Annotations as reflection amplifiers in online learning – An exploratory study

Dominique Verpoorten[1], Wim Westera[2], Marcus Specht[2]

[1]University of Liège, Belgium
{dvepoorten}@ulg.ac.be
[1]Welten Institute, The Netherlands
{wim, spe}@ou.nl

Abstract. In a controlled experiment on the effects of frequent and local digital annotations, 137 volunteers covered an online course at 3 conditions: no/free/question-based electronic annotations. Results show no difference in performance between groups. However, analyses conducted within treatments suggest positive impacts on performance when annotation rates are taken into consideration, and coupled with other reflective enactments.

Keywords: annotations, reflection amplifiers, students set the test, widgets, split screen learning

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 [1]. An annotation is conceived as a personal trace left by students on a pre-existing text or speech. Annotations record readers' efforts to shape their interaction with this content. Research on note-taking has generated debates since Crawford's early studies in this topic [2]. Promoting annotation behaviours has been a long-lasting concern in distance education. From its beginning, and long before the possibility to think about students in terms of "reflective practitioners" [3], it has been constantly recommended to design paper-based course material with large margins. This liberal use of white space [4, 5] is meant to encourage students to make analytical summary notes of what they would identify as worthy of their attention when they revise. In the 90's, a vast body of research [6] discussed the many issues raised when moving annotation from paper-based to screen display reading. In the past few years, a renewed interest emerged for the processes of "writing on the reading" in digital activity systems, due to the novel burgeoning opportunities for searching, sharing, indexing, ordering, tagging, rating annotations in an "information enrichment" perspective [7].

While the effects of note-taking are well documented for paper-based practice [8, 9, 10], the new wave of research on digital annotations develops concerns in several directions: non-linear or linear annotation techniques [11], annotations as checklists [12], annotation sharing mechanisms [13], collaborative annotation [14, 15], tagging

as annotations [16, 17], multiple displays for annotations [18]. Results reveal various conditions under which Web-based annotation mechanisms are beneficial [19].

Beyond their variety, the new alleys of research (for an extended view on recent work, see [20]) endorse to a large extent [21] the two faces of note-taking already identified by Hartley and Davies [9]:

- as a process, annotations help to maintain attention, apprehend the material in a mentally active way and intensify the attendance to the task. By assisting in keeping learning going, they can be tokens of reflective engagement during the study task;
- as a product, annotations are stored for the future, with possibilities to be reviewed, re-structured, and enriched.

Boch and Piolat [8] use a similar distinction but labelled differently: "notes to aid reflection" (process) versus "notes to record information" (product).

1.1 Reflection Amplifiers

In this study, the annotations are conceived as "reflection amplifiers". Following the definition by Verpoorten, Westera, and Specht [22], reflection amplifiers (*RAs*) refer to deliberate prompting approaches that offer learners structured opportunities to examine and evaluate their learning as it develops. Whereas the promotion of reflection is often associated with post-practice methods of experience recapture [23] through portfolios or learning diaries, RAs are nested in the study material and offered to individuals during learning activities. They induce regular mental tingling for evaluating own learning and nurturing internal feedback [24].

The concise reflection they call for further characterises RAs. As support to condensed reflective processes, RAs operate though miniature Web applications (sometimes called "widgets") performing a single task, displaying a very clear and appropriate graphical style, and providing a single interaction point for direct provision of a given kind of data [25], here the personal annotations. In the way they are practised by learners in this study, the annotations meet the common internal characteristics of RAs: brevity, frequency and crisscrossing with the first-order learning activity, i.e. studying the course content.

1.2 Hypotheses

In a comparative study an online course was delivered at 3 conditions: without annotation tool, with annotation tool and free-style notes, with annotation tool and question-based notes (i.e. expressed in the format of a question on the read material). The study investigated the effects of the digital annotations – conceived as multiple short episodes of analytical reflection – upon the enhancement of the quality of learning. Two hypotheses guided the experiment.

Hypothesis 1. "The availability of an annotation tool and the assignment to use it for frequent and local notes induce higher marks at the test and an increased study time". Short but repeated efforts of reflection are predicted beneficial to the content internal-

ization because they are seen as a way to stay analytically engaged with the supplied learning material. It is also speculated that such a reflective approach to learning has a price with regard to time spent on the material.

Hypothesis 2. "The question-based annotation strategy induces higher marks at the test than the spontaneous way of annotating". The study includes a concern for annotation methods by challenging conventional practice of note-taking "as a student" with a different mode wherein the learner is invited to reflect "as an instructor" (details in section "The annotation methods").

