This is the author's postprint version of:

Verpoorten, D., Westera, W., & Specht, M. (2011). Prompting reflection through annotations in formal online learning. In W. Reinhardt, T. D. Ullmann, P. Scott, V. Pammer, O. Conlan, & A. J. Berlanga (Eds.), *Proceedings of the 1st European Workshop on Awareness and Reflection in Learning Networks (ARNets11). In conjunction with the 6th European Conference on Technology Enhanced Learning (EC-TEL 2011): Towards Ubiquitous Learning 2011* (Vol. 790, pp. 90-96). September, 21, 2011, Palermo, Italy: CEUR Workshop Proceedings. Available at http://ceur-ws.org/Vol-790

Promoting reflection through annotations in formal online learning

Dominique Verpoorten, Wim Westera, and Marcus Specht

Learning Media Laboratory, Centre for Learning Sciences and Technologies, Open University in the Netherlands
{dve, wim, spe}@ou.nl

Abstract. This article explores the role of annotations as reflection amplifiers while studying in an Open Educational Resources distance course. A controlled experiment reveals that the treatment groups using frequent and local annotations did not perform better at the test. However, measures within the treatments exhibit a moderate but significant improvement of the mark in the group composed of high annotators.

Keywords: reflection, annotation, online learning, reflection amplifier, metacognition, meta-learning, OER

1 Introduction

Note-taking, either when listening to lectures or reading texts, is a "totem" of teaching and learning. It seems that for centuries tutors have been expecting that students do take notes and that tutees consider note-taking as a natural activity in a scholarly life. But what functions does it exactly fulfill?

According to Hartley & Davies [1], annotations (sometimes called “marginalia”) have 2 faces. As a process, they help to maintain attention and apprehend the material in a cognitively engaged way. They assist in keeping learning going on and, as such, they can be signs of reflection addressed by the self to the self in the present of the interaction. They somehow make learning visible [2]. Annotations are also products. They are stored for the future, with possibilities to be reviewed, re-structured, enriched. Boch & Piolat [3] use a similar distinction but labeled differently: notes to record information (products) versus notes to aid reflection (process).

Despite a renewed interest for digital annotations in the context of Web 2.0. growth and the development of innovative tools likely to take on new annotation functions (tagging, sharing) in the digital world, research concerned with learning aspects of private electronic annotation do not abound.

2 Annotations as reflective micro breaks

An annotation is a personal trace left by a student on a read document. An annotation records readers’ efforts to shape their interaction with the content. This article conceptualizes the making of a digital annotation as a process of personal *reflection*. Annotating is therefore conceived as a “reflection amplifier”. According to the term used by Verpoorten, Westera, and Specht [4] in their structured inventory of reflective techniques, a reflection amplifier is a compact, frequent and focused tingling of reflection about the content and/or about the self-as-a-learner within a particular learning task. Reflection amplifiers contrast with time-consuming and post-practice opportunities for reflection like learning diaries or portfolios. Even though they take only a handful seconds, annotations are conceived as brief episodes of thinking while learning. This action of “writing on the reading” is deemed to enhance the quality of learning.

3 Research questions

The study investigated the outcomes of using digital annotation software in online courses.

First, it was hypothesized that frequent use of the annotation tool and of a dashboard of annotations would be positively reflected in achievement scores because it represented a beneficial active process of content internalization maintained by short but repeated efforts of reflection.

Secondly, it was predicted that some annotation strategies would contribute more significantly to learners‘ performance and overall engagement state.

4 Methodology

4.1 Learning Context

The online course. The learning material of the experiment was the 4-hour online course "Seks en de evolutie" (Sex and the theory of evolution), an OpenER course [5] designed and offered in Dutch by the Open University in the Netherlands. It offered 30 well illustrated pages of 800 words in average (Fig. 1) and 4 interactive animations. It covered quite complex and interrelated notions and mechanisms as defined by Darwin and his followers: mutation, natural selection, genetic drift, gene flow, survival of the fittest, etc. On the whole, the course gave an in-depth account about the evolutionary theory and invited the learner to use it as an interpretation grid of behaviors observable in everyday life. In all treatments, the course was introduced by a welcome video and closed with a test.

The tools. The digital annotation tool presented as a comment box displayed on each page (Fig.1). It kept record of all annotations produced by the learner on this very page, arranged by date. A static reminder was visible on all pages, saying: "Do not forget your annotation". The annotation tool unfolded through a click by the learner. Consistently with the length of the reading material and with the actions requested from the learner (frequent but short notes), the surface of the tool was intentionally not extremely large and its function was deliberately restrained to the basic typing of very localized comments on the pages.

