Calibrating genre: Metacognitive Judgments and Rhetorical Effectiveness in Academic Writing by L2 Graduate Students

Raffaella Negretti

Chalmers University of Technology*

*Part of this work was produced at Stockholm University, Department of English, where Raffaella Negretti worked as tenure-track assistant professor until July 2015. Raffaella Negretti is now associate professor (docent) in academic and scientific writing in English at Chalmers University of Technology, Division for Language and Communication. Her research focuses on academic writing, metacognition, and language learning, and has appeared in the *Journal of Second Language Writing*, *Written Communication*, and the *Journal of Language Learning and Technology*.

Correspondence concerning this article should be addressed to:
Raffaella Negretti, Chalmers University of Technology, Department of Learning and Communication in Science, Division for language and Communication, SE-412 96 Gothenburg, Sweden.
Email: negretti@chalmers.se

Abstract

Several strands of applied linguistic research have emphasized the importance of genre awareness for academic writing students. Although metacognitive behaviors have been linked to L2 writing proficiency and performance, there is still the need for an account of how and why different metacognitive behaviors can help L2 academic writers to apply genre knowledge in authentic situations. Adopting an interdisciplinary approach, this study borrows the framework of calibration from educational psychology to highlight the relationship between the accuracy of graduate students' metacognitive judgments and the quality of their texts. Within an authentic setting, the nature of metacognitive judgments is calibrated against the assessment of rhetorical effectiveness by teacher raters using genre analysis criteria. Findings show that individual differences in rhetorical effectiveness can be better understood when accuracy of metacognitive judgments is considered along two qualitative dimensions: depth and alignment. Differential achievement relates to the ability to apply genre knowledge to the text, and misalignments in task perceptions and criteria. Implications for genre pedagogy and further research are discussed.

Keywords: metacognition, academic literacy, genre awareness, L2 writing, English for academic purposes, individual differences

Introduction

University students' success often depends on their ability to write academic texts such as course papers, project proposals, Masters and doctoral theses, frequently in English. Although research in different fields has investigated how students develop academic writing abilities, we still do not have a theory that 'bridges gaps among textual, cognitive and social dimensions for writing' (Hacker et al. 2009: 154). Recently, scholars have underscored that distinct disciplinary foci seem to converge on a theoretical characterization of (academic) writing expertise as multicomponential (Gentil 2011): whether in a first (L1), second (L2), or foreign language (FL), it entails the development of a multiplicity of cognitive and communicative abilities. Clearly, a renewed effort to cross disciplinary borders and find areas of intersection is necessary to achieve an authentic, comprehensive view of the learning dynamics involved in L1 and L2 writing (Ortega 2012). One such area of intersection may be the study of metacognition.

In applied linguistics, metacognition has been discussed in relation to the strategic behaviour adopted by language learners (see Gao 2007), underscoring its function in learners' ability to self-direct strategies and knowledge. Similarly, in educational psychology metacognitive awareness has been shown to facilitate self-regulation and adaptation of knowledge across domains and tasks (Schraw 1998). L2 writing research has characterised metacognitive knowledge as non-language specific knowledge about writing, helping learners monitor their strategies and fulfil rhetorical/contextual task demands (Schoonen et. al 2011). This characterization is echoed in the concept of genre awareness/knowledge: writers' ability to orchestrate and flexibly deploy rhetorical resources across contexts and languages (Johns 2008; Hyland 2007; Tardy 2009). Yet, it remains unclear how exactly metacognition helps learners of academic writing to achieve rhetorical effectiveness: there is a need to look closely at the *nature* of the metacognitive processes involved in learning to write (Schoonen et. al. 2009).

This study aims to investigate L2 graduate students' metacognitive judgments and their relationship to rhetorical effectiveness in academic writing, using the framework of *calibration*: 'how accurately people's metacognitive judgments match target performance' (Dunlosky and Thiede 2013: 58). The intention is to respond to Pieschl's (2009) call for an extended conceptualization of calibration to achieve ecological validity, by examining the correspondence between individual variation, internal criteria

for metacognitive judgment, and external criteria in naturalistic settings. Recent calibration research in this direction has shown that the *accuracy* of metacognitive judgments can be explained in the light of their qualitative *nature* (Dinsmore and Parkinson 2013): the type of criteria used by students to evaluate the quality of their performance. Considering the communicative nature of academic genres, I thus propose a qualitative 'twist' to the study of calibration: accuracy is investigated by comparing qualitative aspects of metacognitive judgments and rhetorical effectiveness judgments by expert genre teachers, using a genre-based matrix of criteria. These aims might seem ambitious, but in truth the research idea is rather simple, and rests on these general questions: What types of metacognitive judgments seem to help learners produce better texts? How coherent are metacognitive criteria with the contextual, rhetorical requirements of academic genres? I will return to the concept of calibration after a brief review of key theoretical concepts and relevant research on genre and the development of writing expertise.

Theoretical framework and previous research

The concept of metacognition (Flavell 1979) denotes humans' ability to strategically invoke, monitor and control knowledge in a goal-oriented fashion. Despite inconsistencies of conceptualization (Dinsmore et a. 2008), research in educational psychology agrees that metacognition plays an important role in learners' ability to use knowledge, self-regulate their learning, and perform successfully across contexts (Brown 1987, Veenman et. al. 2006), including writing (Zimmerman and Bandura 1994). Essentially, 'the study of metacognition is motivated by the assumption that if metacognition were accurate, people could take effective control of their own learning' (Metcalfe 2009: 159). However, learners' metacognitive judgments are not always accurate: several studies have shown little relation between learners' perceptions of their performance and their actual performance (e.g., Nietfeld et al. 2005; Pressley and Ghatala 1988; Winne and Jamieson-Noel 2002). Research on calibration has attempted to explain why. While accurate calibration determines learners' capacity to meet task demands (Winne and Hadwin 1998), inaccurate metacognitive judgments often lead to estimation bias (under- and overconfidence) and poor performance (Alexander 2013; Kruger and Dunning 1999). Pieschl (2009) has emphasized the need to extend the calibration paradigm to better understand the nature of metacognition in naturalistic learning. With authentic, complex tasks that can be solved in more than one way, the

relationship between metacognitive monitoring, self-regulation and performance is not always straightforward. In these cases, the specific context at hand needs to be taken into account, at least in terms of the learner's perceived task requirements. A research approach that takes into account qualitative dimensions of how learners internalize and negotiate contextual standards provides a point of departure for the study of metacognition in academic writing.