2 Method

2.1 Independent Variables

The intervention variables were the provision of an embedded annotation tool and the exposure to a strategy for frequent and local annotations.

2.2 Dependant Variables

The dependent variable was the subjects' reflective engagement with the content, broken down into seven tangible indices:

- **Index 1**: mark at the final test (FinalTest). This index designated the score obtained at the final test taken after the study session. It measured learners' achievement through 16 multiple-choice questions assessing knowledge and comprehension;
- **Index 2**: time spent in the course (TimeSpent). This index, measured as the number of "active ten-minute periods" in the course, was an estimate. One active period was counted each time that at least one click occurred in a time span of ten min. Longer periods were left out in an attempt to correct for the time students would spend in activities foreign to the study while still being logged into the course;
- **Index 3**: learning efficiency (LearnEff). It is fair to say that the speed of learning is an important achievement (many performance tests, e.g. IQ tests, use time as one of the main indicators). In order to incorporate this temporal dimension in the measures, the marks at the final test were related to the time spent in the course: slow learners got a lower score per unit of time than fast learners. Low-efficiency students did not necessarily receive lower marks, but they needed more time to reach their mark;
- **Index 4**: number of page views (NumberPages). The browsing behaviour, and in this case the action of re-visiting pages, was considered as an index of reflective engagement because it assumed a meta-learning decision about the need of re-reading the material;
- **Index 5**: quantity of annotations (NumberAnnot).
- **Index 6**: total number of characters for the annotations (CharactInAnnot);
- **Index 7**: number of visits (VisitDash) to the Learning Dashboard (see section "Apparatus").

The indices FinalTest, TimeSpent, LearnEff, and NumberPages were common to the 3 conditions. NumberAnnot, CharactInAnnot, and VisitDash were premised upon the annotation tool and therefore only offered in Treatments 2 and 3.

A post-test questionnaire allowed measuring the effects of the intervention on the following additional variables: sense of control and opinion over the annotation experience.

3 Apparatus

3.1 The Online Course

The learning material of the experiment was the four-hour online course "Seks en de evolutie" (Sex and the evolution), a course signed and offered in Dutch by the Open Universiteit on the eLearning platform Moodle [26]. It was made of 30 finely illustrated pages (Fig. 1) of about 800 words each, and 4 interactive animations. It covered quite complex and interrelated notions 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 learners to use it as an interpretation grid of gender-related behaviours observable in everyday life. In the 3 conditions, the course was introduced by a welcome video and closed with the same multiple-question test.

3.2 The tool

The digital annotation tool was a comment box displayed on each page (Fig.1). It kept record of all annotations produced by the learner on this very page. The annotation tool unfolded through a click by the learner. Consistently with the length of the reading material and the action requested from learners (frequent but short notes), the surface of the tool was intentionally not large and its function deliberately restrained to the basic typing. As for pedagogy, the annotation tool was offered to promote analytical scrutiny and internalization of the learning material's meaning by making it possible for learners to capture, within the study task, the gist of what has been read.

Figure 1. An annotation in its local context of a standard Web page of the course

In order to prevent effects of fragmentation and to also support the function of "annotations as products" [9], all local annotations were automatically recorded on a single page called "Learning Dashboard" [27], accessible at any time by the student (Fig. 2). On this dashboard, the annotations were organised by section of the course content.

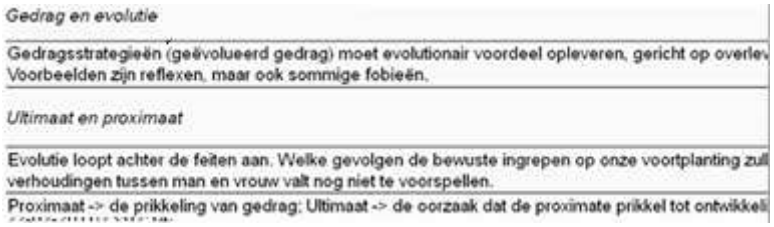

Figure 2. All annotations were displayed within a learning dashboard.

3.3 The Annotation Methods

Subjects in the treatment groups were asked to make an annotation each time they (re-)visited a page. However, participants in one treatment could encode their annotations in the way they preferred ("free annotations") while those in the other treatment were requested to produce annotations as questions ("question-based annotations"). Precisely, these participants were asked 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 at hand. In their inventory of reflective techniques, Verpoorten et al. [22] labelled this reflective strategy: "Students set the test", and described it as: "Learners are asked to make up the questions they might get for their exam".

