Writing processes as situated regulation processes: A context-based approach to doctoral writing

Sala-Bubaré, A.; Castelló, M.; Rijlaarsdam, G.

DOI
10.17239/JOWR-2021.13.01.01

Publication date
2021

Document Version
Final published version

Published in
Journal of Writing Research

Writing processes as situated regulation processes: A context-based approach to doctoral writing

Anna Sala-Bubaré*, Montserrat Castelló* & Gert Rijlaarsdam**

*Facultat de Psicologia, Ciències de l'Educació i de l'Esport Blanquerna, Universitat Ramon Llull | Spain

**University of Amsterdam | The Netherlands

Abstract: Doctoral students face many challenges when writing research articles. However, little is known about how they regulate their writing process in a natural context, due partially to the lack of methods to explore writing regulation from a situated perspective. The present study aims at demonstrating a method to explore doctoral students' writing regulation processes within their context of occurrence in ecological conditions. To do so, we focus on the writing process of Natalia, a second-year doctoral student, while she writes and revises an extended abstract for her first scientific article under natural conditions. Screen-recorder and keystroke logging software, writing logs, an open-ended questionnaire and drafts of her text were used to collect data about the processes and products, and about both her actions and perceptions. Analysis combining these different data allowed us to identify two types of episodes: production and regulation episodes, and six subtypes of regulation episodes, and link them to the section of the text and the challenges the writer addressed with each episode. Results also showed that regulation processes vary between sessions, in terms of frequency and in their goals, and that feedback promoted a problem-solving approach to writing.

Keywords: doctoral writing, regulation processes, writing regulation, research writing, situated process, key-stroke logging

journal of WRITING RESEARCH

Sala-Bubaré, A., Castelló, M., & Rijlaarsdam, G. (2021). Writing processes as situated regulation processes: A context-based approach to doctoral writing. *Journal of Writing Research, 13*(1), 1-30. https://doi.org/10.17239/jowr-2021.13.01.01

Contact: Anna Sala-Bubaré, Universitat Ramon Llull/Facultat de Psicologia, Ciències de l'Educació i de l'Esport Blanquerna, C/ Císter 34, Barcelona, 08022 | Spain – annasb4@blanquerna.url.edu - ORCID: 0000-0003-1733-2063

1. Introduction

Writing research articles (RAs) poses all kinds of challenges to doctoral students. Some of them are related to the need to learn a new genre and the disciplinary discourse (Simpson, 2009; Swales & Feak, 2004), such as the organization of information and the use of terminology and concepts. Others concern the management of the recursivity and complexity of the cognitive processes involved in writing (Castelló, Iñesta, & Corcelles, 2013; Lonka, et al., 2014) or the development of their voice and identity as authors and researchers (Castelló & Iñesta, 2012; Chang & Schleppegrell, 2016; Maher, et al., 2008; Paré, 2011; Pedrazzini, Bautista, Scheuer, & Monereo, 2014). Finally, obtaining and managing support and feedback from supervisors and other researchers are also potential sources of challenge (Aitchison, Catterall, Ross, & Burgin, 2012; Caffarella & Barnett, 2000; Cotterall, 2013). These challenges not only affect the design and planning of RAs but also have a significant impact on the writing process at different textual levels, from word level to the whole text, and require doctoral students to regulate their activity to try to overcome them and succeed in their writing objectives.

We define writing regulation as a highly recursive and dynamic socially situated activity that takes places at all textual levels and throughout the writing process (Ferrari, Bouffard, & Rainville, 1998; Hadwin & Oshige, 2011; van den Bergh, Rijlaarsdam, & van Steendam, 2016). It is composed of explicit decision-making processes, but also implicit adjustments (Iñesta & Castelló, 2012). Research on doctoral writing, although extensive, has mainly relied on writers' self-reports to identify doctoral students' challenges and strategies (Lindsay, 2017; Lonka, et al., 2014; Odena & Burgess, 2017; Wisker, 2015) and to develop pedagogies to support them (Caffarella & Barnett, 2000; Cotos, Huffman, & Link, 2020; González-Ocampo & Castelló, 2018; Maher, et al., 2008; Papen & Thériault, 2018; Pedrazzini, et al., 2014, among others). Less attention has been paid to doctoral students' writing regulation (Sala-Bubaré & Castelló, 2018). More specifically, there is a lack of studies exploring these issues from a micro perspective, that is, observing doctoral students' writing processes synchronously.

In other writing research contexts, studies show the great potential synchronous data collection tools can have in accessing and capturing the complexity of the writing regulation processes (Baaijen, Galbraith, & de Glopper, 2012; Franklin & Hermsen, 2014; Leijten & van Waes, 2013; Sullivan & Lindgren, 2006; van den Bergh & Rijlaarsdam, 2001; van den Bergh, et al., 2016). Of particular interest for the present study is the keystroke logging method, as it 'provides an unobtrusive record of the moment-by-moment creation of the text' (Baaijen, et al., 2012). In particular, it preserves the natural settings and conditions in which doctoral students normally write, in contrast of other synchronous data collection methods such as think-aloud protocols and double tasks (Leijten & van Waes, 2013).

Keystroke logging has been used in writing research for over three decades (Miller & Sullivan, 2006) to uncover the complex cognitive processes underlying writers' behaviour. Pauses and revisions during the writing process have been used to infer different components and aspects, such as planning and revising, which have been also linked to cognitive load and text quality (Dragsted & Carl, 2013; van Waes, & Schellens, 2003). These methods have been used to assess relationships between components, inter- and intra-individual (e.g. L1 vs L2) differences, writing development, writing difficulties and professional writing (Leijten & van Waes, 2013). In recent years, they have also been employed in education to promote students' awareness about their writing processes (Lindgren, Knospe, & Sullivan, 2019). Keystroke studies have focused on a wide variety of writers too, from young primary students (Gnach, Wiesner, Bertschi-Kaufmann, & Perrin, 2007) to elder people (Van Waes, Leijten, Mariën, & Engelborghs, 2017), mostly through experimental designs (Lindgren, et al., 2019; Sala-Bubaré & Castelló, 2018). Overall, these studies have enabled researchers and theorists to revise and expand the models of writing (Lindgren, et al., 2019; MacArthur & Graham, 2016).

Nevertheless, the interpretation of keystroke loggings is not unproblematic and requires varying degrees of inference, as they only provide indirect information about writers' cognitive processes (Baaijen & Galbraith, 2018; Miller, Lindgren, & Sullivan, 2008; Wengelin, et al., 2009). Hence, there is a need to combine keystroke logging with other measures and instruments (Baaijen & Galbraith, 2018; Wengelin, et al., 2009). Keystroke studies, both alone and in combination with other observational techniques such as eye-tracking devices and think-aloud protocols (Leijten & van Waes, 2013), have advanced knowledge about the type of processes involved in writing. However, most of them draw on cognitive perspectives, which focus on the frequency, duration and distribution of writing behaviour, but do not account for the situation (e.g. the communicative situation characteristics, authors' aims, the text and its content or semiotic tools) in which writing processes take place. In particular, they cannot explain how these situated writing characteristics impact on highly context-dependent writing regulation processes (Sala-Bubaré & Castelló, 2018).

Drawing on a socially situated perspective, Iñesta & Castelló (2012) developed a unit of analysis called *Regulation Episode* (RE) aimed at characterising and analysing writing processes as a complex, dynamic, situated and social activity. The authors define the regulation episode as the actions that writers implement to solve a difficulty or challenge identified during the writing process. It addresses the complex, recursive and socially situated nature of the regulation processes (Hadwin & Oshige, 2011; Negretti & Mežek, 2019) while maintaining the focus on the problem-solving process (Flower & Hayes, 1981). In their analysis of the writing process of two experienced academics, Iñesta & Castelló (2012) found that Regulation Episodes helped identify patterns in the writing activity. In particular, they enabled researchers to characterise the recursivity of the writing processes

and to unveil implicit challenges the authors encountered, most of which were related to authors' voice and the social dimension of academic writing.

The present study aims at demonstrating a method to explore doctoral students' writing regulation processes within their context of occurrence and in ecological conditions. We adapted the *Regulation Episode* (Iñesta & Castelló, 2012) to analyse writing regulation processes using keystroke logging in combination with other unobtrusive data collection methods. More specifically, we employed video recordings of participant's computer screen, a questionnaire about writing, writing logs and audio recordings of the discussion about the first draft, and the two drafts of the text. This methodology allowed us to explore an authentic text and preserve the ecological conditions of doctoral writing.

We demonstrate how we explored the micro-regulation processes of a doctoral student while she wrote the initial drafts of her first research article, and whether and how these processes are influenced by text discussion and feedback in the context of a writing workshop.

2. Context: Academic writing course

The study took place in an elective academic writing workshop offered to all second- and third-year doctoral students in a Social Sciences faculty of a Catalan university. The aim of the workshop was twofold: to promote PhD students' understanding of the characteristics of the RA genre and develop new strategies to manage their writing processes[1] and help them write their first research article. To enrol on the course, doctoral students were required to have analysed the data for the article they wanted to write in the workshop. Six doctoral students enrolled on the course. They met fortnightly in eight three-hour sessions.