Applied linguistics research on genre has not explicitly addressed metacognition. Although studies of genre have differed in focus (see Hyon 1996), they converge on the view that developing academic writing expertise means learning to communicate via genres within a disciplinary community (Berkenkotter and Huckin 1995; Johns 2008; Tardy 2009). Several discourse-analytical studies have investigated how writers achieve this goal in their texts through a variety of linguistic and rhetorical features (e.g., Harwood 2005; Hyland and Tse 2004; Pecorari and Shaw 2012; Thompson 2001). In this tradition, learning academic writing, whether in L1 or L2, means developing the ability to modulate a variety of resources to meet academic genres' requirements as situated means of knowledge construction (Hyland 2012).

Metacognition has also been studied in relation to strategic aspects of language learning (e.g., Vandergrift 2005; Wenden 1987, 1998). Early cognitive research on writing viewed writing expertise as the ability to solve the 'rhetorical problem' (Flower and Hayes 1980), and discourse knowledge as fundamental in writers' ability to evaluate their text (Scardamalia and Paris 1985). Longitudinal studies on the development of expert writing across languages suggest important metacognitive and metalinguistic dimensions (Sasaki 2011; Kobayashi and Rinnert 2013). Research has shown that metacognitive knowledge helps transfer genre knowledge across tasks and contexts (Reiff and Bawarshi 2011) and that it mediates between L1 and FL/L2 writing proficiency (Victori 1999; Schoonen et al. 2011).

Research on cognitive and metacognitive processes in L2 writing has indicated a relationship between the temporal distribution and frequency of these processes and text quality (Breetvelt et al. 1994; Ong 2014; Ong and Zhang 2013; Manchón and Roca de Larios 2007; Tillema et al. 2011; van den Bergh and Rijlaarsdam 2001). This research has offered important answers about the type of online processes L2 writers engage in; however, it has typically entailed experimental conditions and the manipulation of variables such as planning time and task complexity. Thus, it is important to add

research that examines how 'metacognitive processes may be influenced by individual differences, types of language tasks, context of elicitation, and learners' awareness and knowledge of these processes' (Ong 2014: 28). For instance, a recent study (Khuder and Harwood 2015) comparing test (timed) and non-test (non-timed) situations, showed that independently from distribution and effectiveness of cognitive processes, students wrote significantly better in more naturalistic conditions. Thus, a limitation of the cognitive-oriented body of writing research, from a genre perspective, is that it has not addressed the writing of authentic, advanced academic tasks, in naturalistic situations. Pieschl (2009) emphasizes a lack of sufficient ecological validity in studies of metacognitive processes that do not capture students' calibration to *external* demands, since learners 'translate these perceptions into adequate internal standards for further self-monitoring and self-regulation' (Pieschl et al. 2012: 287). A qualitative approach is needed to illuminate this connection between learners' perceived task requirements, metacognitive processes and contextual demands.

The goal of this study is to understand the relationship between calibration and more or less felicitous rhetorical choices, and identify potential trajectories for further study of metacognition in academic writing. It builds on previous work (Negretti 2012) showing that different types of metacognitive awareness mediate between task perceptions, the evaluation criteria used by novice academic writers, and the self-regulatory strategies they use. That study highlighted students' tendency to self-regulate more effectively when they reached conditional metacognitive awareness, tied to each specific task and its rhetorical demands. Similarly, research involving L2 pre-service teachers suggested an interplay between the participants' metacognitive awareness of how to use genre knowledge and their ability to adapt their texts to specific tasks (Negretti and Kuteeva 2011). Taken together, these studies suggest that it is beneficial to investigate metacognition from an ecological perspective (van Lier 2010): taking into account the criteria and expectations of the context where learning takes place.

The main issue in adapting the calibration framework to writing research is measuring writing performance: text quality is by nature tied to task, posing problems of score generalizability (Schoonen et al., 2009). It has been measured as writing fluency (Larsen-Freeman 2006) or as a function of a composite score based on raters' assessment (e.g. Breetvelt et. al 1994). However, the notion of text quality needs to acknowledge the dialogicity of performance (van Lier 2010): quality according to

whom? Thus, it is here operationalized as rhetorical effectiveness, evaluated by expert teachers through genre analysis criteria (Swales 1990). This study aims to answer these questions:

a) What is the nature of the criteria upon which graduate academic writers base their metacognitive judgments?

b) What is the relationship between the nature of students' metacognitive judgments and the rhetorical effectiveness of their texts?

Method

Setting and participants

Data were collected from a group of graduate students in the humanities and social sciences, participating in an English for Academic Purposes course at a major Nordic university (see Table 1). All students but one had English as an L2. The language of the course was English, and all the data were collected in English by the researcher, who was also the teacher of the course. Eight out of twenty-three students consented to participate and provided complete data.

Table 1

Participants' background

	Name	Age	Education	Discipline
1	Bea	38	PhD (completing)	Education
2	Tim	26	MA 2nd term	Middle-East studies
3	Mat	26	MA 3rd term	Archeology
4	Lina	29	PhD	Conservation studies
5	Leo	27	MA	English linguistics (previous degree in Physics)
6	Viola	25	MA 2nd term	Middle-East studies
7	Sara	25	MA 2nd term	Performing arts
8	Amelia	52	PhD	Education

The ESP genre-based course aims to foster an awareness of academic genres in the students' disciplines. Over six weeks, students critically analyze genre models of their choice (a thesis and published articles), discuss disciplinary differences and similarities in collaborative analysis tasks, and use their insights to write an individual research proposal.

Data collection

In order to capture qualitative aspects of calibration in academic writing, metacognitive judgments were elicited through written retrospections: two self-evaluations assigned respectively after the first draft of the proposal (seminar 4) and before the final submission (seminar 6). These techniques require learners to make judgments and offer explanations for their choices (Chaudron 2003). The retrospections aimed at eliciting 'retrospective accuracy judgments' (see Schraw 2009: 416-417 for a taxonomy of calibration) about how well one performed on a task (and why) after the task is completed.