Fig. 1. – The annotation in its local context of a standard web page of the course

However, in order to prevent effects of fragmentation and to support the function of annotations as products, all marginalia were also recorded on a single page called "dashboard", available at any time by the student. On this page, the annotations were organized by section of the course content. By combining an annotation tool to a dashboard, this research attempted to differently treat the effects of annotations from the effects of reviewing them.

The annotation strategies. Treatment 1 and 2 used the same annotation tool located on each page but in a different way. In both conditions, subjects were asked to take an annotation each time they (re-) visited a page. However, participants in condition 1 could encode their marginalia in the way they preferred while those of condition 2 were requested to produce annotations as questions. Precisely, participants were requested to put themselves in the shoes of the teacher and to craft questions likely to be

used in a final test about the content of the page. Verpoorten & al. label this reflective strategy: “Students set the test” and describe it as “Learners are asked to make up the questions they could get for their exam” [4].

4.2 Sample and Schedule

Invitations to participate were displayed on electronic and paper communication channels of the Open University. Announcements of the study were also sent to Dutch dailies and women magazines, as well as to a psychology popular publication. 247 subjects, randomly distributed into the 3 conditions, entered the course at least once but only 137 completed the final test and answered the evaluation questionnaire. They compose the final sample: 34 persons for condition 1 (control), 54 for condition 2 (free annotations) and 49 for condition 3 (annotations as questions).

4.3 Measure instruments

In this comparative study, the online course was delivered at 3 conditions:

- no annotation tool (control group);
- frequent free annotations;
- frequent structured annotations (“students set the test”).

The intervention variables were the provision of an embedded annotation tool and the exposure to a strategy for frequent and local annotations. The dependent variable was the subjects’ cognitive engagement with the content, broken down in 5 quantitative indices of performance:

- index 1: score at the final test. This index designated the score obtained at the final test taken straight after the study session. It measured learners’ achievement through 16 multiple-choice questions controlling knowledge and comprehension;
- index 2: number of pages (re)visited;
- index 3: time spent in the course. This index was measured as the number of “active ten-minute periods” in the course. A period is considered as “active” when it records one click in the 10 minutes time span, between the arrival on the page and the departure;
- index 4: number of annotations;
- index 5: total number of characters for the annotations;

(Indices 1, 2, 3 were common to all conditions. The others were logging information available only for conditions 2 and 3).

Prior to the access to the course, participants filled in a pre-questionnaire comprising questions about note-taking habits, a shortened version of the MAI (Metacognitive Awareness Inventory) and self-reported evaluations of familiarity with the topic and with ICT.

5 Results

An alpha level of .05 was used for all statistical tests. Levene's test of homogeneity of variance preceded parametric tests and degrees of freedom were adapted if necessary.

5.1 Inter-group comparisons

Background questionnaire. To ensure equivalence between conditions at baseline, one-way ANOVAs were performed on the elements of the background questionnaire. The procedure indicated an even distribution regarding meta-cognitive capacities, $F(2, 134) = .27$, $p = .76$, familiarity with the topic, $F(2, 134) = .18$, $p = .83$, and familiarity with eLearning, $F(2, 134) = 1$, $p = .119$. Descriptive statistics also showed an equal distribution for age, sex and education level.

Index 1: score at the final test (3 groups). An ANOVA procedure exhibited no significant differences between groups regarding mean results at the final test, $F(2, 134) = .44$, $p = .64$.

Indices 2 and 3: logging information (3 groups). Significant differences (Table 1) emerged between conditions with regard to the total time spent on the course, $F(2, 134) = 3{,}494$, $p = .033$, and the number of page views, $F(2, 134) = 5{,}291$, $p = .006$.

Table 1. Means for the 3 performance variables common to the 3 conditions

	Final score at the test			Total time spent on course			Number of page views			
	1	2	3							
	(N=34)	(N=54)	(N=49)	1	2	3	1	2	3	Total
Mean	**6,462**	**6,059**	**6,464**	**245,00**	**322,41**	**333,67**	**57,09**	**73,19**	**84,18**	**73,12**
Std. Deviation	**2,3195**	**1,7320**	**1,8850**	**115,240**	**171,188**	**172,539**	**23,020**	**36,881**	**44,961**	**38,477**
Minimum	**,0**	**1,9**	**2,5**	**100**	**50**	**100**	**29**	**31**	**29**	**29**
Maximum	**10,0**	**9,4**	**10,0**	**510**	**810**	**970**	**110**	**222**	**252**	**252**

Post-hoc tests revealed that the amount of time and page views was higher for treatment groups compared to the control group but equivalent between treatment groups.