The workshop was structured around the writing of their first research article, the only writing task participants completed during the workshop (with the exception of the writing logs). This was an authentic task, as it is a requirement for doctoral students to finish their PhD studies. They have - and want - to write and publish an article. Participants write, revise and discuss texts and the writing process in class with the instructors and peers. They start by writing an extended abstract of the article and revise and extend their text throughout the eight sessions of the workshop with the help of their peers and instructors. The use of extended abstract as initial drafts of their paper has proved useful in previous editions of the workshop as, since it had to comprise all the sections of a paper, it served as an outline of the article they wanted to write. Moreover, it prompts writers to keep a global, integrated view of their text and allows for discussion about the coherence of, and relationship between, the different parts of a research article.

More specifically, in the first session of the workshop, after a brief presentation of the objectives and syllabus of the course, participants were asked to write the extended abstract (around 1,000 words). In the next workshop session,

peers and workshop facilitators commented on all participants' first drafts in an open discussion, and then they revised their text based on the feedback they received.

This writing-feedback-revision cycle was done three times throughout the workshop. Due to the complexity of the methodological apparatus and the amount of data collected, both described below, the present study focuses on the writing processes of one doctoral student, Natalia, in the first of one such writing-feedback-revision cycle.

3. Method

We designed an exploratory mixed-method case study, which allowed us to get a deeper understanding of a doctoral student's writing regulation processes under ecological conditions: a research setting without time restrictions and interruptions and in which they had access to all the resources they would normally use (e.g., computer, academic articles, internet…). Doctoral researchers work and write in a variety of sites (McAlpine & Mitra, 2015), such as labs, home and libraries. The setting for this study resembles the kind of setting of writing retreats: participants writing their individual papers together in a room for a given time (see, for instance, Murray & Newton, 2009).

Moreover, the study was designed to use multiple methods and data of different nature. We collected data about both the writing process and its products, and about both the participant's actions and perceptions about these actions, which allowed us to get a contrasted and comprehensive understanding. We used the four methods described by Hyland (2016): elicitation (initial questionnaire and group discussion about the feedback), introspection (authors' writing logs), observation (keystroke logging and screen-recorder software), and text data (drafts).

3.1 The writer: Natalia

Natalia[2] was a 26 years-old full-time doctoral student enrolled in the second year of the doctorate in education. According to her responses on the initial questionnaire[3], she liked writing and saw herself as a writer and a researcher-member of a research group. She anticipated difficulties in 'improvising' and using her own words, as she did in her personal writing, due to feelings of insecurity. Her motives to write a research article related to fulfilling the requirement of the doctoral program, but also to disseminate her research and learn how to write better articles. She was oriented towards an academic career.

She had not published a research paper before, although she had presented partial results of her studies in conferences, and therefore, she had some partial texts of the article she wanted to write. Natalia saw writing as essential in the research process, but only related to the products and communication of results. At the time

of the data collection, she still had not started thinking about which journal she would submit her article. She gave her consent to participate in the study.

3.2 Data sources

Various instruments and procedures were used to capture both the writing processes and products, and Natalia's perceptions and actions. The resulting data sources were:

- *Initial questionnaire "Certainties and doubts"* (adapted from Castelló, et al., 2013): fourteen open-ended questions exploring her perceptions and feelings about academic writing and RA, and of herself as a research writer (see Appendix A).
- *Pre- and post-writing log*: semi-structured questions about the objectives and expectations about each session (pre-); and about the problems faced, the attempted solutions, the resources used and the level of satisfaction with their work at the end of each session (post-) (see Appendix B).
- *'Inputlog'* (Leijten & van Waes, 2013): Inputlog is a keystroke logging piece of software used in writing research to capture keystrokes and mouse movement. It is a non-intrusive instrument that allows writers to work in a Microsoft Word processor, starting either with a new or a pre-existing document. Inputlog collects fine-grained data about all the key and mouse actions, as well as the shifts between windows (e.g. browser, computer folders, etc.). It also features other modules and options to refine, analyse and integrate data.
- *Screen recorder 'SnagIt'®* (Techsmith Corporation ®, Techsmith, n.d.): *Snagit* is a screen capture and video recording piece of software that can be used to non-intrusively record writers' computer screen while they are composing their texts. Screen recordings complemented Inputlog data. Keystroke loggings of long writing sessions are difficult to interpret without additional data and provide no information about the content of the sources used by writers. Screen recordings allowed us to better understand and interpret the huge amount of data in the keystroke logging, to see the moment-by-moment unfolding of the text, and provided direct access to the sources the participant used and their content.
- *Audio recording* of the feedback session. All participants and workshop facilitators were involved in a discussion about the texts in which they provided feedback to each other and this conversation was recorded.
- *Drafts of the extended abstract*. We kept the draft of the extended abstract produced in each writing session.

The combination of data sources was crucial to gain a full picture of the regulation processes within the context. The synchronous data collection techniques (keystroke loggings and screen capture software) provided information about the moment-by-moment creation of the text and the resources used to that end, while the initial questionnaire and the writing logs provided crucial insights about the context of writing: the mental context of the individual writer (e.g. personal objectives and fears), her assessment and perception about the process and the product of writing and, more generally, her perception of herself as a research writer. The audio recording of the feedback session also provided access to the problems Natalia encountered and the rationale for some of the decisions taken, as well as to the feedback comments.

3.3 Data collection

The data collection process took place in the first three out of eight sessions of the writing workshop. In this section, we explain the how we collected information about Natalia's writing process of the first drafts of the extended abstract of the article she wanted to write in the course. This type of text was also suitable for our research purposes because it is a relatively long text and challenging enough for doctoral students, so regulation processes were likely to appear. The extended writing process and the multi-authored nature of doctoral students' articles make it impossible to monitor the writing process of the whole article.

Session 1. Natalia was told in advance to bring all the documents she had about the paper she wanted to write. She had access to all these documents and any resources she needed, including internet. Two in-class hours were allocated to writing the text. She could ask for more time if she needed (*first writing session*). Inputlog and the screen recording software were activated on her computer to record her writing activity. Natalia completed the pre-writing log and wrote her text. When she finished writing, she completed the post-writing log and the initial questionnaire "Certainties and doubts"[4]. During this session, workshop facilitators sat in the same room and solved technical problems with the software if needed.

Session 2. After two days, workshop participants including Natalia received a peer's text and were asked to read and comment it. In the second session of the workshop (*feedback session*), two weeks later, peers and workshop facilitators, discussed and commented on each participant's text.

Session 3. Finally, Natalia revised her text with the support of the feedback provided in the previous session (*second writing session*). The same procedure was followed as in session 1. She completed the pre- and post-writing log and revised her text while keystroke logging and screen recording software were running in her computer.

3.4 Data analysis

Inputlog provides various options and features of data analysis, including pause, S-notation and fluency analysis among other advanced analyses which are based on algorithmic processing. As we wanted to conduct a qualitative and in-depth analysis of the participant's process and regulation episodes, these analyses were not suited for this study. Instead, we used the *general analysis*, which provides a log file in a .csv extension that we exported to Excel. In this excel document, each row contains an input action (a keystroke, a mouse movement, a change of window) with detailed information, including the output, position in the document, start and end time and pause before (see Figure 1). Logs contained 17,792 input actions in Natalia's first session and 15,432 input actions in the second writing session.