In mid-term retrospections, students wrote a brief self-evaluation of their first draft, determining which parts required more work and why. No criteria for evaluation were suggested, to avoid directing the students' attention to specific 'desirable' aspects of rhetorical effectiveness. Final retrospections explicitly elicited judgments on the text, and asked students to respond to these questions: 1) Which aspects of the text did you put most work into, and why? 2) Did you face any specific challenges? 3) Overall, what were the most important criteria for you in writing your proposal? Again, the aim was to trigger metacognitive reflection without suggesting specific criteria. Although retrospective accounts pose undeniable limitations in terms of memory accuracy, this study did not aim to map online cognitive writing processes, but rather to elicit the learners' accounts of approaching the writing task (Greene and Higgins 1994).

Analysis of metacognitive judgments

The analysis proceeded in several stages; systematicity and validity of data treatment were ensured through constant engagement with the data, verification of initial coding with a peer-debriefer, recurrent cross-comparison of coded data and focus on overall trends rather than selected, 'exotic' bits (Silverman 2001). Techniques of constructivist grounded theory were adopted (Charmaz 2006): codes derived from the data rather than a-priori categories, active coding using Nvivo, and use of detailed memos of how interpretation of the data was achieved to encourage reflexivity and validity.

Initially, open coding line-by-line through active verbs described the actions expressed in the data (Charmaz 2006). For example, a statement such as 'My article is based on a course I did last year' (Lina) was coded as '*Explaining* the background of the research project'. The second stage was focused coding, i.e. the inductive procedure of

deriving analysis categories from the data itself: codes created at stage one were reviewed to derive an initial categorization of metacognitive judgments. The above example (Lina) for instance does not convey evaluation, just a description of background, and thus was not coded as a metacognitive judgment. Students' comments often described writing strategies and self-regulation, and were coded as such in the data. Three main categories of criteria for metacognitive judgments were derived:

- Rhetorical criteria. Eg: 'The introduction … is a crucial part of the text and has to be comprehensible for people who are not familiar with the topic' (Mat)

- Content criteria. Eg: 'The literary review is currently too slim, largely due to the trouble of finding useful material' (Viola)

- Language criteria. Eg: 'Wordiness and informal language: phrases of everyday language just kept on sneaking their way into the text' (Lina)

The three categories often blended in the same comment, but if a student explicitly indicated a rhetorical purpose for content and/or language features, the comment was coded as 'Rhetorical criteria'. The comparison of coded data across students yielded an overview of the categorization of metacognitive judgments, including relations to the text and to self-regulation, and suggested the qualitative dimension of depth. Students' retrospections were then reviewed to map individual profiles using this categorization, which resulted in a table summarizing the content of the students' retrospections (see online supplementary material).

Finally, the content of each student's retrospections was compared with the comments by three raters about rhetorical effectiveness, suggesting the dimension of alignment. These two dimensions of accuracy—depth and alignment of metacognitive judgments—seemed connected to differences between the most and the least effective writers.

Analysis of text quality: matrix of rhetorical effectiveness

As this is a naturalistic study, it was important to assess text quality coherently with the genre-based pedagogy adopted in the course and the requirements of the final assignment (available online). This final assignment, a research proposal, aims to help students develop an argument consistent with published academic genres in their fields, yet is accessible to a broad academic audience: the learning objective is not to teach students to write a specialized text, but rather to 'sensitize students to rhetorical effects

and structures that tend to recur in genre-specific texts' (Swales 1990: 213). The assignment is modeled on the research article as the center of a web of academic genres, and thus a viable model for genre pedagogy (Swales 1990: 177). The connection between the course assignment and the research article was highlighted repeatedly in the seminars and the course tasks, especially for the doctoral students, who were essentially writing the same assignment as the MA students but could include a brief paragraph reporting results (i. e. a 'mini research article' rather than a proposal).

A matrix was created to assess text quality. The goal was to elicit the raters' perceptions of the textual characteristics that help or hinder rhetorical effectiveness. The criteria were thus based on the primary communicative action of 'creating a research space' (Swales 1990): the fact that the text 'leads to the conclusion that the research is relevant' (Feak and Swales 2011: 10). The criteria shown in Table 2 are derived from ESP genre analysis research[i]: each criterion was chosen to reflect the rhetorical function of a key feature of academic genres (see explanation in the end note) both at the macro level (i.e. overall organization and argumentation) and the micro level (i.e. use of reporting structures and stance markers). Furthermore, these criteria reflected closely the course requirements. The advantage of using criteria derived from genre analysis is that they are general enough to accommodate disciplinary variation, yet at the same time capture the common communicative goal underlying different rhetorical features of academic texts (Feak and Swales 2011: 56).

Table 2

Matrix of criteria for the analysis of rhetorical effectiveness

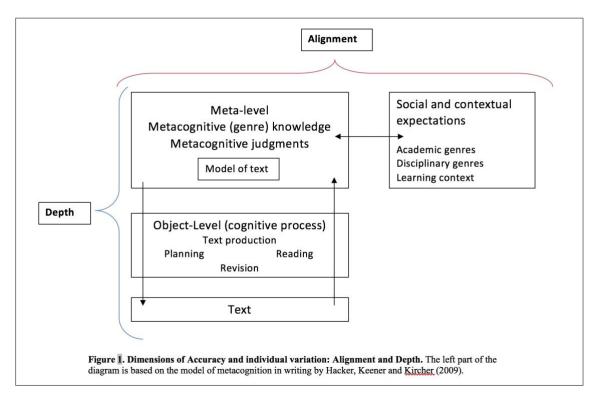| Macro level: Purpose and argumentation		
<u>Criteria</u>	<u>Check</u>	<u>Comments and examples</u>
a) The introduction and literature review tell a "research story" by creating a specific disciplinary niche for the study		
b) The aims and significance of the study are explained and supported through secondary sources and references to previous research in the area		
c) The methodology for the study is connected to its aim and supported through references to theory or previous research		
Micro Level: Discourse and style		
a) In the text, paragraphs are connected logically by indicating their link to the niche		
b) References to secondary sources (citations) are used to create a research context and indicate stance, not simply listed		
c) The style is consistent with the academic register and with disciplinary conventions		

Three raters, including the researcher, independently evaluated the final texts using this matrix. The raters were all applied linguists and experienced genre teachers, who had taught this course and therefore had extensive experience in evaluating this type of research proposal. The choice of raters was motivated by their role as stakeholders in this learning context. The aim of the matrix was to allow raters to respond to the texts with an open agenda; thus, it was used as a departure point to elicit evaluations in a de-briefing interview rather than as a rating protocol: this poses a further limitation in terms of reliability, while it brings to the surface the raters' spontaneous response to the texts. Raters independently assessed each text; they were instructed to check each criterion when pertinent and then write comments explaining their assessment. An in-depth de-briefing meeting (1 hour) was then held with each rater separately to explore what motivated their judgments. Raters agreed on 33 out of 48 possible evaluation instances; the de-briefing interviews elicited very similar overall views, even when occasional matrix ratings diverged, and prompted insightful information about how each rater conceptualized rhetorical effectiveness.

