1111/jcal.12045

Joksimović, S., Gašević, D., Kovanović, V., Adesope, O., & Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The Internet and Higher Education*, *22*, 1–10. http://doi.org/10.1016/j.iheduc.2014.03.001

Kalyuga, S. (2007). Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educational Psychology Review*, *19*(4), 509–539. http://doi.org/10.1007/s10648-007-9054-3

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge, ; New York, NY, USA: Cambridge University Press.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86. http://doi.org/10.1207/s15326985ep4102_1

Kuhn, D. (1995). Microgenetic Study of Change: What Has It Told Us? *Psychological Science*, *6*(3), 133–139. http://doi.org/10.1111/j.1467-9280.1995.tb00322.x

López-Pérez, M. V., Pérez-López, M. C., & Rodríguez-Ariza, L. (2011). Blended learning in higher education: Students' perceptions and their relation to outcomes. *Computers & Education*, *56*(3), 818–826. http://doi.org/10.1016/j.compedu.2010.10.023

Lust, G., Elen, J., & Clarebout, G. (2013). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, *60*(1), 385–395. http://doi.org/10.1016/j.compedu.2012.09.001

Lust, G., Juarez Collazo, N. A., Elen, J., & Clarebout, G. (2012). Content Management Systems: Enriched learning opportunities for all? *Computers in Human Behavior*, *28*(3), 795–808. http://doi.org/10.1016/j.chb.2011.12.009

Malmberg, J., Järvelä, S., & Kirschner, P. A. (2014). Elementary school students' strategic learning: does task-type matter? *Metacognition and Learning*, *9*(2), 113–136. http://doi.org/10.1007/s11409-013-9108-5

McGill, T. J., & Klobas, J. E. (2009). A task–technology fit view of learning management system impact. *Computers & Education*, *52*(2), 496–508. http://doi.org/10.1016/j.compedu.2008.10.002

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.

Mirriahi, N., & Dawson, S. (2013). The pairing of lecture recording data with assessment scores: a method of discovering pedagogical impact. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 180–184). Retrieved from http://dl.acm.org/citation.cfm?id=2460331

Mu, X. (2010). Towards effective video annotation: An approach to automatically link notes with video content. *Computers & Education*, *55*(4), 1752–1763. http://doi.org/10.1016/j.compedu.2010.07.021

Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. *Phi Delta Kappan*, *76*(7), 561.

Perfetti, C. A. (2000). Comprehending written language: a blueprint of the reader. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 167–208). Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198507932.001.0001/acprof-9780198507932-chapter-6

Perkins, D. N. (1985). The Fingertip Effect: How Information-Processing Technology Shapes Thinking. *Educational Researcher*, *14*(7), 11–17. http://doi.org/10.3102/0013189X014007011

Pinheiro, J., & Bates, D. (2009). *Mixed-Effects Models in S and S-PLUS* (1st ed. 2000. 2nd printing 2009 edition). New York: Springer.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second Edition edition). Thousand Oaks: SAGE Publications, Inc.

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. http://doi.org/10.1111/bjet.12146

Risko, E. F., Foulsham, T., Dawson, S., & Kingstone, A. (2013). The Collaborative Lecture Annotation System (CLAS): A New TOOL for Distributed Learning. *IEEE Transactions on Learning Technologies*, *6*(1), 4–13. http://doi.org/10.1109/TLT.2012.15

Salomon, G., & Perkins, D. N. (1989). Rocky Roads to Transfer: Rethinking Mechanism of a Neglected Phenomenon. *Educational Psychologist*, *24*(2), 113–142. http://doi.org/10.1207/s15326985ep2402_1

Sánchez, R. A., & Hueros, A. D. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, *26*(6), 1632–1640. http://doi.org/10.1016/j.chb.2010.06.011

Sarfo, F. K., Elen, J., Clarebout, G., & Louw, F. (2010). Innovative instructional intervention and the need for a better insight into instructional conceptions. In M. Zuljan & J. Vogrinc (Eds.), *Facilitating effective student learning through teacher research and innovation* (pp. 151–174).

Seltman, H. J. (2012). *Experimental design and analysis*. Online. Retrieved from http://www.stat.cmu.edu/hseltman/309/Book/Book.pdf

Siemens, G., & Gašević, D. (2012). Special Issue on Learning and Knowledge Analytics. *Educational Technology & Society*, *15*(3), 1–163.

Snow, C. (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. Rand Corporation.

Tausczik, Y. R., & Pennebaker, J. W. (2009). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. http://doi.org/10.1177/0261927X09351676

Trevors, G., Duffy, M., & Azevedo, R. (2014). Note-taking within MetaTutor: interactions between an intelligent tutoring system and prior knowledge on note-taking and learning. *Educational Technology Research and Development*. http://doi.org/10.1007/s11423-014-9343-8

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, *37*(1), 57–70. http://doi.org/10.1023/A:1003548313194

Weinstein, C. E. (1988). Assessment and training of student learning strategies. In R. R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 229–271). New York: Plenum Press.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, *8*(4), 327–353. http://doi.org/10.1016/S1041-6080(96)90022-9

Winne, P. H. (2001). Self-Regulated Learning Viewed from Models of Information Processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah: Lawrence Erlbaum Associates.