id	Part	type	output	positionFul	ColumnsI	docLengthFul	charProductNr	startTime	startClock	endTime	endClock	timeError	actionTime	pauseTime	pauseLocation	pauseLocationFul	RevisionType	x	y	RawStart	RawEnd	positior	docLength	Colum	Colum
5860		focus	Wordlog	1367	0	1600	1861	1841638	0.0213153	1841638	0.0213153	0	0	0	10	CHANGE	PRODUCTION			1841638	1841638				extended abstract
5861		mouse	Movemei	1367	-1367	1600	1861	1841638	0.0213152	1841794	0.0213171	4.20718E-05	156	0	16	MOUSE	PRODUCTION	750	511	1841638	1841794				extended abstract
5862		mouse	Movemei	1367	-1367	1600	1861	1845273	0.0213573	1845866	0.0213642	8.85417E-06	593	3479	16	MOUSE	PRODUCTION	777	615	1845273	1845866				extended abstract
5863		mouse	LEFT Clicl	1367	-1367	1600	1861	1846038	0.0213668	1846241	0.0213685	1.53472E-05	203	172	16	MOUSE	PRODUCTION	777	615	1846038	1846241				extended abstract
5865		mouse	Movemei	1367	-1367	1600	1861	1847364	0.0213813	1848986	0.0214003	2.14815E-05	1622	1123	16	MOUSE	PRODUCTION	869	549	1847364	1848986				extended abstract
5866		mouse	LEFT Clicl	1367	-1367	1600	1861	1849220	0.0214001	1849407	0.0214052	5.96565E-06	187	234	16	MOUSE	PRODUCTION	869	549	1849220	1849407				extended abstract
5868		mouse	Movemei	1367	-1367	1600	1861	1849735	0.0214087	1852715	0.0214435	3.61921E-05	2980	328	16	MOUSE	PRODUCTION	839	560	1849735	1852715				extended abstract
5869		keyboard	ALT	1368	-1367	1601	1862	1852862	0.0214416	1853022	0.021447	9.83766E-07	160	147	12	COMBINATION KEY	PRODUCTION			1852862	1853022	1368	1601		extended abstract
5870		keyboard	TAB	1368	-1368	1601	1862	1852947	0.0214415	1853103	0.021448	1.80556E-06	156	85	12	COMBINATION KEY	PRODUCTION			1852947	1853103	1368	1601		extended abstract
5872		focus	PaperNor	1368	-1368	1601	1862	1853103	0.0214479	1853103	0.021448	1.45949E-05	0	0	10	CHANGE	PRODUCTION			1853103	1853103				newmodel paper
5873		keyboard	ALT	1368	-1368	1601	1862	1854364	0.0214625	1854613	0.0214654	1.41204E-06	249	0	12	COMBINATION KEY	PRODUCTION			1854364	1854613				newmodel paper
5874		keyboard	TAB	1368	-1368	1601	1862	1854486	0.0214636	1854640	0.0214657	1.78241E-06	154	122	12	COMBINATION KEY	PRODUCTION			1854486	1854640				newmodel paper
5876		focus	Wordlog	1368	-1368	1601	1862	1854640	0.0214657	1854640	0.0214657	1.09239E-05	0	0	10	CHANGE	PRODUCTION			1854640	1854640				extended abstract
5877		keyboard	t	1368	-1368	1601	1862	1855584	0.0214767	1855701	0.021478	1.22685E-06	117	0	2	BEFORE WORDS	PRODUCTION			1855584	1855701	1368	1601		extended abstract
5878		keyboard	h	1369	-1368	1602	1863	1855890	0.0214739	1855801	0.0214792	1.40594E-06	111	106	1	WITHIN WORDS	PRODUCTION			1855690	1855801	1369	1602		extended abstract
5879		keyboard	e	1370	-1369	1603	1864	1855811	0.0214729	1855928	0.0214806	9.83796E-07	117	121	1	WITHIN WORDS	PRODUCTION			1855811	1855928	1370	1603		extended abstract
5880		keyboard	SPACE	1371	-1370	1604	1865	1855896	0.0214808	1855996	0.0214814	2.71991E-06	100	85	3	AFTER WORDS	PRODUCTION			1855896	1855996	1371	1604		extended abstract
5881		keyboard	p	1372	-1371	1605	1866	1856131	0.021483	1856242	0.0214843	9.83796E-07	111	235	2	BEFORE WORDS	PRODUCTION			1856131	1856242	1372	1605		extended abstract
5882		keyboard	r	1373	-1372	1606	1867	1856216	0.0214839	1856311	0.0214851	4.23611E-06	95	85	1	WITHIN WORDS	PRODUCTION			1856216	1856311	1373	1606		extended abstract
5883		keyboard	BACK	1374	-1373	1607	1868	1856582	0.0214882	1856643	0.0214889	1.33132E-06	63	366	11	REVISION	PRODUCTION			1856582	1856643	1374	1607	1	extended abstract
5884		keyboard	BACK	1373	-1374	1606	1868	1856697	0.0214855	1856792	0.0214906	2.1412E-06	95	115	11	REVISION	PRODUCTION			1856697	1856792	1373	1606	1	extended abstract
5885		keyboard	BACK	1372	-1373	1605	1868	1856882	0.0214919	1856935	0.0214923	7.8247E-06	53	185	11	REVISION	PRODUCTION			1856882	1856935	1372	1605	1	extended abstract
5886		keyboard	SPACE	1371	-1372	1604	1868	1857558	0.0214951	1857655	0.0215006	9.14352E-07	97	676	3	AFTER WORDS	PRODUCTION			1857558	1857655	1371	1604		extended abstract
5887		keyboard	CAPS LOC	1372	-1371	1605	1869	1857637	0.0215043	1857750	0.0215017	1.81713E-06	113	79	15	UNKNOWN	PRODUCTION			1857637	1857750	1372	1605		extended abstract
5888		keyboard	F	1372	-1372	1605	1869	1857794	0.0215025	1857891	0.0215054	2.49356E-07	97	157	2	BEFORE WORDS	PRODUCTION			1857794	1857891	1372	1605		extended abstract
5889		keyboard	CAPS LOC	1373	-1372	1606	1870	1857815	0.0215024	1857938	0.0215039	1.55093E-06	123	21	15	UNKNOWN	PRODUCTION			1857815	1857938	1373	1606		extended abstract
5890		keyboard	r	1373	-1373	1606	1870	1857949	0.0215049	1858055	0.0215053	1.51362E-06	106	134	2	BEFORE WORDS	PRODUCTION			1857949	1858055	1373	1606		extended abstract
5891		keyboard	a	1374	-1373	1607	1871	1858080	0.0215056	1858196	0.0215069	1.97917E-06	116	131	1	WITHIN WORDS	PRODUCTION			1858080	1858196	1374	1607		extended abstract
5892		keyboard	c	1375	-1374	1608	1872	1858251	0.0215075	1858394	0.0215092	1.22685E-06	143	171	1	WITHIN WORDS	PRODUCTION			1858251	1858394	1375	1608		extended abstract
5893		keyboard	t	1376	-1375	1609	1873	1858357	0.0215087	1858436	0.0215097	1.95602E-06	79	106	1	WITHIN WORDS	PRODUCTION			1858357	1858436	1376	1609		extended abstract
5894		keyboard	i	1377	-1376	1610	1874	1858526	0.0215072	1858616	0.0215118	1.15741E-06	90	169	1	WITHIN WORDS	PRODUCTION			1858526	1858616	1377	1610		extended abstract
5895		keyboard	c	1378	-1377	1611	1875	1858626	0.0215188	1858732	0.0215131	6.56481E-07	106	100	1	WITHIN WORDS	PRODUCTION			1858626	1858732	1378	1611		extended abstract
5896		keyboard	u	1379	-1378	1612	1876	1858700	0.0215273	1858785	0.0215137	2.11806E-06	85	74	1	WITHIN WORDS	PRODUCTION			1858700	1858785	1379	1612		extended abstract
5897		keyboard	m	1380	-1379	1613	1877	1858883	0.0215145	1859004	0.0215163	1.27315E-06	121	183	1	WITHIN WORDS	PRODUCTION			1858883	1859004	1380	1613		extended abstract
5898		keyboard	SPACE	1381	-1380	1614	1878	1858993	0.0215161	1859110	0.0215175	7.06019E-06	117	110	3	AFTER WORDS	PRODUCTION			1858993	1859110	1381	1614		extended abstract
5899		keyboard	g	1382	-1381	1615	1879	1859603	0.0215318	1859710	0.0215244	7.98611E-07	107	610	2	BEFORE WORDS	PRODUCTION			1859603	1859710	1382	1615		extended abstract
5900		keyboard	r	1383	-1382	1616	1880	1859672	0.0215298	1859778	0.0215252	1.6088E-06	106	69	1	WITHIN WORDS	PRODUCTION			1859672	1859778	1383	1616		extended abstract
5901		keyboard	a	1384	-1383	1617	1881	1859811	0.0215259	1859944	0.0215271	1.22685E-06	133	139	1	WITHIN WORDS	PRODUCTION			1859811	1859944	1384	1617		extended abstract
5902		keyboard	d	1385	-1384	1618	1882	1859917	0.0215262	1860044	0.0215283	2.37269E-06	127	106	1	WITHIN WORDS	PRODUCTION			1859917	1860044	1385	1618		extended abstract
5903		keyboard	e	1386	-1385	1619	1883	1860122	0.0215319	1860228	0.0215304	6.71296E-07	106	205	1	WITHIN WORDS	PRODUCTION			1860122	1860228	1386	1619		extended abstract
5904		keyboard	SPACE	1387	-1386	1620	1884	1860180	0.0215286	1860286	0.0215311	0.000159155	106	58	3	AFTER WORDS	PRODUCTION			1860180	1860286	1387	1620		extended abstract
5905		mouse	Movemei	1387	-1387	1620	1884	1873931	0.0218902	1877706	0.0217327	5.54282E-05	3775	13751	16	MOUSE	PRODUCTION	54	17	1873931	1877706				extended abstract
5906		mouse	LEFT Clicl	1387	-1387	1620	1884	1878720	0.0217444	1878970	0.0217473	0	250	1014	16	MOUSE	PRODUCTION	54	17	1878720	1878970				extended abstract
5907		replacen	1368:138	1387	-1387	1620	1884	1878720	0.0217444	1879220	0.0217502	9.79635E-06	0	250	10	CHANGE	PRODUCTION							2	extended abstract
5909		mouse	LEFT Clicl	1387	-1387	1620	1884	1879048	0.0217824	1879204	0.02175	0	156	78	16	MOUSE	PRODUCTION	54	17	1879048	1879204				extended abstract
5910		replacen	1356:136	1387	-1387	1620	1884	1879048	0.0217824	1879360	0.0217519	2.70835E-06	0	156	10	CHANGE	PRODUCTION							2	extended abstract
5912		mouse	LEFT Clicl	1387	-1387	1620	1884	1879282	0.0217595	1879438	0.0217528	0	156	78	16	MOUSE	PRODUCTION	54	17	1879282	1879438				extended abstract

Figure 1: Excerpt of Inputlog's general analysis output of Natalia's first session.