Findings

In this section, I will first present the two qualitative dimensions of metacognitive judgments' accuracy that were identified in this study. Then, the first sub-section presents an overview of the coding categorization illustrating the qualitative nature of the students' metacognitive judgments. It should be noted that this categorization does not report in detail the content of the retrospections, which is presented in a supplementary table online. In the second sub-section, the content of each student's retrospection is summarized and compared with the raters' evaluations, with supporting quotes and comments from the students and the raters, from the most to the least effective writers.

Accuracy: depth and alignment of metacognitive judgments

Two qualitative dimensions of accuracy emerge from the data that seem related to rhetorical effectiveness: *metacognitive depth* and *alignment* with external requirements (see Figure 1 below).

Figure 1. Dimensions of Accuracy and individual variation: Alignment and Depth. The left part of the diagram is based on the model of metacognition in writing by Hacker, Keener and Kircher (2009).

Depth refers not only to metacognitive awareness of cognitive writing processes, but also to the extent to which metacognitive judgments engage with the text and in turn feed back into self-regulation strategies. To better explain this dimension, the model of metacognition in writing proposed by Hacker and colleagues (2009) has been simplified

on the left hand side of the diagram. Although this model refers to online metacognitive processes in experimental settings, it clearly shows the way metacognitive dynamics engage with the written text through monitoring and control processes. In my study, the best writers critically engaged with the text and used these metacognitive evaluations as a springboard for self-regulation. The notion of *alignment* refers to the coherence between the students' criteria for judgment and the actual requirements posed by the genre and the immediate writing context. Alignment comprises the students' genre awareness—the ability to use genre knowledge acquired in the course—but also the students' ability to meet the course criteria and learning outcomes. While the most effective writers seem to display a closer alignment and a greater depth, other students' lack of effectiveness could be explained as metacognitive inaccuracy either along the dimension of metacognitive depth, metacognitive alignment, or both.

The writers with the most effective texts were Tim, Lina, Mat and Amelia. Leo and Sara received mixed ratings, Viola and Bea low ratings. All the texts produced by these students showed a more or less successful effort of adaptation to disciplinary discourse, and therefore were different in content and form.

Depth of metacognitive judgments

Tables 3 and 4 provide a comprehensive overview of how the retrospection data was categorized. These numbers can only provide a general picture, considering that both coded segments and data sources varied in length. It should be noted that the percentages reported in tables 3 and 4 should be added together, because they indicate the proportion of coverage of the coded segment in relation to *all* the coded data in the source. Furthermore, with such small numbers these percentages can only provide a roughly grained picture of how much each student focused on the coded criteria. Nevertheless, the tables show interesting differences in focus among the students, and tendencies towards certain types of metacognitive judgments.

Table 3

Overview of coding categorization across learners – writing strategies

Coding categories	Number of segments in the data coded under each category																							
	Learner																							
	Tim			Lina			Mat			Amelia			Leo			Sara			Bea			Viola		
Writing strategies	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot
Planning	0	0	0	0	1	1	0	0	0	0	1	1	2	1	3	1	0	1	2	0	2	0	0	0
					13%	4%					13%	5%	10%	5%	6%	8%		4%	39%		20%			
Drafting	0	0	0	0	0	0	0	1	1	2	1	3	3	2	5	1	0	1	0	0	0	0	2	2
								9%	5%	32%	16%	27%	20%	13%	15%	11%		5%					36%	22%
Revising	4	0	4	3	0	3	2	0	2	0	1	1	0	1	1	2	1	3	1	1	2	1	0	1
	15%		9%	20%		14%	11%		5%		18%	6%		9%	7%	13%	15%	14%	10%	21%	15%	12%		5%

T1: Mid-term retrospection. T2: Final retrospection. Percentages indicate coverage in proportion to all the coded data in the source, and should therefore be added together with percentages reported in Table 4.

Learners described a variety of self-regulation strategies connected to writing processes such as planning, drafting and revising. These descriptions were coded under the categories illustrated in Table 3. Considering the nature of the data collected, this categorization can only be considered superficially indicative, yet it shows interesting temporal dimensions. For example, while Tim, Lina and Mat–the most effective writers–were already engaged in revision when the first retrospection was elicited (T1)—15%, 20% and 11% of the coded data respectively was devoted to descriptions of revision strategies—the other learners seem to engage primarily in drafting and planning, with revision strategies described if at all in the second retrospection (T2).