Indices 4 and 5: logging information (2 groups). Table 3 provides information about the use of the annotation tool in conditions 2 and 3. From the observation of the logs, it turned out that the participants in the treatments displayed quite different annotation behaviours, some learners made a large number of annotations (more than 20.000 characters in condition 2 and more than 10,000 characters in condition 3),

while others did with a few hundreds. These differences in approach may not become visible in the total time spend (no significant difference between condition 2 and 3), but clearly they cannot be ignored.

5.2 Intra-group comparisons (profiles)

At this stage, the analysis moved its focus from inter-group comparisons to measures based on intra-groups profiles. In this context, each participant to the 2 treatment groups was labeled “high” or “low” for each index or “protein” of performance: low/high annotator, low/high total number of characters, low/high browser. Profiles were built on the ratio between the absolute number of “annotations”, “characters”, “page views” obtained by a learner and the total time spent by this learner in the course. The frequencies of these different kinds of enactments on the learning material quantified the reflective engagement with this material. Relating these high and low behaviors to the performance at the test exhibits significant differences only for index 4 (number of annotations), $t(101) = 2.146$, $p = 0.034$, $d = 0.37$ and for index 5 (total number of characters in annotations), $t(101) = 2.76$, $p = 0.007$, $d = 0.35$. High timers (time spent in the course) and high browsers (page views) did not make better than their low peers regarding test performance.

6 Discussion

Going back to the underpinning hypotheses of this study, it must be concluded that:

- average score at the final test does not differ between control and treatments groups. Offering an embedded annotation tool for frequent and local annotations and a synoptic dashboard for these annotations does not create any observable leap in learners' results compared to a plain distance course;
- the structured annotation strategy did not produce any significant enhancement of learners‘ performance compared to its free counterpart.

Regardless of the denial of its two main hypotheses, this study nevertheless delivered some results when the focus was put on high annotators versus low annotators. In this case, it appeared that annotations can be a vehicle for a reflection, traceable in the learners' achievement at the final test. Unsurprisingly, students who took advantage of the annotation tool, in number and length of annotations, learnt more from the texts that those who did not (see similar results regarding the number of handwritten annotations in [6]).

The study also invites to refine the notion of “reflection”. The word “reflection amplifier” was here used to point at the intended effect of the annotations. But a more neutral label like “thinking amplifier” might be better. Yet, the only secured observation is that high achievers in the treatment group show a higher level of “physical” activity (annotating) while learning. That this “active reading” can be equated to “active reflection” remains an open question. The performance enhancement might also be credited to indirect effects of the annotations on ownership, commitment or atten-

tion. Further research is needed to disentangle these notions and their connection to reflection.

Lastly, the study indicate that minimal tools and interventions can already help for learning online. Without spending huge amounts of time and resources (the technical development of the note-taking tool and the associated dashboard took one week), it is possible to equip Open Educational Resources courses with a basic support to reflection that can make a difference when adequately used.

7 Conclusion

This article investigated the possible links between an optimal standard of learning (reflective, autonomous) and the annotation process. Are students right to make frequent and local annotations? Should that practice of "writing on the reading" be recommended to everyone ? With what intensity? Should a teacher worry in case of no-annotation? Do these reflective breaks alter learning? This research provides some indications that frequent and local electronic annotations, conceived as short and repeated episodes of reflection on the content, can be positively related to learners' performance at the final test.

8 References

1. Hartley, J., Davies, I. K.: Note-taking: A critical review. Innovations in Education & Training International, 15(3), 207-224 (1978)
2. Hattie, J.: Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement. London: Routledge (2008)
3. Boch, F., Piolat, A.: Note taking and learning: A summary of research. The WAC Journal, 16 (September), 101-113 (2005)
4. Verpoorten, D., Westera, W., Specht, M.: Reflection amplifiers in online courses: a classification framework. Journal of Interactive Learning Research, 21(4), 654-666 (2010)
5. Eshuis, J. H., Goltstein, G. P. H.: Seks en de evolutie (2007), http://www.ou.nl/eCache/DEF/2/22/850.html
6. Slotte, V., Lonka, K.: Review and process effects of spontaneous note-taking on text comprehension. Contemporary Educational Psychology, 24, 1-20 (1999)