Initially, to get a deeper sense of the participant's writing process and to better understand Inputlog's logs, we watched the screen recordings several times in parallel to the corresponding Inputlog logs in the Excel file. Then, the first author of this paper coded Natalia's writing activity in two hierarchical levels: actions and, most importantly for this study, regulation episodes, following an emergent and iterative coding process in seven steps. For each step, the coding was discussed by two researchers until consensus was reached about the identification and the definition of each category.

Step 1. Keystroke loggings were segmented into actions that the participant conducted during writing. Actions were segmented when a change in behaviour was identified in the keystroke logging (e.g., log changes from mouse to keyboard) and/or in the screen recording (e.g. while moving the mouse, opens a pdf file -action 1-, scrolls up and down -action 2-, and selects a paragraph of the text -action 3-). Additionally, actions were also segmented into two or more when there were pauses longer than 3 seconds within the same action (e.g. types -action 1-, pauses

for four seconds, types -action 2.). An action finished when the writer shifted to another action.

Step 2. Segmented actions were emergently coded based on Natalia's behaviour (e.g. source shifting, pasting text or looking for a document) combining the keystroke loggings and screen recordings data. Since the focus of the analysis is on the Regulation Episodes, in Appendix C we provide only a non-comprehensive list of actions for illustrative purposes.

Step 3. We grouped strings of actions addressing the same aim or challenge into *Regulation Episodes*, the actions that writers implement to solve a problem – a difficulty or challenge identified during the writing process (Iñesta & Castelló, 2012; Zanotto, Monereo, & Castelló, 2011). As it is challenge-oriented, actions are always taken in context and, as such, the analysis not only considers writers' behaviour, but also the content and context of the actions and the text (e.g., a regulation episode aimed at clarifying the aims of the study). Notably, the same type of action could serve for different types of episodes. Instances in which there was no sign of regulation were coded as *Production Episodes*.

Step 4. Production and Regulation episodes were coded for two dimensions: a) *type of episode*, that is, what the writer did and aimed at in each episode; and b) *section of the text* involved in each episode; that is, which part of their prospective RA they were writing about in each episode: *title, introduction, objectives, method, results and discussion*. In addition, the episodes involving two or more sections were coded as *global*, while those performed in the sources were coded as *source*. Frequency and duration of the types and sections of the episodes were calculated.

Step 5. All discursive data (initial questionnaire, writing logs and feedback discussion) were gathered in a participant file and coded to identify Natalia's reported problems and challenges in each instrument (e.g., questions 1 and 7 in the initial questionnaire regarding their feelings towards writing, or the questions about the problems faced during the session in the post-writing log).

Step 6. Researchers linked the challenges and problems to the episodes identified not only to account for the quantitative data but also and most importantly to fully understand their goals and perceptions.

Step 7. Finally, we analysed and compared both drafts to identify the changes introduced in the second writing session and assess whether, and to what extent, the issues raised in the feedback session had been addressed and solved. A qualitative assessment of the final version of each draft was conducted since texts, in terms of the content and the changes introduced, are essential in our analysis as they are intrinsic to the unit of analysis.

The entire analysis took the context of episodes into account; thus, all information available of the individual writer available from outside and inside the process was taken into account. We created charts and graphics to visualize the writing process in each session.

4. Results

Table 1 provides a description of all the types of Regulation Episodes that were identified in the analysis, as well as the Production Episodes.

Table 1. Types of episodes and definition

Types of Episodes	Description
1. PRODUCTION	Linear production, where no regulation processes can be observed.
2. REGULATION	Adjustments and changes in the activity (regulation) are observed.
2. 1 Sources	Looking for, searching within, or reading sources (either own texts, others' papers, etc.).
2. 2 Editing	Surface-level revisions, such as correcting grammar and spelling mistakes and adjusting format.
2. 2 Revision	Deep-level revisions that affect the meaning of a word, sentence or paragraph. They involve substituting and adding information.
2. 4 Reading	Reading the text written so far*.
2. 5 Deleting	Eliminating parts of the text (words, sentences, paragraphs) or undoing, without any addition of information. When writers delete and add new information, the episodes fall into the 'revision' category.
2. 6 Recursive reformulations of the intended text (RRIT)	Revising the text while it is being produced, that is, immediate and recursive revisions at the sentence being written/point of inscription.

**Note*: episodes were coded as reading using the screen recordings under two circumstances: when the mouse pointer moved along the text lines in a linear way, and when the writer spent some time slowly scrolling down the text.

A first result has to do with the emergence of a particular type of writing episode, which we called a Production Episode (PE). Such episodes are aimed at linear text production, without any evidence of regulation processes (see Table 1). By contrast, we identified six different types of *Regulation Episodes* (RE). The first five of these included aims such as looking for *sources, editing, revising, reading* the written text or *deleting* some of its parts. It is worth highlighting, however, the emergence of the final type of episode, the Recursive Reformulation of the Intended Text (RRIT). This episode is somewhere in between production and revision episodes and is generally aimed at producing text, although the writing process is not as fluent as in production episodes. Rather, the writer revises at the word or sentence level shortly after writing them, similar to what Lindgren and Sullivan (2006) called revisions at the point of transcription or point of utterance. The difference is that

these authors emphasize the revision behaviour, while the RRIT episode includes both the production and the immediate and recursive revisions of such production. This type of episode is evidence of the complexity of research writing and can only be identified using synchronous instruments, such as keystroke logging and screen captures.

In the following three subsections, we report the results of the analysis of Natalia's writing process. We provide methodological notes and comments to highlight the utility and contribution of the methods used. First, we describe in detail her writing process in the first writing session, the episodes identified in the analysis, the challenges she mentioned and the relationship between challenges and episodes. Detail of the feedback she received on her first draft is then presented to facilitate the understanding of the process she followed in the second session, which is presented afterwards following a similar order.

4.1 First session

Overall writing session development

Globally, a big part of the session was devoted to alternating *sources* and *production* episodes. *Production* episodes were the longest type of episode and accounted for 50% of the time (see Table 2).

Sources episodes were the most frequent and prevalent among the regulation episodes, which resonates with the challenge Natalia mentioned about adapting previous texts to the idea she had of this new article. Episodes of *revision* were quite frequent, but short, whereas *editing* episodes were rare. The other types of episodes (*RRIT*, *reading* and *deleting*) accounted for less than 2.3% of the session.

In relation to the text sections development, Natalia progressed quite linearly through the various sections of a prototypical research article, with a few instances of recursivity, identified thanks to the combination of the screen recording and keystroke loggings.

Figure 2 shows the distribution of Natalia's episodes in the first writing session. The X axes represent time, whereas the Y axes represent the type of episodes (the top half of the figure) and the sections of the text (the lower half of the figure). Together, the two parts show what Natalia's focus was at any given moment of the writing session: what she did, and when she did it in the process, what preceded and what followed. This provides crucial context to interpret episodes and Natalia's struggles with certain parts of the text. The same information is depicted in Figure 3, which specifically highlights recursivity in Natalia's writing process.

Table 2. Frequency and duration of episodes by type and section of the text in Natalia's first session

		N	%	Duration (h:mm:ss)	%	Mean duration
Types of episodes	PRODUCTION	35	38.0	50:21	50.1	1:26
	REGULATION	57	62.0	50:09	49.9	0:53
	Source	29	31.5	32:20	32.2	1:07
	Editing	6	6.5	06:34	6.5	1:06
	Revision	11	12.0	05:22	5.3	0:29
	Reading	2	2.2	01:37	1.6	0:49
	Deleting	7	7.6	02:05	2.1	0:18
	RRIT*	2	2.2	02:11	2.2	1:06
	TOTAL	92	100	1:40:30	100	1:06
Sections of the text	Title	4	4.3	01:42	1.7	0:25
	Introduction	5	5.4	05:57	5.9	1:11
	Objectives	24	26.1	14:55	14.8	0:37
	Method	31	33.7	41:29	41.3	1:20
	Results	9	9.8	09:40	9.6	1:04
	Discussion	10	10.9	08:22	8.3	0:50
	Text	3	3.3	06:06	6.1	2:02
	Sources	6	6.5	12:19	12.3	2:03
	TOTAL	92	100	1:40:30	100	01:06

* RRIT = Recursive Reformulation of the Intended Text.

Challenges

During this first session, Natalia reported four challenges in the writing log and the initial questionnaire: (1) finding the way to write in her own words, (2) using and adapting previous texts of her thesis to meet the idea she had for the current article, (3) synthesising all the information she had about the study and her thesis, and (4) writing fluently and correctly in English. Adapting previous texts for the present article (challenge 1) was not only a challenge and but was also the objective Natalia reported in the writing log for the first writing session (see Appendix B).