An in-depth characterization of the nature of learners' metacognitive judgments is offered in Table 4, which illustrates primarily the dimension of *depth*. Metacognitive judgments focused on three categories of criteria: content, rhetoric and language. Furthermore, two sub-categories were identified in the data: judgments that explicitly mentioned elements of the text, and judgments that were followed by a description of a self-regulatory strategy stemming from the evaluation of textual characteristics. Segments coded in this way (last row in Table 4) followed a typical structure MJ on text + SR (Metacognitive Judgment + Self-Regulation) as in the following example from Mat, one of the writers with high ratings:

'[in my paper] I miss a little the framework, the general approach. I guess, I can solve this problem by searching more in the literature'

Table 4

Overview of coding categorization across learners – metacognitive judgments

Coding categories	Number of segments in the data coded under each category																							
	Learner																							
	Tim			Lina			Mat			Amelia			Leo			Sara			Bea			Viola		
Metacognitive judgments	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot	T1	T2	Tot
Content	2 11%	1 12%	3 12%	1 8%	1 19%	2 11%	1 13%	0	1 6%	0	1 17%	1 6%	2 18%	1 7%	3 10%	3 23%	1 23%	4 23%	1 15%	2 45%	3 30%	3 48%	2 36%	5 41%
Rhetoric	3 34%	2 34%	5 34%	3 30%	1 13%	4 25%	2 18%	3 25%	4 21%	1 46%	0	1 31%	2 21%	5 30%	7 27%	2 13%	2 24%	4 19%	1 17%	1 21%	2 19%	1 16%	1 8%	2 11%
Language	1 3%	1 3%	2 3%	1 2%	1 18%	2 7%	3 17%	2 21%	5 19%	1 22%	2 36%	3 26%	0	1 1%	1 1%	0	1 15%	1 8%	1 19%	1 13%	2 16%	0	0	0
Focus on text	2 8%	2 15%	4 11%	2 8%	0	2 6%	2 12%	1 10%	3 11%	0	0	0	0	3 8%	3 6%	1 9%	0	1 4%	0	0	0	1 16%	1 20%	2 18%
Focus on text + SR	3 29%	2 35%	5 31%	3 32%	1 37%	4 33%	3 31%	3 36%	6 33%	0	0	0	2 31%	5 27%	7 28%	2 23%	1 23%	3 23%	0	0	0	1 8%	0	1 3%

T1: Mid-term retrospection. T2: Final retrospection. Percentages indicate coverage in proportion to all the coded data in the source.

The data categorization reported in Table 4 suggests tendencies towards different types of metacognitive behaviors in the students with higher and lower rhetorical effectiveness. It also illustrates differences in focus between the first (T1) and the second retrospection (T2). In the best writers, the largest proportion of coded data focused on aspects of rhetoric: 34% (Tim), 25% (Lina), 21% (Mat) and 31% (Amelia), respectively. One of the mixed-rating writers' metacognitive judgments (Leo) also fell for the most part under this category (27%), whereas the other mixed-rating writer (Sara) and the lower-rating writers (Bea and Viola) seemed primarily concerned with content. Furthermore, a good portion of the metacognitive judgments made by the writers with high and mixed ratings focused on the text and/or included a follow-up, self-regulatory strategy, whereas these two codes, and especially the latter, were nearly absent in Bea and Viola, the writers with low ratings.

The comparison of metacognitive judgment categories at both times shows some further differences possibly related to different aspects of the writing process, as is also suggested in Table 3. Rhetorical concerns were paramount for the highly effective writers (Tim, Lina, Mat and Amelia) at T1, whereas in the second elicitation some of them focused also on other aspects such as language (Mat 21%, Amelia 36%). The self-regulatory dimension, presented in Table 4 by the number of Text +SR judgments and possibly connected to the process of revision (see Table 3) is already present for these writers in the first retrospection, and occurs in connection with metacognitive evaluation at both times, except for Amelia. The metacognitive behavior of the other writers seems however more diverse. Content was the primary concern at both times for Viola and

Bea, the writers with lower evaluations; these students also made very few or no judgments on the text or text + SR (Viola a total of 18% and 3%, Bea none). Sara, who had mixed ratings, seemed to shift focus from content in T1 to rhetoric and language in T2. Additionally, she made many judgments followed by a description of a self-regulation strategy (23% in total). Puzzling are the cases of Leo and Amelia: whereas Leo, who received mixed ratings, presents a metacognitive profile very close to that of the best writers, Amelia, who received high ratings, did not comment on rhetorical elements in T2, and none of her judgments seem to deal explicitly with the text.

As mentioned earlier, the categorization illustrated in these tables offers an interesting but insufficient explanation of metacognitive accuracy and its possible connections with rhetorical effectiveness. Another dimension of metacognitive accuracy, *alignment*, was identified in the data to explain differences among students. This dimension could be understood as the coherence between students' genre-related internalized criteria and the actual criteria they are expected to meet. Alignment can be appreciated by comparing the content of the students' retrospections with the comments made by the three raters, which is the focus of the following section.

Metacognitive judgments and rhetorical effectiveness: the individual cases

To gain insight into why some students' judgments may have a relationship with the rhetorical effectiveness of their texts, it is necessary to compare the students and raters' comments. This comparison highlights issues of alignment that the metacognitive categorization presented earlier does not completely capture. Table 5 gives an overview of how the texts were assessed across macro and micro criteria. The raters were not asked to score each criterion but to put a checkmark and write a comment, because the purpose of this matrix was to elicit perceptions of quality in the de-briefing interview.

Table 5

Overview of text ratings across learner and criteria

	Learner																							
	Tim			Lina			Mat			Amelia			Leo			Sara			Bea			Viola		
Macro criteria	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
Research Story	x	x	x	x	x	x	x	x	x	x	x	x	x	x	--	x	x	x	--	--	--	--	--	--
Niche	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	--	--	x	--	--	x
Method	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	--	x	--	--	--	--	--	x
Micro criteria																								
Connecting paragraphs	x	x	x	--	x	x	x	x	--	x	x	x	x	x	--	--	x	x	--	--	--	--	--	--
Use of citations	x	x	x	x	x	x	--	x	x	x	--	x	x	x	--	x	x	x	x	x	x	x	x	x
Style	x	x	x	x	x	x	x	x	x	x	x	--	--	--	--	--	--	--	--	x	x	x	x	--

R1: Rater 1, R2: Rater 2, R3: Rater 3. Criteria are listed in summarized form (see Table 1); Xs indicate checkmarks in the matrix.

Writers with the highest ratings. As shown in Tables 3 and 4, the writers with the highest ratings—Tim, Lina, Mat and Amelia— critically appraised aspects of language, content, and rhetoric, but gave more attention to the latter. Furthermore, Tim, Mat, and Lina engaged early on with revisions, showed a tendency to focus on the text, and combined metacognitive judgments with a follow-up self-regulatory strategy. These writers could be characterized as having a greater metacognitive depth in their judgments, and adopting criteria more closely aligned with genre and contextual expectations.