3.4 Sample and Schedule

Invitations to participate to the experiment were displayed on electronic and paper communication channels of the Open University in the Netherlands, including the homepage of the used course. Dutch dailies and magazines, as well as a psychology popular publication, also received announcements of the study. The registered persons were randomly distributed over the 3 conditions and received credentials for one version of the online course. They had one month to fill in a background questionnaire (15 min), cover the course (4 hr), take the final test (15 min) and answer the evaluation questionnaire (20 min). Out of the 361 initial respondents, 282 entered the course once at the very least but only 137 completed all steps of the study. They composed the final sample: 34 participants in Condition 1 (control group), 54 in Condition 2 (free annotations) and 49 in Condition 3 (annotations as questions). As a reward for their cooperation, they received either an iTunes voucher of 10 euros, or a three-month premium access to a mind-mapping tool (http://www.mindmeister.com), or a USB stick containing applications dedicated to eLearning (http://eduapps.org), or a free entrance to a workshop organised by the Open Universiteit.

4 Results

An alpha level of .05 was used for all statistical tests.

4.1 Measures Between Groups

Background Questionnaire. To ensure equivalence between treatments, statistical tests were performed on the data collected in the background questionnaire. The procedure exhibited an even distribution in the 3 conditions regarding:

- meta-cognitive capacities, measured with a shortened version [28] of the Meta-cognitive Awareness Inventory [29], $F(2, 134) = .27$, $p = .76$, $\eta p^2 = .004$;
- self-reported familiarity with the topic, measured with a 3-point Likert scale, $\chi^2(2, N = 137) = .36$, $p = .83$;
- self-reported familiarity with eLearning, measured with a 3-point Likert scale, $\chi^2(2, N = 137) = 3.94$, $p = .13$;
- demographics: age $F(2, 134) = .4$, $p = .92$, $\eta p^2 = .07$ ($X = 39$, $SD = 11$), sex $\chi^2(2) = .73$, $p = .69$ (56% female, 44% male), and education level $\chi^2(2, N = 137) = 4.8$, $p = .09$ (75% of the sample ticked the category "Higher education").

Indices. An ANOVA procedure (Table 1) exhibited no significant difference between conditions regarding mean marks obtained at the final test, $F(2, 134) = .44$, $p = .64$, $\eta p^2 = .007$. Significant differences emerged between conditions with regard to the:

- total time spent on the course, $F(2, 134) = 3.49$, $p = .03$, $\eta p^2 = .05$;
- number of page views, $F(2, 134) = 5.29$, $p = .006$, $\eta p^2 = .07$;
- learning efficiency (mark at the test/time spent in the course), $F(2, 134) = 4.76$, $p = .01$, $\eta p^2 = .01$.

Table 1. Mean and standard deviation for the indices common to the three conditions

	Mark at the test			Total time spent on course (in minutes)			Page views			Learning efficiency		
	1 (*N*=34)	2 (*N*=54)	3 (*N*=49)	1	2	3	1	2	3	1	2	3
M	6.4	6	6.4	250	320	330	57	73	84	0.032	0.024	0.022
SD	2.3	1.7	1.8	120	110	110	23	36	44	0.018	0.014	0.012

Additional Fisher contrast tests disclosed that the differences were significant only against the control group and not between the treatments. This lack of observable divergence made it reasonable and beneficial to statistical power and clarity to redefine the treatment conditions as one single group ($N = 103$) for the following analyses.

4.2 Measures Within Treatment Group

Amount of Reflective Enactments. No correlation was found between the mark at the test and the absolute number of annotations (Index 5), characters (Index 6), page views (Index 4) and dashboard views (Index 7).