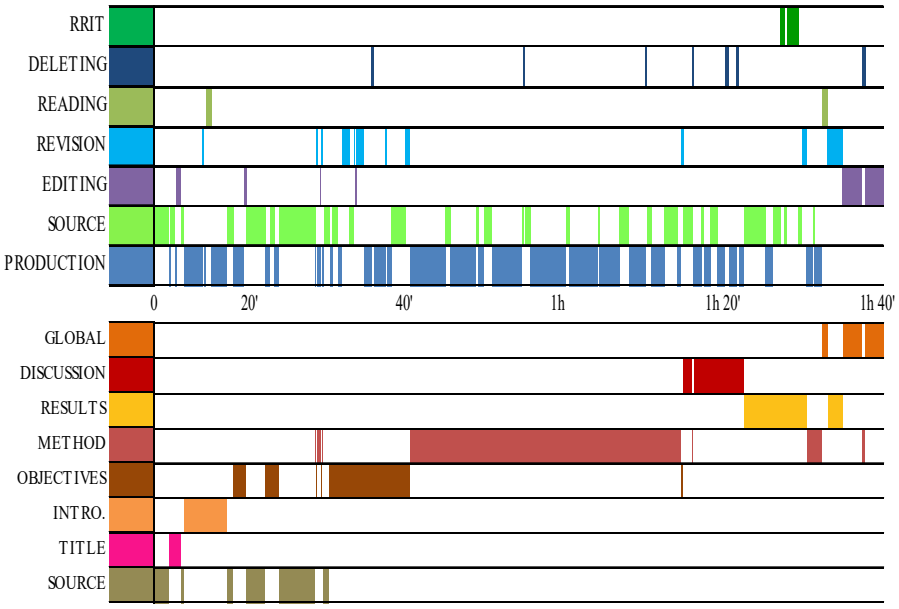

Figure 2: Natalia's episodes during the first writing session. * RRIT = Recursive Reformulation of the Intended Text.

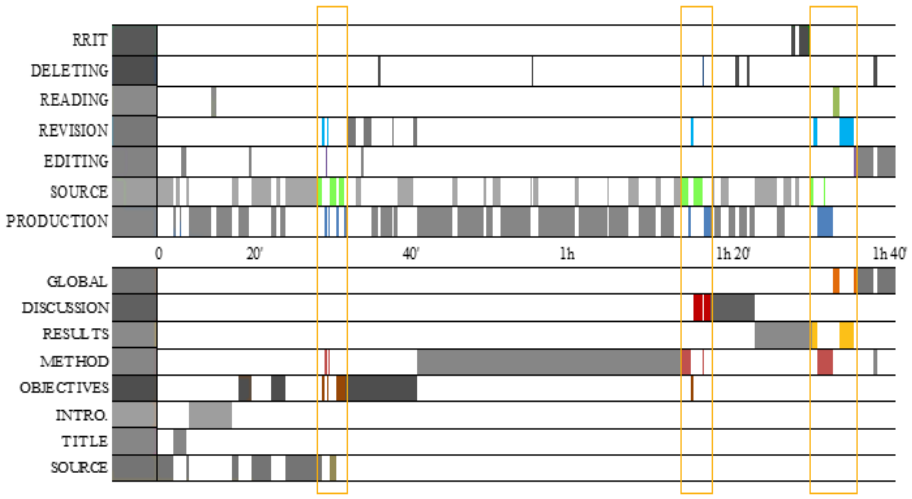

Figure 3: Instances of higher recursivity during Natalia's first writing session. * RRIT = Recursive Reformulation of the Intended Text.

Writing processes and text development: Relationship between episodes and challenges

Only regulation episodes were linked to the challenges, as displayed in Table 3. Even though production episodes could also be related to specific challenges, this relationship was impossible to identify as there are no signs of struggle within these episodes. In the description of the writing process, the names of the types of episodes are written in italics and the sections are underlined to make the explanation easier to follow. In brackets, we mention the number of the challenge listed in Table 3.

Table 3. Natalia's challenges in the first writing session and her attempts to solve them

Challenges	Source	Regulation episodes aimed at solving challenges
Writing in her own words.	Questionnaire	Recursive Reformulation of Intended Text (RRIT) Creating new text.
Adapting texts of her own to the concept she had of the article.	Log 1	Deleting / Sources Excluding (omitting or deleting after inserting) information from the sources. Revision Revising the text written so far.
Synthesising the information.	Log 1	Deleting / Sources Excluding (omitting or deleting after inserting) information from the sources.
Writing in English*.	Log 1	Editing Correcting spelling mistakes

*Although all her writing was in English, Natalia described this challenge in terms of writing fluency and, especially, vocabulary and grammar.

Natalia devoted 1 hour and 42 minutes to writing a first draft, in which we identified 92 episodes. As displayed in Figure 2, Natalia started with a blank page and first wrote the title, introduction, and objectives, using a combination of *production* and *sources* episodes, with minor *editing* and *revision* episodes in between. She then moved on to the method, but when she started *producing* text - mainly copying what she had in the sources (as seen thanks to the screen recording) - she had to go back to the objectives a few times to revise them, and then stayed in the objectives, *producing*, *deleting*, and especially *revising* the objectives she had previously written. This part of the session seemed to be aimed at producing text while *adapting it to the new article* (challenge 2 in Table 3). This seemed particularly

challenging in writing the objectives, because for this section she repeatedly consulted the *sources* and chose which information to include and which to leave out of the new text. Although she initially copied the ones she had in the sources, she had to revise them afterwards to adapt them to the new text, in one of the few instances of recursivity observed during the first session (depicted in Figure 3). In this part of the session, the struggle between the new and former text and objectives that Natalia described in the writing log was evident in her writing process.

Then, around minute 37 and until 1h 25min (see Figure 2), Natalia composed the method section of her paper by alternating *production* with *sources* episodes. The screen recording showed that she mostly selected and transcribed parts of the methods she had in the other texts, with minimal adaptations or changes. The only exceptions were four *deleting* episodes (two when writing the method). Regulation episodes (both *sources* and *deleting*) involved deciding what information to include and exclude in the new paper and synthesising the information (challenge 3), as well as adapting text to the new article (challenge 2). In this part of the session, Natalia's intentions (especially regarding the 'adaptation' challenge) were not so evident when observing her process as there were fewer revision episodes. However, her description in the writing log helped us interpret her process.

She then moved to the discussion, skipping the results, again alternating *production* with *sources* episodes, with a short instance of recursivity among the discussion, the method, and the objectives (see Figure 3) to briefly *revise* and *delete* information in these sections. Again, these episodes were aimed at adapting the text to the new article (challenge 2).

At the end of the session, she focused on the results section, which she had skipped earlier. This part of the session looks quite different. Instead of production episodes, Natalia performed *recursive reformulation of the intended text* (RRIT) episodes, the only two of this type in the whole session. Unlike for the other sections, the only source Natalia had available to write the results was a diagram of the main results of the study. She thus had to find her own words to write this section instead of relying so much on the sources, which she said in the initial questionnaire was a challenge for her (challenge 1). The combination of the screen recording and the initial questionnaire were crucial to interpret this part of the session, as we were able to observe the content of the sources and understand what the underlying challenge was.

After the results, she briefly went back to the method section to add information about the analysis related to the results she had just written, and then *revised* the constructs and measures of the study in the results section. This was the third and last instance of recursivity observed during Natalia's first revision. After this *revision*, she screened the text from top to bottom, making some minor *editing* for typos and spelling errors (challenge 4 about writing in English), and one significant change in the method section, where she *deleted* a sentence about data analysis.

4.2 Feedback

The week after, in the feedback session, workshop participants discussed Natalia's text. Table 4 shows the issues raised and the connection with the challenges Natalia mentioned in the first writing session. Although participants raised many issues, most were connected to the challenges Natalia mentioned in her writing log and questionnaire. They raised concerns about the unclear role of the theory used in the study, and whether it was important or used in the analysis (challenge 7). Natalia explained that the theory was central to her thesis, but not so much to this study and she was struggling to distinguish this paper from other texts in her thesis (challenge 2). This challenge was connected to the need to clarify the focus of the paper and reduce the number of objectives (challenge 6). Natalia's peers also suggested making her authorial voice stronger (challenge 5), by writing in her own words (challenge 1) and also reducing the method section, especially the description of the instrument (challenge 8), which, in turn, was connected to the challenge of synthesising the information that Natalia mentioned in the writing logs (challenge 3). She admitted being aware that this section was too long, but said she focused on that because she had a clearer picture of the method than the rest of the paper. Since the aim of the study was not clear yet, neither was the gap, which was missing from the text (challenge 10). Natalia's peers also asked her to provide a clear justification of the context and participants of the study (challenge 11) and more information about the analysis conducted and the software used (challenge 13).

Finally, participants also raised a few comments regarding formal aspects: the need to revise the spelling and grammar (challenge 9), which was connected to Natalia's challenge of writing in English (challenge 4) and the lack of citations to support the statements (challenge 12), especially in the introduction.

Table 4. Relationship between challenges mentioned by Natalia and the feedback comments

Challenges	Feedback comments
(1) Writing in her own words.	Making author's voice stronger in the text, especially by reducing method section.
(2) Adapting texts of her own to the concept she had of the article.	Too many objectives, lack of clear focus. Clarifying the role theory plays in the study.
(3) Synthesising the information.	Reducing the description of the instrument.
(4) Writing in English.	Correcting spelling and grammar mistakes.
No mention	Lack of gap in the literature.
No mention	Justifying the selection of the context and participants.
No mention	Lack of necessary citations.
No mention	Providing more information about the analysis.