Tim, the social scientist, made metacognitive judgments that adopted genre criteria to analyze precisely his text's strengths and weaknesses, evaluate parts of his text, explain what worked, and what revisions were needed. This MJ on text + SR structure was frequent in both retrospections, for example in T1:

(1) 'I have highlighted the purpose in a sensible way, however my "method" section needs more work … My research question is very relevant; however I need much more. I feel that I have been critical, but I could have taken that further'

Tim was at the first stages of his MA research, and in his final retrospection he explained how he addressed his perceived limitations in content knowledge by focusing on clarity and style:

(2) 'The literature review: I struggled to not make it look like a list of books. I don't think I have succeeded in writing one that was as good as I wanted it to be. So I tried to be as concise and academic as possible'

The criteria adopted in Tim's metacognitive judgments seemed to align with the expectation of the raters. All agreed that his paper highlighted a clear gap by creating a critical and convincing research story with appropriate support in previous work. Similarly to his appraisal in (2), the raters agreed that his work suffered from some methodological and theoretical shortcomings, yet its quality lay in the clarity of argumentation and his ability to create a niche 'through the paragraphs; his arguments highlight the need for this study' (R2).

Lina, a conservationist who had just completed her PhD program, also received positive evaluations from all three raters. Lina's final text examined a new methodological procedure, and therefore created the niche without an extensive review of previous research: this peculiarity raised minor concerns in rater 3 ('the method is the

study'), but otherwise her text received unanimous praise. Analogously to Tim, she critically examined her text in light of genre expectations, and often her evaluations were combined with descriptions of self-regulation strategies to address the areas needing improvement. For example, in her mid-term retrospection Lina commented on the need to balance content and brevity against rhetorical persuasiveness and contextualization of the research:

(3) 'The results section I fear it is slightly weak. Exclusion is problematic … I want this to be clear to the reader without underrating what I have done, which I find difficult to balance'

In her final retrospection Lina showed again metacognitive awareness of the need to address genre criteria as well as language criteria, combining evaluation with self-regulation:

(4) 'I struggled with wordiness and informal language . . . I also knew I had to seriously cut down the amount of words and I found it challenging to prioritize what was important enough to be included'

The ability to create a convincing story was the primary reason why Mat and Amelia's texts were considered rhetorically effective, although the raters' comments occasionally diverged.

Mat, an archaeologist, also used a variety of criteria to make judgments of why specific parts of his text were more or less effective. For instance, Mat commented on the link between conciseness and relevance:

(5) 'Problems in writing short, compact texts. It is hard for me to decide between important and unimportant information, to know how deep to go'

In his final retrospection, Mat focused on rhetorical concerns, showing sensitivity to audience expectations:

(6) 'I put most work in the introduction . . . from a more general overview to the narrow, specific research topic in a few sentences.. (it) should be comprehensible for people who are not familiar'

Mat's focus on creating a rhetorical niche in the introduction was coherent with the raters' assessment, and his deliberate choice to give less attention to the literature review was also captured by the raters' comments, who mentioned this aspect in

relation to micro criteria: rater 1 felt that the literature review read like a 'laundry list'; rater 3, on the other hand, commented on the deviation from the typical organization of academic texts: 'the structure does not follow the background-gap-aim . . . It disrupts the flow'.

Finally Amelia, a doctoral student in a practice-oriented field of education, also obtained positive assessments on macro criteria by all three raters. As opposed to the other effective writers, Amelia seemed engaged all along with the 'bigger picture': her descriptions centered on drafting rather than revision (table 3), and although aspects of rhetoric were equally important as aspects of language, most of her metacognitive judgments did not engage directly with the text (nor follow-up self-regulation strategies). Despite this apparent lack of depth—of course the type of data elicited cannot represent actual online processes taking place during transcription—her focus on rhetoric and her evaluation criteria seemed to align with genre expectations, resulting in an effective text. In her first retrospection Amelia expressed difficulty with the L2, but mostly remarked on the need for congruence between literature review and methodology:

(7) 'I have the most difficulty writing the literature review and the methodology . . . I have to base my studies on a theoretical framework that is new to me, and I have only bad experience in trying to make the literature fit in. . . . I have difficulty with expressing myself in English, it takes time to write'

Similarly, in her final retrospection, Amelia underscored the need for more conceptual clarity and conciseness in the literature review and methodology. In terms of rhetorical effectiveness, raters agreed about the text's effectiveness in terms of construction of an argument, suggesting that Amelia's concerns did indeed align with genre and context expectations. Amelia's self-perceived struggle with language and informal style were also echoed in the raters' comments, although their opinions slightly differed: R3 was critical of the style as 'too personal and unscholarly'; whereas R1 and R2 thought this was in line with the style of practitioner-oriented publications.

Writers with mixed ratings. Sara and Leo received mixed reviews from the raters, albeit for very different reasons.

Sara, although still in the first stages of her MA program, showed metacognitive depth by analyzing what needed improvement in her work and why. Despite self-

admitted deficits in disciplinary knowledge, Sara attempted to use what she knew to achieve a rhetorical function (argue, justify), and she often displayed a MJ on text + SR structure in her judgments:

(8) 'I struggle writing my argumentation and theoretical frame … I could argue more under justification. I could also develop my method under theoretical framework'

This sensitivity to rhetorical needs was also evident in a comment about her introduction:

(9) 'I can see what differs in my work from theirs (the classmates). I would like to keep that because I think it helps the reader to quickly go through the information'

In her final retrospection, Sara once again seemed aware that her text reflected her lack of content knowledge:

(10) 'I have struggled a lot with my theoretical framework and my literature review . . . I have not had enough material to work with in my writing'

Rather than depth or alignment of metacognitive judgments, Sara's lack of rhetorical effectiveness appeared to lie in her inexperience and her lack of content knowledge, as suggested in Table 4 by the fact that she devoted slightly more focus to content (23% of coded data). Nevertheless, she was accurate in her self-assessment, which largely coincided with the raters' comments. They agreed that her text showed that she was still 'green' (R2), but also that it represented a 'good effort . . . she thought critically about how to use the theories she discussed' (R3); although 'support is a bit weak' (R1), her text seemed 'to have potential' (R2).