Rates of Reflective Enactments. Beyond the mere amount of reflective actions (NumberAnnot, CharactInAnnot, NumberPages, Dashvisits), the rates at which these enactments occur while studying might be an important aspect of the reflective activity. For this reason, "reflection rates" were calculated to express the displayed reflection per unit of time (minute) for different indices. These rates were obtained for each individual by dividing the quantity of reflective enactments (the different indices) by the individual time spent in the course (Index: TimeSpent). Based on these ratios, post-hoc splits were applied: subjects were categorized against the mean of the group as either high/low annotators (HA/LA via Index 5), high/low producers of annotation characters (HC/LC via Index 6), high/low browsers (HB/LB via Index 4) and high/low visitors of the learning dashboard (HD/LD via Index 7). For instance, participant 45 took 87 annotations (against an average of 43 for the whole group), produced 13958 characters (against an average of 4792), visited a content page 56 times (against an average of 78), and paid 2 visits to the dashboard (against an average of 3). According to the ratios obtained by dividing these indices by the study time (410 minutes for participant 45 against an average 328 minutes), participant 45 was labelled: HAHCLBLD (High Annotator – High producer of Characters – Low Browser – Low Dashboarder). It was assumed that this fourfold "learning DNA" captured different facets of the participant's reflective engagement with the learning material. Assigning such a multivariate reflective engagement profile to the 103 participants revealed some new insights.

One-index Learning DNA and Mark at the Test. Table 2 displays the performance of high and low groups for each reflective enactments. When treated independently from each others, the indices deliver significant differences for index 4 (annotations), $t(101) = 2.146$, $p = 0.034$, $d = 0.35$ and for index 5 (characters in annotations), $t(101) = 2.76$, $p = 0.007$, $d = 0.35$.

Table 2. High annotators and high producers of characters outperform their low counterparts. This is not the case for high browsers and high dashboarders (index 6).

	Annotations		Characters		Dashboard		Page views	
	Low	High	Low	High	Low	High	Low	High
N	59	44	61	42	71	32	53	50
Score	5,8	6,5	5,7	6,7	6	6,3	6,1	6,1
SD	1,81	1,7	1,8	1,4	1,7	1,9	1,5	2

Two-index Learning DNA and Mark at the Test. Twofold combinations of reflective rates, for instance HA+HB (high annotation rate + high browsing rate) versus HA+LB (high annotation rate + low browsing rate) exhibited significant differences at the omnibus ANOVA, $F(3, 99) = 3.19$, $p = .027$, $\eta p^2 = .088$. Table 3 shows the data for significant cases. Post hoc comparisons using the Fisher LSD test located significant mean differences between HA+HB and LA+HB (p = 0,002), and HA+HB and LA+LB (p = 0,022), but not between HA+HB and HA+LB (p = 0,082).

Table 3. The effect of a high rate of annotation outweighs the effect of browsing rates in twofold learning DNAs.

Twofold learning DNA	Mean mark at the test	SD	N
HA+HB	7.1	1.6	23
HA+LB	6	1.7	21
LA+HB	5.6	1.9	29
LA+LB	5.9	1.6	30

Three-index Learning DNA and Mark at the Test. The attempts made with a profile combining 3 reflection rates gave a significant mark advantage to the most reflective profile (HA+HB+HD) onto all other combinations. However, the creation of such additional combinations induced more numerous groups (Table 4) and quickly created a problem of statistical power that hampered significance tests.

Table 4. Descriptive statistics for more complex learning DNAs

Threefold learning DNA	Mean score at the test	SD	N
HA+HB+HD	7,8	0,9	5
HA+HB+LD	7	1,7	15
HA+LB+HD	5,6	2,1	8
HA+LB+LD	6,4	1,3	15
LA+HB+HD	6,2	1,7	11
LA+HB+LD	5,2	2,1	18
LA+LB+HD	6,3	2,1	8
LA+LB+LD	5,8	1,4	23

4.3 Qualitative Results

The explored qualitative aspects - overall satisfaction, sense of control, perceived intensity of reflection, - were self-reported on 5-point Likert scales in the evaluation questionnaire.

Sense of Control. Mann-Whitney test on the sense of control of the high versus low annotation rates (HA/LA) did not disclose significant differences, $U = 1225$, $p = .61$, $r = .4$. But when the browsing rate was added in the profile, the highly engaged people (HA+HB) reported a significantly higher level of control ($Mdn = 4$) compared to HA+LB ($Mdn = 3$), LA+HB ($Mdn = 3$), LA+LB ($Mdn = 3$), $\chi^2(3, N = 103) = 7.69$, $p = .04$.

Stimulation of Reflection by the Annotation Process. When asked about the effect of taking frequent annotations, 71.2% of the sample answered that reflection increased, 24.6% that it was not influenced and 4.2% that it diminished.