4.3 Second session

Overall writing session development

In the second writing session, Natalia aimed at revising the text to apply modifications and improve text coherence and cohesion. The session lasted one hour and one minute. It contained 47 episodes and presents a very different distribution of episodes types than the first one (see Figures 4 and 5). The writing process was still quite linear with respect to the sections, although Natalia paid far more attention to the global level. She proceeded by solving the feedback comments one by one. Globally, *Regulation episodes* were more prevalent in this second session than in the previous one (see Table 5), especially those of *reading*, *editing* and *revision*.

Table 5. Frequency and duration of episodes by type and section of the text in Natalia's second session.

		N	%	Duration (h:mm:ss)	%	Mean duration
Types of episodes	PRODUCTION	5	10.6	08:28	14.9	1:42
	REGULATION	42	89.4	48:17	85.1	1:09
	Source	3	6.4	05:27	9.6	1:49
	Editing	3	6.4	11:49	20.8	3:56
	Revision	15	31.9	10:36	18.7	0:42
	Reading	8	17.0	12:06	21.3	1:31
	Deleting	11	23.4	03:36	6.3	0:20
	RRIT*	2	4.3	04:43	8.3	2:22
	TOTAL	47	100	56:45	100	1:12
Sections of the text	Title	2	4.3	00:21	0.6	0:11
	Introduction	9	19.1	09:15	16.3	1:02
	Objectives	7	14.9	04:46	8.4	0:41
	Method	17	36.2	10:32	18.6	0:37
	Results	2	4.3	00:45	1.3	0:23
	Discussion	3	6.4	02:36	4.6	0:52
	Text	4	8.5	23:03	40.6	5:46
	Sources	3	6.4	05:27	9.6	1:49
	TOTAL	47	100	56:45	100	1:12

*RRIT = Recursive Reformulation of the Intended Text.

Natalia again spent most of her time on the method section, but episodes in this section were now devoted to reading (n = 6), revising (n = 4) and deleting (n = 4) information instead of sources and production.

Challenges

For this second session, Natalia reported again that synthesising the information and writing fluently in English were major challenges, and that she aimed at addressing the comments provided by her peers and workshop facilitators (the complete list of challenges is depicted in Table 6).

Writing processes and text development: Relationship between episodes and challenges.

As depicted in Figure 4, at the beginning of the session, like in the first one, Natalia spent some time in the sources and then started switching between the introduction and the global level mostly to *revise* her paper, especially focusing on the description and role of the theory (challenge 2 and 7) and the lack of citations (challenge 12). She then *revised* the text to delete synonyms (challenge 4). During this part of the session, we observed two instances of recursivity among the title, introduction and global level (see Figure 5). These instances were brief and involved different types of episodes. Around minute 20, Natalia moved to the objectives and the method. *Revision* and *deleting* episodes were alternated frequently, first to reduce the number and scope of the objectives (challenge 2 and 6) and later to synthesise the information about the instrument (challenge 3 and 8). She then performed two episodes of *RRIT*, aimed at adapting the method to this new article (challenge 2) and including information about the data analysis (challenge 13). After that, she spent a couple of minutes *producing* text to provide some information about the theory in the discussion (challenge 2 and 7). During the last 20 minutes, Natalia worked at the global level again, as she went through the text from start to end looking for and correcting spelling and grammar errors (challenge 4 and 9), and then *reading* the whole text again without changing anything.

Natalia's draft was 940 words long, 122 words shorter than the first draft. Text analysis showed that she extended and revised the introduction to provide information and citations about the role of the theory, justify the study and define the gap of the study. These revisions improved the first part of the text, where she made some general claims about the topic and the theory and the (reduced and revised) objectives, which clarified the focus of the paper. The synthesis and revision of the instrument and procedure provided clarity to the method and made Natalia's authorial voice stronger. Voice was also introduced using the pronoun 'we'[5].

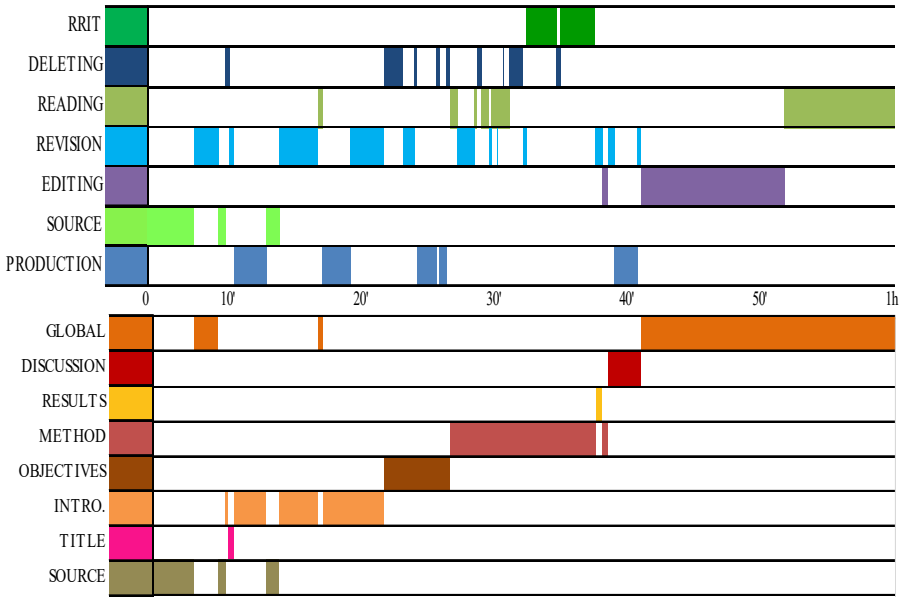

Figure 4: Natalia's episodes during the second writing session. * RRIT = Recursive Reformulation of the Intended Text.

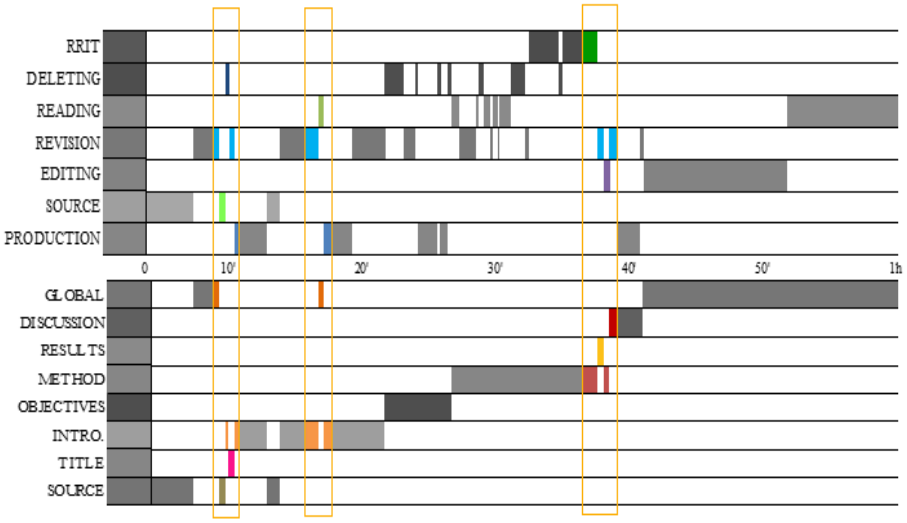

Figure 5: Instances of higher recursivity during Natalia's second writing session. * RRIT = Recursive Reformulation of the Intended Text.

That said, the text analysis revealed that Natalia's voice was still missing with regard to the dialogue with other authors, both in the introduction and the discussion, with the latter only reflecting about the contribution of the study and the theory. Despite her efforts, the role of the theory was still unclear due to lack of connection with the method and results.

Overall, during this session, each episode addressed one issue raised in the feedback session, especially those related to the challenges she mentioned in the first session. New issues raised by Natalia's peers, such as the lack of a clear research gap and citations received less attention, and very few episodes aimed at further developing or revising other aspects of the text. As we can see in Table 6, the variety of attempted solutions and the type of episodes linked to each problem were higher than in the first session, especially in relation to the adaptation of the text (challenge 2). Moreover, in this session *revision* and *deleting* episodes were aimed at solving not just one but various challenges.

5. Discussion

In this study, we explored the writing regulation processes of Natalia using an innovative and ecological methodology. The methodology combined synchronous (screen recordings and keystroke loggings) and asynchronous instruments (writing logs, questionnaire and texts) to collect data about Natalia's writing process and products, and about her actions and perceptions. All data were analysed together and in context. We argue that this approach provides a much deeper understanding of the complex writing processes involved in research and doctoral writing (Castelló, et al., 2013; Chang & Schleppegrell, 2016; Cotterall, 2013) and helps us unpick why, when, and how writing challenges are addressed and, in some cases, solved.