In contrast, Leo, a former physicist now phonetician, had the opposite problem. Although he displayed a metacognitive profile very similar to the best writers, with much attention to rhetorical criteria, the text and self-regulation, his perception of the task and his criteria for evaluation misaligned with the raters'. This misalignment problem is illustrated by Leo's goals for writing, which were very much tied to the research problem he was attempting to define:

(11) [to come across as]'An honest, competent researcher and that I have thought this matter through myself. That's my strategy'

Leo's rhetorical choices were motivated by this goal; the following quote illustrates Leo's choice to give priority to his personal aim rather than the course content:

(12) 'I have not thought so much about the things we learnt on the course yet because I need to write my text first'

Only in his final retrospection did Leo mention the need to meet genre and disciplinary community expectations, albeit briefly and hesitantly:

(13) 'It was a mistake to discuss the weaknesses of previous research but I did it anyway because it was the rule . . . The rest of the text was logical and honest, those stupid examples there weakened that spirit of the pioneer or the innocent scientist who's just doing his job and happens to question old rules . . . Still, I don't know if it will be accepted by the readers'

The raters' comments reflected this misalignment. His deliberate strategy to use a personal tone to foreground logic and scientific honesty was a cause of puzzlement for all three, and rater 3 commented that the text was telling 'the wrong story, a fairy-tale story' (as opposed to a research story). They recognized the value of Leo's research within his discipline (macro criteria), but all three agreed that the register and especially the style were problematic: 'although convincing, this could never work as a publishable text' (R1); 'the whole story reads like a personal diary of a scientist' (R2); 'this is not how phoneticians write: the rhetoric is normally very cold, formal, dry' (R3). Leo's strategy to achieve credibility through atypical rhetorical choices backfired because of misalignments with genre and contextual requirements.

Writers with the lowest ratings. The raters unanimously agreed in their assessment that Bea and Viola, for very different reasons, wrote the least effective texts.

Bea, a doctoral student in the final stages of her applied education program, displayed a developing understanding of academic genres:

(14) 'To write an article that is not "simply" presenting my study, but discussing one of the results in dialogue with other resent [sic] research on the topic is new to me'

However, concerns for how the text met genre requirements were overtaken by concerns for content and conceptualization of the text (as shown in Table 3 and 4):

(15) 'The result is utterly interesting, which leaves me in need to both find and read new literature/articles on the topic. Do I have time to do this thoroughly before the course is over?'

Bea seemed to struggle with conceptualization even in her second retrospection, where 30% of her coded comments concerned content (Table 3), and rhetorical concerns about genre conventions were expressed in general terms:

(16) 'trying to "cover" all parts of what should be addressed . . . fit the format of an Academic text'

Bea's concern for content rather than genre is reflected in the raters' assessment: all three could not recognize the genre of her text. Problems were identified both at the macro level and micro-level, but its main flaw was the lack of a niche to create coherence and convey significance: 'I cannot see a clear niche; I am not sure what the aim is' (R2); 'The study lacks a clear aim and a well-argumented niche' (R1); 'not much connection between aim and previous research: what did she actually do?' (R3). Bea's metacognitive judgments were not over-confident and denoted a certain awareness of problems, but overall remained on a superficial level and did not engage with the text.

Viola also seemed to focus on content at both times (see Table 4). In her judgments, she mentioned several important genre criteria, showing general genre awareness. However, on the metacognitive level she often made positive and over-confident judgments on the quality of her text, which resulted in almost no description of self-regulatory strategies for revision:

(17) 'I believe the study is well thought-through and cohesive with a clear idea of how it will be executed and with problematizing research questions. The research gap is clear and the implications and applications are defined'

When a weakness was mentioned, Viola indicated the lack of content rather than rhetorical shortcomings:

(18) 'The literary review is currently too slim, largely due to the trouble of finding useful material'

Positive and rather uncritical evaluations of several rhetorical criteria also characterized Viola's final retrospection: the theory and the method sections 'discuss appropriate theoretical frames' and show that her research is 'viable' (her words).

Finally, in terms of rhetorical effectiveness the raters described Viola's text as having all the right moves but not ringing true: 'it seems that she went through a checklist of requirements of an academic text, but that she did not really think about them' (R2). All three raters made almost identical comments about the fact that these 'required elements' did not coalesce into an argument: 'it tells you that it is important, but it does not show you why' (R1, R2), showing that rhetorical effectiveness is more than the sum of its parts:

'It imitates the genre, rather than making it her own' (R2)

'It feels like someone *performing* the genre, but not critically bringing it together' (R3)

Discussion

This study aimed at providing an understanding of qualitative differences in metacognitive accuracy in graduate L2 writers and possible connections with more or less successful academic writing in English. In addition, the goal was to identify useful analytical tools for further study of metacognition and the development of academic writing competence in authentic settings through a) an account of the criteria adopted by L2 writers in their metacognitive judgments, and b) an explanation of how the accuracy of these judgments may connect with rhetorical effectiveness. In the following I will review the key findings and discuss implications for further investigation and pedagogical practice.

To summarize, accuracy in this study has been framed along two dimensions: depth of metacognition and alignment of judgment criteria. Accurate calibration in academic writing seems connected to both. The most metacognitively accurate students in the study were also—but not always—the writers with the most effective texts (which were very different in content and style). Sara for instance, who received mixed evaluations, was accurate in assessing that her lack of content knowledge could undermine the rhetorical effectiveness of her text in the raters' eyes. The qualitative analysis of the nature of metacognitive judgments illuminates interesting differences among the students. In terms of the criteria adopted, three major categories were identified: content, rhetoric, and language. A further refinement of this categorization

identified judgments with a focus on text, and text plus a description of a follow-up self-regulatory strategy (MJ on Text + SR).

The dimension of *depth*, represented in Fig. 1 and supported by the metacognitive model of writing by Hacker et al. (2009), shows that effective metacognitive monitoring in writing critically engages *with the text*. The students with the most positive raters' reviews invoked genre-derived criteria of content, rhetoric and language to evaluate specific aspects of their texts and determine how to improve them. This metacognitive maturity helped them overcome lack of experience and conceptualization challenges: although three of the most effective writers—Lina, Mat and Amelia—could rely on their content knowledge (as opposed to Sara and Tim, with mixed and high ratings respectively), all of them perceived as inadequate their experience of academic genres, and yet managed to muster resources to compensate for these conceptual and/or experiential shortcomings. Critical metacognitive judgments for instance helped Sara achieve some kind of rhetorical effectiveness despite her disciplinary inexperience. Tim, the L1 writer, displayed a metacognitive profile similar to the other effective writers, which aligns with findings suggesting that metacognitive knowledge accounts for writing performance in both L1 and L2 (see Schoonen et.al. 2009, 2011). These dynamics support educational psychology findings indicating that conditional metacognitive knowledge—knowing how and why to apply knowledge and strategies to the specific situation—'enables students to . . . compensate for low ability or lack of relevant prior knowledge' (Schraw 1998: 117).