5 Discussion

"Put simply, reflection is about maximising deep and minimising surface approaches to learning." [30, p. 3]. As a strategy to promote deep learning, this study asked learners to use annotations as "reflection amplifiers", i.e. brief and repeated reflection affordances, interspersed in the learning material and activated, through the support of a dedicated widget, in support to the first-order learning task at hand. These stop-and-think episodes could be seen as a tentative instantiation of "split screen learning" [31] that consists in maintaining a dual focus on the content of the lesson and the acquisition processes that are in play. Overall, the results obtained are disappointing.

RAs do expand time on task without delivering benefit for learning achievement: the control group gets the same mark while using less time (see Table 1). Yet, the first hypothesis is not confirmed. From a strict performance-oriented viewpoint, frequent and local annotations are counter-productive.

No noticeable performance appeared between distinct uses of the annotation tool: free annotation versus question-based annotation (see section Indices). Yet, the second hypothesis is not confirmed.

This study delivers however one unexpected and intriguing pattern when the analysis operates 2 shifts of focus (see section 4.2): a) from one single reflective enactment taken in isolation (writing annotations) to multiple reflective enactments (producing a longer annotation, navigating amongst course pages, visiting one's learning dashboard), and b) from the mere amount of these reflective enactments to their rates of use. In this case, results provide insights about ways students balance and combine the primary activity (studying the course) and the secondary reflective activities (annotations, page re-visits, dashboard views). Here, a different pattern emerges: students who write more personal annotations per unit of time than the average get a higher mark. Combinations of this reflection rate with other reflective enactments (page views, dashboard views) bring extra benefits to performance. The qualitative data also

seems to be affected by combined reflective rates: a significant effect on student's sense of control is obtained only from high aggregated rates of reflective enactments (see section "Sense of Control").

On this basis, it can be advanced that the dynamics of reflective commitment to a study task encompasses and interweaves several reflective enactments performed at a certain rhythm. It is possible that the reflective passivity of some students might be counteracted by inviting them not only to deploy more reflective actions on the material but also to accelerate their frequency.

6 Further Work

Four main issues raised in this study call for further research.

RAs and Performance. Although performance tests are not the only way to measure learning, it remains a legitimate and largely-practised way to assess mastery of content. In this perspective, final scores should reasonably be expected to reflect benefits resulting from using RAs. It has not happened here, at least in comparison with the control group. This lack of benefits from annotations contrasts with other studies in the field [32, 33, 34]. Further empirical studies can help to sort out what the effects of annotations "ought to be" from what they actually accomplish, and most importantly, in what instructional context and for what kind of learners.

RAs and Intellectual Dynamics. On a more fundamental level, the study findings, and especially those related to the effects of combined reflective enactment rates (annotations, revision of annotations, page re-visits), highlight the intellectual dynamics at work in a deep approach of study material. Further investigation is required to establish whether it could be a characteristic of high achievers and a hallmark of intellectual life in general to operate an "active study", defined as an ongoing crisscrossing between a primary learning activity and secondary reflective or meta-cognitive enactments. This periodic and persistent to-and-fro mental move offers a very different aspect of reflection than the one conveyed by portfolios or learning diaries wherein the experience and the reflection thereon are temporally distinct. If metaphors can be invoked, their complementarity presents as Lego blocks while the real-time dynamics of learning would look more after pinballs. Practical ways to evidence and sustain this interplay between cognitive and reflective landscapes [31] must be explored. Investigation of these constant shifts from a primary activity to secondary reflective activities might also benefit from the literature on interrupted tasks [35] since a stop-and-think reflective break can in certain circumstances break the flow or productively coalesce. In this context, interruption rates and optimums [36] become a pivotal notion.

RAs and Personalized Learning. An important issue for future research on annotations is also tied to the selection of relevant frames to ascertain their effects. More research needs to be undertaken to see if prompts used to amplify reflective appraisal of the study material can be related to ownership of learning and sharper feelings that learning is or becomes a "personal matter". In such an approach, personalised learn-

ing might be seen more as a consequence of seizing action and reflection affordances (37, 38, p. 151) and less as the result of a decision taken by an external agent like a teacher or an adaptive system.

RAs and vocabulary. Research in the field will benefit from a closer inspection of the use of the word "reflection" to describe cognitive operations in which the students engage. An observation of the literature devoted to reflection-in-action prompts is up to show how varied the mental processes called reflection can be. For instance, on the one hand can the kind of process invoked in the annotations be called "reflection" or does it resort to "analytical scrutiny". On the other hand, it does not seem unacceptable to consider the step backward implied by an annotation as a form of reflection. Beyond "processes" – since one always reflects "on" something – improved specifications of "objects" of reflection critical for the growth of professional learners are needed to guide instructors in concrete instructional designs.