The analysis provided evidence of regulation taking place at all the textual levels and throughout the writing session (Ferrari, et al., 1998; Hadwin & Oshige, 2011; van den Bergh, et al., 2016) and we were able to connect these processes to the writer's aims and perceived challenges. In the first writing session, the triangulation of the observation of Natalia's writing processes and her perceptions allowed us to understand that the identified pattern of production-source-production was aimed at producing a rough first draft of the text based on the sources. The analysis showed that she regulated her writing process to adapt the text, especially by selecting information and modifying the objectives of her paper. Her activity in this first session was frequently aimed at knowledge telling (Baaijen & Galbraith, 2018; Bereiter & Scardamalia, 1987; Lonka, et al., 2014), with fewer instances aimed at knowledge transforming. Moreover, the analysis of the moment-by-moment creation of the text (done by combining screen recording and keystroke loggings) allowed us to identify several instances of recursivity, in which Natalia diverted from the linear production of her text, and to explore whether they were merely a shift

Table 6. Natalia's challenges in the second writing session and her attempts to solve them

Challenges	Source	Regulation episodes aimed at solving the challenges
Writing in her own words.	Questionnaire & Feedback	Deleting Reducing description of the instrument. Revision Including first-person pronouns. Adding connectors to organize paragraphs.
Adapting texts of her own to the concept she had of the article.	Log 1 & Feedback	Deleting Reducing the number of objectives. Revision Revising some of the existing objectives. Adding the name of the theory in the title. Recursive Reformulation of Intended Text (RRIT) Rewriting the measures used in the study. Production Adding information about the theory (contribution and relevant authors).
Synthesising the information.	Log 1 & 2 & Feedback	Deleting Deleting information about the instruments and the data collection procedure.
Writing in English.	Log 1 & 2 & Feedback	Editing Proofreading the text. Revision Revising the use of synonyms.
(10) Lack of clear gap.	Feedback	Production Adding a sentence about the lack of literature on the topic.
(11) Justifying the selection of the context and participants.	Feedback	Revision Moving the implications to the introduction to serving as a justification of the context. Adding information about the reasons for the selection.
(12) Lack of necessary citations.	Feedback	Revision Adding citations in the introduction.
Providing more information about the analysis.	Feedback	Recursive Reformulation of Intended Text Adding information about the analysis.

of focus or were episodes related to a more global representation of the text (e.g., writing the results triggered the need to add information in the method). The fact

that these instances mostly involved the objectives is also proof that Natalia's biggest challenge in this session was adapting previous texts to the new article.

The analysis also shed light on different patterns and processes between writing sessions. Data about the feedback together with Natalia's processes revealed a shift in her aims, which in turn implied changes in her regulation processes. Results showed that, in the second session, Natalia focused on addressing the feedback comments, rather than producing new text, as very few episodes were not directly related to the feedback. Thus, her writing regulation processes were aimed at specific objectives (e.g. feedback comments), in contrast to the ones she mentioned in the first session (e.g. adapting the new text). As a result, her regulation episodes were more easily traced and connected to the challenges she was addressing.

This effort to connect Natalia's writing behaviour to its focus revealed that in the second session, she regulated more and in a more flexible way compared to the first session. In contrast to the first session, in the second writing session the same type of episode (e.g. deleting) served to solve different challenges, and different types of episodes were aimed at solving one challenge. Interestingly, however, in this session we did not identify more instances of recursivity than in the first session, a characteristic that has been widely associated with regulation (Castelló, et al., 2013; Lonka, et al., 2014). These results indicate that recursivity might not exclusively be seen as changes among different parts of the text, but also as shifts among a diversity of writing regulation processes, such as production, revision and reading.

In sum, the design of this study focused on contrasting and complementing the observation of the student's processes and activities with the outcomes and, more importantly, with her perceptions about the challenges she faced. Accessing the writer's personal goals, rather than only the general goals connected to the characteristics and instructions of the task, as well as her anticipated and experienced challenges, was crucial to understanding her writing processes. This information not only helped us interpret the keystroke loggings and overcome the most common limitations of this method (Baaijen & Galbraith, 2018; Miller, Lindgren, & Sullivan, 2008); most importantly, it guided the interpretation of the episodes identified and the connections between the episodes, as well as the understanding of the process as a whole and the writer's strengths and weaknesses.

We were able to analyse Natalia's processes not by looking at quantitative measures, such as fluency or pause durations, but by exploring and understanding the regulation processes in the context of the challenges that she faced. Moreover, using the Regulation Episode (Iñesta & Castelló, 2012) as the unit of analysis allowed us to understand the evolution of the processes in relation to specific challenges and their attempted solutions, and to understand the relationships between episodes and their distribution through the session. This is particularly important as regulation is mainly a problem-solving process (Flower & Hayes, 1981; Hadwin & Oshige, 2011). In this sense, the method was useful to understand writing processes

oriented by broad objectives, such as those formulated by Natalia in the first session, as well as by the specific objectives guiding Natalia's second writing session.

However, while we always triangulated the observations with the participant's perceptions, it must be acknowledged that the relationship between challenges and processes was established by the researchers. Future studies could explore these issues and use the methodology presented here, in addition to stimulated recall interviews or retrospective think-aloud protocols, to gather more information about this relationship and participants' awareness of the decisions they made during the writing process.

Our study is a first attempt to explore doctoral students' writing regulation processes at a micro- and macro- level in ecological and authentic settings using synchronous and asynchronous data collection instruments. These methods and the analysis conducted, although laborious, showed how these processes can vary between and within writing sessions and could also be used to explore common patterns and variation among individuals in future studies. The method is designed to capture the unique and situated writing regulation processes. However, this in-depth analysis of individual variation could also guide the identification of common patterns across writers and texts. The evidence of complex relationships between regulation processes, the evolution of drafts, feedback, and the writer's own challenges and aims can contribute to our understanding of complex writing situations, such as research writing, in ecological conditions.

Aside from theoretical and methodological implications, this study also may have pedagogical implications. The analysis revealed that feedback prompted changes in the writer's regulation processes, promoted a problem-solving approach and more flexible and frequent uses of regulation processes. This suggests that feedback can trigger changes not only in the text, as proposed by previous studies (Aitchison, et al., 2012; Castelló, et al., 2013), but also in writers' regulation processes (Cotos, et al., 2020; Negretti & Mežec, 2019). Writing instructors and doctoral supervisors might want to consider providing feedback at early stages of the writing process to stimulate students to better regulate their writing processes and improve their texts.

Notes

1. For a more detailed description of the workshop, see Castelló et al. (2013).
2. Participant is referred to by a pseudonym.
3. The information the participant provided in this instrument is presented here as it allows as to provide a much richer description of Natalia in regards to their approach to the writing task. The questionnaire was also used to identify challenges faced when writing research texts.
4. To facilitate the dynamics of the session, we asked the participant to respond the questionnaire after she finished to write their text, so the instructions of

the writing task and the use of the data collection tools would be done for all workshop participants at the same time right before starting to write. In this way participants could devote as much time as they needed to write the text and responding to the questionnaire without interruptions or delay in the instructions.

5. It must be noticed that, although she was the sole author of the extended abstracts, she was planning to write the RA the extended abstract referred to with her doctoral supervisor, hence the plural pronoun.

Acknowledgements

This work was supported by the Spanish Ministry of Education, Culture and Sports [grant number FPU13/06957]; and the European Erasmus+ Programme under the project 'Researcher Identity Development: Strengthening Science in Society Strategies' [grant number 2017-1-ES01-KA203-038303].

References

Aitchison, C., Catterall, J., Ross, P., & Burgin, S. (2012). 'Tough love and tears': learning doctoral writing in the sciences. *Higher Education Research & Development, 31*, 435-447. http://doi.org/10.1080/07294360.2011.559195

Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction, 36*(3), 199-223. https://doi.org/10.1080/07370008.2018.1456431

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication, 29*(3), 246-277. http://doi.org/10.1177/0741088312451108

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Caffarella, R. S., & Barnett, B. G. (2000). Teaching doctoral students to become scholarly writers: The importance of giving and receiving critiques. *Studies in Higher Education, 25*(1), 39–52. http://doi.org/10.1080/030750700116000

Castelló, M., & Iñesta, A. (2012). Texts as artifacts-in activity: Developing authorial identity and academic voice in writing academic research papers. In M. Castelló & C. Donahue (eds.), *University writing: Selves and texts in academic societies* (pp. 179-200). Bingley, UK: Emerald Group Publishing Limited.

Castelló, M., Iñesta, A., & Corcelles, M. (2013). Learning to Write a Research Article: Ph. D. Students' Transitions toward Disciplinary Writing Regulation. *Research in the Teaching of English*, 442-477.

Chang, P., & Schleppegrell, M. (2016). Explicit Learning of Authorial Stance-taking by L2 Doctoral Students. *Journal of Writing Research, 8*(1), 49-80. https://doi.org/10.17239/jowr-2016.08.01.02

Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the Research Writing Tutor during revision. *Journal of Writing Research, 12*(1), 187-232. https://doi.org/10.17239/jowr-2020.12.01.07

Cotterall, S. (2013). More than just a brain: emotions and the doctoral experience. *Higher Education Research & Development, 32*, 174–187. http://dx.doi.org/10.1080/07294360.2012.680017

Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and keylogging data. *Journal of Writing Research, 5*(1), 133-158. http://dx.doi.org/10.17239/jowr-2013.05.01.6

Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional science, 26*(6), 473-488. http://doi.org/10.1023/A:1003202412203

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication, 32*(4), 365-387. https://doi.org/10.2307/356600

Franklin, S. V., & Hermsen, L. M. (2014). Real-time capture of student reasoning while writing. *Physical Review Special Topics - Physics Education Research, 10*(2), 20121. http://doi.org/10.1103/PhysRevSTPER.10.020121

Gnach, A., Wiesner, E., Bertschi-Kaufmann, A., & Perrin, D. (2007). Children's writing processes when using computers: Insights based on combining analyses of product and process. *Research in Comparative and International Education, 2*(1), 13-28. http://doi.org/10.2304/rcie.2007.2.1.13

González-Ocampo, G., & Castelló, M. (2018). Writing in doctoral programs: examining supervisors' perspectives. *Higher Education, 76*(3), 387-401. https://doi.org/10.1007/s10734-017-0214-1

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240-264.

Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, *59*, 116-125. https://doi.org/10.1016/j.system.2016.05.002

Iñesta, A., & Castelló, M. (2012). Towards an integrative unit of analysis: Regulation Episodes in expert research article writing. In C. Bazerman, C. Dean, J. Early, K. Lunsford, P. Null, S.Rogers, & A. Stansell (Eds.), *International Advances in Writing Research: Cultures, Places, Measures* (pp. 421–448). Fort Collins, CO: WAC Clearinghouse.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358-392. https://doi.org/10.1177/0741088313491692

Lindgren, E., Knospe, Y., & Sullivan, K. P. (2019). Researching Writing with Observational Logging Tools from 2006 to the Present. In Lindgren, E., & Sullivan, K. P. H. (Eds.). *Observing writing: Insights from keystroke logging and handwriting* (pp. 1-29). Leiden, Netherlands: Brill Publishing.

Lindgren, E., & Sullivan, K. P. H. (2006). Analysing on-line revision. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging: Methods and applications* (Vol. 18, pp. 157–188). Oxford: Elsevier.

Lindsay, S. (2015). What works for doctoral students in completing their thesis?. *Teaching in Higher Education, 20*(2), 183-196. https://doi.org/10.1080/13562517.2014.974025

Lonka, K., Chow, A., Keskinen, J., Hakkarainen, K., Sandström, N., & Pyhältö, K. (2014). How to measure PhD. Students' conceptions of academic writing - and are they related to well-being? *Journal of Writing Research*, *5*(3), 245–269. https://doi.org/10.17239/jowr-2014.05.03.1

MacArthur, C. A., & Graham, S. (2016). Writing research from a cognitive perspective. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 24-40). New York: Guilford Press.

Maher, D., Seaton, L., McMullen, C., Fitzgerald, T., Otsuji, E., & Lee, A. (2008). 'Becoming and being writers': the experiences of doctoral students in writing groups. *Studies in Continuing Education, 30*(3), 263-275. https://doi.org/10.1080/01580370802439870

McAlpine, L., & Mitra, M. (2015). Becoming a scientist: PhD workplaces and other sites of learning. *International Journal of Doctoral Studies, 10*, 111-128. https://doi.org/10.28945/2112

Miller, K. S., Lindgren, E., & Sullivan, K. P. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *Tesol Quarterly, 42*(3), 433-454. https://doi.org/10.1002/j.1545-7249.2008.tb00140.x

Miller, K. S. & Sullivan, K. P. H. (2006). Keystroke logging: an introduction. In G. Rijlaarsdam (Series Ed.) and K. P. H. Sullivan, & E. Lindgren. (Vol. Eds.), *Studies in Writing, Vol. 18, Computer Keystroke Logging: Methods and Applications* (pp. 1–9). Oxford: Elsevier.

Murray, R. & Newton, M. (2009). Writing retreat as structured intervention: margin or mainstream?, *Higher Education Research & Development, 28*, 541-553. https://doi.org/10.1080/07294360903154126

Negretti, R., & Mežek, Š. (2019). Participatory appropriation as a pathway to self-regulation in academic writing: The case of three BA essay writers in literature. *Journal of Writing Research, 11*(1), 1-40. https://doi.org/10.17239/jowr-2019.11.01.01

Odena, O., & Burgess, H. (2017). How doctoral students and graduates describe facilitating experiences and strategies for their thesis writing learning process: A qualitative approach. *Studies in Higher Education, 42*(3), 572-590. https://doi.org/10.1080/03075079.2015.1063598

Papen, U., & Thériault, V. (2018). Writing retreats as a milestone in the development of PhD students' sense of self as academic writers. *Studies in Continuing Education, 40*(2), 166-180. https://doi.org/10.1080/0158037X.2017.1396973

Paré, A. (2011). Speaking of writing: Supervisory feedback and the dissertation. In L. McAlpine & C. Amundsen (Eds.), *Doctoral education: Research-based strategies for doctoral students, supervisors and administrators* (pp. 59–74). Dordrecht: Springer.

Pedrazzini, A., Bautista, A., Scheuer, N., & Monereo, C. (2014). Review by (non) peers as an opportunity for learning: a case study on the editorial process of papers by junior researchers. *Infancia y Aprendizaje, 37*(4), 851-901. https://doi.org/10.1080/02103702.2014.977531

Sala-Bubaré, A., & Castelló, M. (2018). Writing regulation processes in higher education: a review of two decades of empirical research. *Reading and Writing, 31*(4), 757-777. https://doi.org/10.1007/s11145-017-9808-3

Simpson, Z. (2009). "Totally in the Zone": Using metaphor to 'glimpse' writer-identity, *Education as Change, 13*(1), 195-209. https://doi.org/10.1080/16823200902945028

Sullivan, K. P., & Lindgren, E. (2006). *Computer keystroke logging and writing: Methods and applications*. Oxford: Elsevier.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (Vol. 1). Ann Arbor, MI: University of Michigan Press.

TechSmith (n.d.). Snagit. Retrieved from http://www.snagit.com/

van den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text quality. *Educational Psychology*, *21*(4), 373-385. https://doi.org/10.1080/01443410120090777

van den Bergh, H., Rijlaarsdam, G., & van Steendam, E. (2016). Writing process theory: A functional dynamic approach. MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.). *Handbook of writing research* (2nd ed.). New York: Guilford Press.

Van Waes, L., Leijten, M., Mariën, P., & Engelborghs, S. (2017). Typing competencies in Alzheimer's disease: An exploration of copy tasks. *Computers in Human Behavior, 73*, 311-319. https://doi.org/10.1016/j.chb.2017.03.050

Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of pragmatics*, *35*(6), 829-853. https://doi.org/10.1016/S0378-2166(02)00121-2

Wengelin, Å., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior research methods, 41*(2), 337-351. https://doi.org/10.3758/BRM.41.2.337

Wisker, G. (2015). Developing doctoral authors: Engaging with theoretical perspectives through the literature review. *Innovations in Education and Teaching International, 52*(1), 64-74. https://doi.org/10.1080/14703297.2014.981841

Yin, R. K. (2003). *Case study research design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Zanotto, M., Monereo, C., & Castelló, M. (2011). Estrategias de lectura y producción de textos académicos: Leer para evaluar un texto científico. *Perfiles Educativos*, *33*(133), 10–29.

Appendix A: Writing articles: certainties and uncertainties

Think about the key aspects in the way you face research writing, the certainties you have and the uncertainties that disturb you. The following questions can help you think about how you write. We ask you to reflect on all the questions but feel free to add anything else you deem appropriate.

1. Do you like writing? Is it easy and satisfactory for you or, on the contrary, it is a problem that often you don't know how to solve? Does this also happen when you write articles? Is writing a research article different from writing other texts? Why?
2. How do you see yourself as a writer? Try to define yourself with a metaphor.
3. If you had to use a metaphor to define what writing a research article is, what would you say?
4. What are the characteristics of a research article? What type of text does it resemble? Why?
5. What would you say are the characteristics of a good research articles writer?
6. What you written any research article? How did it go?
7. How do you feel when writing an article? What are the most frequent feelings?
8. What role writing plays in your research?
9. Do you see yourself as a research author? Why?
10. What are your motivations to write this article?
11. What are your expectations regarding this workshop?

Appendix B: Writing log

Before starting:

- I hope the writing session will go/be…
- My objectives are…
- I will probably face problems in relation to…

After the session ended:

- The session went/was… (please explain briefly the process you followed)
- While I was writing, I felt… because…
- I face the following problems:
  1.
  2.
  3.
  …
- To solve them I had to… (or I didn't solve them)
  1.
  2.
  3.
  …
- I used the following resources (e.g. articles, websites, mail…)
  1.
  2.
  3.
  …
- And I used them to…
- Now I feel (*satisfied/unsatisfied/nervous/happy…*) with the result, because…

Appendix C: List of observable actions

- Changing text format
- Closing document
- Copying from the source
- Copying from the text (keyboard shortcuts or clicks)
- Deleting at the point of utterance
- Deleting away from the point of utterance
- Navigating computer folders
- Opening document
- Pasting at the point of utterance
- Pasting away from the point of utterance
- Pausing
- Reading the source
- Reading the text
- Replacing text at the point of utterance
- Replacing text away from the point of utterance
- Scrolling up and/or down the source
- Scrolling up and/or down the text
- Shifting between sources
- Typing at the point of utterance
- Typing away from the point of utterance
- Undoing
- Using web browser