Winne, P. H. (2006). How Software Technologies Can Improve Research on Learning and Bolster School Reform. *Educational Psychologist*, *41*(1), 5–17. http://doi.org/10.1207/s15326985ep4101_3

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15–32). New Work and Oxon: Routledge.

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, *9*(2), 229–237. http://doi.org/10.1007/s11409-014-9113-3

Winne, P. H., Gupta, L., & Nesbit, J. C. (1994). Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics. *Alberta Journal of Educational Research*, *40*(2), 177–93.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Wormald, B. W., Schoeman, S., Somasunderam, A., & Penn, M. (2009). Assessment drives learning: an unavoidable truth? *Anatomical Sciences Education*, *2*(5), 199–204.

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, *22*(22), 3527–3541. http://doi.org/10.1002/sim.1572

Yi, M. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies*, *59*(4), 431–449. http://doi.org/10.1016/S1071-5819(03)00114-9

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419. http://doi.org/10.1016/j.learninstruc.2012.03.004

Zimmerman, B. J. (1998). Academic studing and the development of personal skill: A self-regulatory perspective. *Educational Psychologist*, *33*(2-3), 73–86. http://doi.org/10.1080/00461520.1998.9653292

Appendix A

Table 4. Inferential statistics for the model fit assessment

Model	χ^2	df	R^2 marg.	R^2 cond.	AICc	ICC student	ICC course
			Videos annotated				
null			0	.165	365.508		
fixed	41.940***	2	.282	.402	331.691		
full	**4.345***	**1**	**.307**	**.439**	**329.801**	**.19**	
			Annotations				
null			0	.712	1065.563		
fixed	10.603**	2	.302	.870	1047.699		
full	**4.254***	**1**	**.341**	**.854**	**1039.414**	**.231**	**.547**
			Annotations per video				
null			0	.724	611.464		
fixed	20.286***	2	.438	.892	592.977	.684	.123
full	**6.083***	**1**	**.430**	**.884**	**586.832**	**.687**	**.090**
			Time watched*				
null			**0**	**.378**	**2320.858**		
fixed	4.227	2	.025	.435	2290.939		
full	0.373	1	.026	.450	2276.006		
			Pause				
null			0	0	1378.878		
fixed	1.313	2	.011	.039	1367.352		
full	**6.541***	**1**	**.066**	**.165**	**1354.425**	**.105**	
			Rewind				
null			0	.080	1295.973		
fixed	1.276	2	.011	.140	1285.954		
full	**0.034***	**1**	**.050**	**.258**	**1275.715**	**.219**	
			Fast$^{¥}$				
null		-	-	-	-	-	-
fixed		-	-	-	-	-	-
full		-	-	-	-	-	-
			Word count				
null			0	.065	2049.023		
fixed	**87.105*****	**2**	**.497**	**.627**	**1942.597**	**.259**	
full	0.475	1	.499	.616	1930.541		
			Word count per annotation				
null			0	.856	1006.966		
fixed	**20.633*****	**2**	**.004**	**.923**	**985.320**	**.92**	
full	0.850	1	.003	.922	981.745		
			Narrativity‡				
null			**0**	**.690**	**256.384**		
fixed	4.963	2	.042	.831	261.369		
full	2.751	1	.061	.657	261.831		
			Deep cohesion				
null			0	.369	284.802		
fixed	7.821*	2	.051	.463	285.893		
full	**6.168***	**1**	**.102**	**.502**	**282.644**	**.445**	
			Referential Cohesion‡				
null			**0**	**.0002**	**149.963**		
fixed	3.945	2	.033	.034	156.703		

full	1.344	1	.042	.042	159.267		
				Syntax simplicity			
null			0	.706	214.936		
fixed	**6.112***	**2**	**.060**	**.749**	**219.151**	**.718**	**.015**
full	0.329	1	.061	.732	222.163		
				Word concreteness†			
null			0	.716	308.818		
fixed	**0.860**	**2**	**.010**	**.691**	**315.321**	**.654**	**.052**
full	**5.619***	**1**	**.057**	**.657**	**313.639**		
				Density			
null			0	.645	-290.276		
fixed	**6.676***	**2**	**.388**	**.652**	**-280.717**		**.431**
full	0.013	1	.438	.604	-295.207		

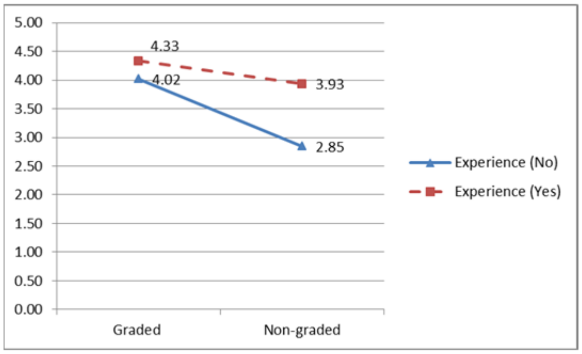
Legend: Models written in the boldfaced font found to be best fit. *The null model had the best fit of the three models. That is, the dependent variable is explained by individual characteristics rather than fixed and interaction effects. †Although the full model was better than the fixed one according to the χ^2 test, the fixed model is kept as the ful model had ICC score of 0 for the level of course. $^{¥}$ Model could not be fitted and the variance for the random effects is 0.