7 Limitations of the Study

While the results related to hypotheses 1 and 2 have been obtained through acknowledged statistical procedures and are, on this basis, totally sound, those related to the compound score of reflective rates must be taken with caution as they emerge from the pooling of the treatment groups (free annotations/question-based annotations) and the use of dichotomized data (e.g. High versus low users of the dashboard), two methodological decisions that can be discussed. With regard to pooling, Verma, Gagliardin, and Ferretti [39] highlight that a "sufficient" degree of comparability is a precondition for such pooling to be meaningful". In this case, the comparability of the treatment groups at baseline, the identical assignment of taking a note at each page visit made to both groups, and the similar results obtained against the control group were considered sufficient to justify the pooling in an exploratory perspective. With regard to dichotomization, DeCoster, Iselin, and Gallucci [40] conclude their comprehensive review on the issue by pointing specific circumstances wherein cut-off points are acceptable, among which "the purpose of the research is to investigate how a dichotomized measure will perform in the field" (p. 364). This is the case with the work on reflective rates which has no other purpose than exploring whether frequent annotations, with no effect taken separately (hypotheses 1 and 2 are not confirmed), may, combined to other reflective enactments, reach a certain threshold, at which point they have tangible effect. The use of cut-off points, as obtained through dichotomized indicators, was also influenced by readability concerns: "big categories" (high versus low) help to prevent a "drowning by numbers" effect which can go along with unusual and complex variables such as browsing or annotation rates. Based on these motives, it appeared reasonable to proceed as described even though the use of regression analysis might have been a more straightforward method.

The decision of not to analyse the content of the annotations to compare it to the course content can also be considered as a second caveat, especially when Natural Language Processing tools, like Latent Semantic Analysis (LSA) or Latent Dirichlet Allocation (LDA), provide sufficiently good measures of it. This approach has been left to further studies for 2 reasons. On the one hand, measuring the quality of an an-

notation is difficult in itself because the learner's cognitive context around it does exist but cannot always be grasped by the researcher. Furthermore, the neglect of the fine-grained qualitative aspects was a decision flowing from the initial scope of study. Its chief postulate is that quality learning is encouraged by a permanent criss-crossing between an ongoing learning processes and explicit/structured episodes of reflection. The experimental design attempted to reproduce this intertwine somewhat artificially (that experimental design simply failed at supporting this dynamic cannot be excluded) with the annotation RA. The work was therefore more acquainted with quantitative measures than with qualitative ones. To address the latter, other instruments and methodologies – out of the initial scope of the study – should enrich the research effort. They would give insights into the actual engagement with the RA and into the quality level of the prompted reflection. Some quality check of the content of the annotations, by students and/or researchers, could also enrich the set-up, for instance by have a group where students try to answer the annotation questions just before passing the test or through annotation sharing mechanisms.

The static nature of the RA used in this study is also a limitation. The annotation tool presents as a neutral artefact that becomes available in a pre-defined way. The RA remains ignorant of the profile or the learning activities carried out by the student. In this context of self-instruction, the reflective activity is also deprived of feedback. All these limitations may qualify this RA in this context as simply inappropriate or insufficient.

Lastly, the present study takes place in a real-world context with the highs and lows of this approach. Yet, the direction of the effect in each hypothesis may be put in question as some hidden variables (e.g. motivation, self-efficacy, availability for a course taken on a voluntary basis, in the changing situations of life, etc.), beyond those controlled (see section Background questionnaire), may work as moderators.

8 Concluding remarks

A growing literature extols the importance to instil reflection and deep approaches to learning in tuition. However, practical and systematic ways to operate are not conspicuous, at least when it comes to reflection in methods of learning considered as traditional or transmissive [41], in contrast to constructivist methods (problem-based learning, collaborative learning) wherein reflection is claimed to be "built-in" [42, 43, 44]. This study inquired the question: how to induce a more thoughtful autonomous study of learning material? To answer, the experimental setting artificially increased the number of annotations, conceptualised as frequent tinglings for reflection while reading and purposed to support a persistent dynamic mental engagement with the reading material. An assumption guided this work: that such a kind of active and reflective posture to learning, which constantly articulates the cognitive and the metacognitive landscapes, is a key feature of intellectual life. The experimental setting presented here was a simplified attempt to mimic and externalize such fundamental inner dynamic processes via an annotation tool. Eventually, annotations taken alone did not really measure up. However, some elements of the study suggest that the frequency and the aggregation of different reflective behaviours can be worth exploring further in connection to quality learning.