Note: In cases of videos annotated, pause, rewind, word count, word count per annotation, and deep cohesion, models were built with student as a random effect. Models for time watched, narrativity, and referential cohesion could not be fitted with either student or student within a course random effects. All other models included student nested within a course as a random effect specification.

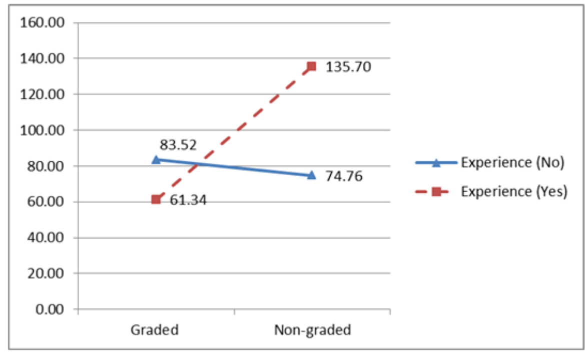
Appendix B

This appendix reports on the results of the hierarchical mixed model analysis on the secondary dependent variables used as proxies of the amount of the learning tool use; the primary dependent variable was count of annotations created by learners and was reported in the main body of paper text.

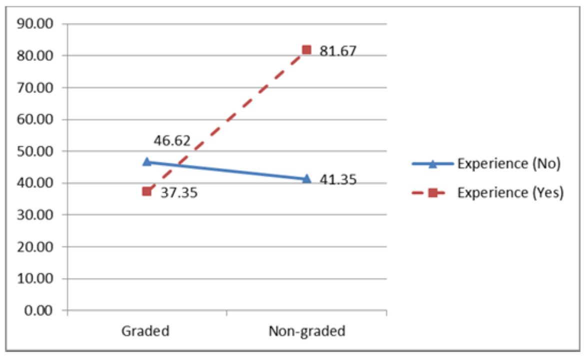
The likelihood ratio test for ***videos annotated*** models yielded significantly better fit of the *final* model (i.e., the model that included fixed, interaction, and random effects) than the *null* model. The model showed significant effects of previous experience (F(1, 72.68) = 15.80, p < .001), instructional condition (F(1, 73.86) = 19.94, p < .001), and the interaction of instructional condition and experience (F(1, 110.89) = 4.29, p < .041) on the number of videos watched. The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 6a. Observing these results, it could be concluded that the students with previous experience with the tool use tend to annotate more videos, than those students who encounter the tool for the first time. Likewise, the students who were in the graded instructional condition annotate more videos than those who were in the non-graded condition.



a) Videos annotated



b) Pause events



c) Rewind events

Figure 6. Effects of instructional conditions and experience with the tool on the total number videos annotated, time spent watching videos, and counts of pause and rewind events

The likelihood ratio test for the ***videos annotated*** models yielded significantly better fit of the *final* model (i.e., the model that included fixed, interaction, and random effects) than the *null* and fixed models. The model showed significant effects of previous experience ($F(1, 72.68) = 15.80$, $p < .001$), instructional condition ($F(1, 73.86) = 19.94$, $p < .001$), and the interaction of instructional condition and experience ($F(1, 110.89) = 4.29$, $p = .041$) on the number of videos watched. The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 6b. Observing these results, it could be concluded that the students with previous experience with the tool use tended to annotate more videos, than those students who encounter the tool for the first time. Likewise, the students who were in the graded instructional condition annotated more videos than those who were in the non-graded condition.

The likelihood ratio test for the ***rewind*** models yielded significantly better fit of the *final* model (i.e., the model that included fixed, interaction, and random effects) than the *null* and fixed models. The model showed a non-significant effect of previous experience ($F(1, 79.60) = 15.80$, $p = .213$) and significant effects of instructional condition ($F(1, 81.51) = 4.50$, $p = .036$), and the interaction of instructional condition and experience ($F(1, 103.36) = 6.70$, $p = .011$) on the count of pause events. The estimated mean values, calculated as a result of this model, are shown in the diagram in Figure 6c. Observing these results, it could be concluded that the students with previous experience with the tool use did not have more pause events than those students who encounter the tool for the first time. Likewise, the students who were in the graded instructional condition had more pause events than those who were in the non-graded condition.