References

1. Jackson, H. J.: Marginalia: Readers writing in Books. Yale University Press, New Haven, (2001)
2. Crawford, C. C.: Some experimental studies of the results of college note-taking. Journal of Educational Research,12, 379-386 (1925)
3. Schön, D.: The Reflective Practitioner: How professionals think in action. Temple Smith, London (1983)
4. Commonwealth: Creating learning materials for open and distance learning - A handbook for authors and instructional designers. Commonwealth of Learning, Vancouver (2005)
5. Open University: Guide for the Appointment of Tutorial and Counselling Staff. Open Teaching Toolkits, Open University (1991-1994)
6. Dillon, A.: Reading from paper versus screens: a critical review of the empirical literature. Ergonomics 35(10), 1297-1326 (1992). doi: 10.1080/00140139208967394
7. Priolli, P. (2007): Information Foraging Theory: Adaptive Interaction with Information. Oxford University Press, Oxford-New York (2007)
8. Boch, F., Piolat, A.: Note taking and learning: A summary of research. The WAC Journal 16, 101-113 (2005)
9. Hartley, J., Davies, I. K.: Note-taking: A critical review. Innovations in Education & Training International 15(3), 207-224 (1978)
10. Slotte, V., Lonka, K.: Note-taking review – Practical value for learners. Arob@se 1-2, 79-86 (2003)
11. Makany, T., Kemp, J., Dror, I. E.: Optimising the use of note-taking as an external cognitive aid for increasing learning. British Journal of Educational Technology 40(4), 619-635 (2009). doi: 10.1111/j.1467-8535.2008.00906.x
12. Degani, A., Wiener, E. L.: Human factor of flight-deck checklists: The normal checklist. NASA, Moffett Field (1990).
13. Van der Baaren, J., Schuwer, R., Kirschner, P., Hendriks, M.: Finding your way into an open online learning community. Journal of Interactive Media in Education, 4 (2008).
14. Hwang, W.-Y., Chen, N.-S., Shadiev, R., Li, J.-S.: Effects of reviewing annotations and homework solutions on math learning achievement. British Journal of Educational Technology 42(6), 1016-1028 (2011).
15. Kam, M., Wang, J., Iles, A., Tse, E., Chiu, J., Glaser, D., Canny, J.: Livenotes: a system for cooperative and augmented note-taking in lectures. Paper presented at the CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA (2005)
16. Glahn, C.: Contextual support of social engagement and reflection on the Web. PhD series, Open Universiteit, Heerlen, The Netherlands (2009). https://www.ou.nl/Docs/Onderzoek/Promoties/2009/Christian_Glahn_thesis.pdf
17. Verpoorten, D., Glahn, C., Chatti, A., Westera, W., Specht, M.: Self-Reported Learning Effects of a Tagging Activity Carried out in a Personal Learning Environment (PLE) by Secondary-School Pupils. International Journal for Cross-Disciplinary Subjects in Education 2(1), 276-284 (2011).
18. Schilit, B., Golovchinsky, G., Price, M.: Beyond paper: supporting active reading with free form digital ink annotations. Paper presented at the Conference on Human factors in computing systems, Los Angeles, CA, USA (1998). http://orbi.ulg.ac.be/handle/2268/151959
19. Kawase, R., Papadakis, G., Herder, E., Nejdl, W.: Creation and Use of Annotations in Different Usage Contexts. L3S Research Center Hannover (NA). http://stellarnet.eu/d/1/2/images/a/ae/Sss1.pdf

20. Hwang, W.-Y., Hsu, G.-L.: The Effects of Pre-Reading and Sharing Mechanisms on Learning with the Use of Annotations. Turkish Online Journal of Educational Technology – TOJET 10(2), 234-249 (2011). http://files.eric.ed.gov/fulltext/EJ932242.pdf
21. Glover, I., Xu, Z., Hardaker, G. Online annotation – Research and practices. Computers & Education 49(4), 1308-1320 (2007). doi: 10.1016/j.compedu.2006.02.006
22. Verpoorten, D., Westera, W., Specht, M.: Reflection amplifiers in online courses: a classification framework. Journal of Interactive Learning Research 21(4), 654-666 (2010) http://orbi.ulg.ac.be/handle/2268/151374
23. Boud, D., Keogh, R., & Walker, D.: Reflection, Turning Experience into Learning. Kogan Page, London (1985)
24. Butler, D. L., Winne, P. H.: Feedback and Self-Regulated Learning: A Theoretical Synthesis. Review of Educational Research 65(3), 245-281 (1995)
25. Verpoorten, D., Westera, W., Specht, M.: Infusing reflective practice in eLearning courses - can widgets help? International Journal of Technology Enhanced Learning 3(1), 93-109109 (2011). http://orbi.ulg.ac.be/handle/2268/151348
26. Eshuis, J.H., Goltstein, G.P.: Seks en de evolutie. Open Universiteit Nederland, Heerlen (2007). http://www.ou.nl/eCache/DEF/2/22/850.html
27. Verpoorten, D., Westera, W., Specht, M.: A first approach to "Learning Dashboards" in formal learning contexts. Paper presented at the ADVTEL workshop (1st International Workshop on Enhancing Learning with Ambient Displays and Visualization Techniques) at EC-TEL 2011, Palermo, Italy (2011). http://orbi.ulg.ac.be/handle/2268/151988
28. Bijker, M., Van der Klink, M., Boshuizen, H.: Modeling self-efficacy, self-regulation, self-directed learning and career processes of adult professionals and relations with learning outcomes and labour market success. Paper presented at the 5th EARLI-SIG14 Learning and Professional Development, Munich, Germany (2010)
29. Schraw, G., Dennison, R.: Assessing metacognitive awareness. Contemporary Educational Psychology 19, 460-475 (1994).
30. Hinett, K.: Improving Learning through Reflection.York: Higher Education Academy, York (2002)
31. Claxton, G.: Expanding the capacity to learn: a new end for education? Opening keynote address presented at the British Educational Research Association Annual Conference, University of Warwick, UK. (2006). http://www.buildinglearningpower.co.uk/includes/pdf_library/GC_BERA_Keynote.pdf
32. Hwang, W.-Y., Wang, C.-Y., Sharples, M.: A Study of Multimedia Annotation of Web-Based Materials. Computers and Education, 48(4), 680-699 (2007)
33. Nokelainen, P., Miettinen, M., Kurhila, J., Floréen, P., Tirri, H. : A shared document-based annotation tool to support learner-centred collaborative learning. British Journal of Educational Technology 36(5), 757-770 (2005)
34. Slotte, V., Lonka, K.: Review and process effects of spontaneous note-taking on text comprehension. Contemporary Educational Psychology, 24, 1-20 (1999)
35. Mark, G., Gudith, D., Klocke, U. The cost of interrupted work: More speed and stress. Paper presented at the 26th annual SIGCHI Conference on Human factors in computing systems, Florence, Italy (2008). doi: 10.1145/1357054.1357072
36. Palanque, P., Winckler, M., Martinie, C.: A Formal Model-Based Approach for Designing Interruptions-Tolerant Advanced User Interfaces. Computational Intelligence 340, 143-169 (2011). doi 10.1007/978-3-642-14562-9_8
37. Verpoorten, D., Glahn, C., Kravcik, M., Ternier, S., Specht, M.: Personalisation of learning in virtual learning environments. Learning in the Synergy of Multiple Disciplines. Pro-

ceedings 4th European Conference on Technology Enhanced Learning, EC-TEL 2009, 52-66 (2009). http://orbi.ulg.ac.be/handle/2268/151458
38. Crawford, M.: The world beyond your head. How to flourish in an age of distraction. Viking, Great Britain (2015)
39. Verma, V., Gagliardi, F., Ferretti, C.: On pooling of data and measures. DMQ 84 (2009)
40. DeCoster, J., Iselin, A.-M. R., Gallucci, M.: A conceptual and empirical examination of justifications for dichotomization. Psychological Methods 14(4), 349-366 (2009). doi: 10.1037/a0016956
41. Terhart, E.: Constructivism and teaching: a new paradigm in general didactics? Journal of Curriculum Studies 35(1), 25-44 (2003).
42. Kirschner, P. A., Sweller, J., Clark, R. E.: 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 (2006)
43. Hmelo-Silver, C. E., Duncan, R. G., Chinn, C. A.: Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). Educational Psychologist 42(2), 99-107 (2007). doi: 10.1080/00461520701263368
44. Sweller, J., Kirschner, P. A., Clark, R. E.: Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries. Educational Psychologist 42(2), 115-121 (2007). doi: 10.1080/00461520701263426