The dimension of depth helps clarify why other students were less successful: they did not seem to be metacognitively aware of doing something that did not contribute to writing an effective text. Students seemed to engage with different aspects of the writing experience—planning (conceptualization), drafting and revising—at different paces, which connects with cognitive research on L2 writing emphasizing that the distribution of metacognitive behaviors at different points in time may have an impact on writing quality (Tillema et al. 2011). At the same time, as pointed out by a recent study comparing timed and non-timed writing (Khuder and Harwood 2015), text quality is not necessarily always a result of the distribution or efficiency of cognitive processes. Viola and Bea, the least rhetorically effective writers, did not make in-depth comments about their texts and seemed 'stuck' on generating ideas and conceptualizing the content of their papers; whereas Bea expressed some uncertainty about her

performance, Viola showed a 'false sense of confidence' (Hirvela 2011): hers is a classic case of miscalibration, caused by difficulty in recognizing one's shortcomings and resulting in bias towards overconfidence (Kruger and Dunning 1999).

These findings support previous work suggesting that effective self-regulation in writing is achieved only when writers develop metacognitive awareness in connection with both situated and rhetorical aspects of the task (Negretti 2012). Metacognitive depth can only offer a partial explanation: writers need to connect the text they are producing (and their writing strategies) to their emergent genre knowledge (Tardy, 2009) and the communicative demands of the task. The dimension of alignment is therefore necessary to appreciate accuracy of metacognitive judgments in academic writing: 'even the best writers with the best intentions can produce words that are meaningful to them but will fail to be meaningful to another' (Hacker et al. 2009: 156).

Alignment refers to the students' ability to use their developing genre awareness to negotiate the conventions of academic genres in their disciplines *and* the specific requirements of the learning situation, although in this study the disciplinary dimension was evaluated only by academic genre experts. The writers with the most successful texts (Lina, Mat, Amelia, Tim), in their judgments were able to apply genre knowledge to the specific conditions of the task: the final assignment and its requirements. To use Cumming's (1989) famous distinction, the less effective writers seem to have declarative knowledge of genre, but not procedural knowledge; to use the metacognitive awareness framework from educational psychology (Schraw and Dennison 1994), they may have declarative and even procedural awareness of how to perform the genre, but lack the conditional awareness of how to use it in this situation. While for the writers with lower ratings (Bea and Viola) problems of metacognitive depth and alignment seemed to intertwine, misalignment is especially evident in the case of Leo, one of the two writers who received mixed evaluations. In his case, accuracy seems affected by the misalignment between his criteria and the actual standards upon which performance was assessed in context. Leo was resisting genre conventions to convey an 'honest, innocent scientist' persona and was frustrated by the 'dumb' conventions of academic rhetoric (his words). His conflict recalls Cumming's (2013: 135) discussion of the challenges tied to expressing personal identity in L2 academic writing, with conventions 'bounding possibilities of selfhood', and Hirvela's (2011: 54) observations about the difficulties for learners of establishing L2 'writerly selves'. The dimension of alignment shows that

the connection between metacognition and effectiveness cannot be explained only in terms of type or frequency of metacognitive behaviors: other factors, such as personal goals, task perceptions, and the coherence between internal and contextual standards merit consideration (Pieschl 2009).

In terms of further research, this small-scale study suggests the importance of combining a focus on the cognitive, textual and social dimensions of writing to achieve ecological validity (Dinsmore and Parkinson 2013). In authentic settings the assessment (and self-assessment) of text quality needs to reflect rhetorical and contextual demands. Further studies could adopt a matrix of context-relevant criteria such as the one used in this study, ensuring that both criteria and raters are connected to the learning situation. Because the students in this study were a disciplinary heterogeneous group taking an ESP course, ESP teachers were recruited as raters. With more homogeneous groups of students, it would be important to enlist the collaboration of disciplinary experts: the study of metacognitive accuracy, when approached in naturalistic settings, cannot overlook the question of who has the right to establish what constitutes acceptable writing criteria. These suggestions do not necessarily mean adopting only qualitative approaches. A limitation of this study is that it assumes that the raters—being experienced applied linguists and teachers of the course—could be presumed to make reliable judgments. This approach to the assessment of text quality was valuable to elicit their stakeholders' perspectives on rhetorical effectiveness in the de-briefing interviews; however, this same matrix could be used as a rating protocol together with reliability-enhancing procedures (c.f. van den Bergh and Rijlaarsdam 2001) to provide points for comparison across raters, both EAP/ESP teachers and discipline experts, especially with a larger data set. Furthermore, providing such a matrix to students, academic writing teachers and discipline experts would allow for a mixed-method investigation of calibration in writing, possibly combining traditional quantitative measures (Schraw 2009) with qualitative data such as writing journals and/or stimulated recall interviews. Overall, this study outlines possible trajectories for an investigation of metacognition and self-regulation in connection with the development of genre knowledge.

In terms of pedagogical practice, this investigation suggests that even advanced graduate L2 writers may need support in applying their emergent genre knowledge to the immediate task. Johns (2008: 246) underlines both the social and the conventional nature of academic genres, and reminds us that the purpose of genre pedagogy is 'to

educate, rather than train, novice academic students through the promotion of genre awareness and rhetorical flexibility'. Opportunities for metacognitive reflection and self-regulation on genre-relevant criteria provide students with a ladder to climb towards successful negotiation of social practices and personal voice, especially from a formative assessment perspective (Graham, Harris and Hebert 2011). Pedagogical interventions could take the form of individual or peer evaluation tasks for goal setting and further work, or 'metalinguistic conversations' (Ofte 2014) about how genre knowledge can be used by the students to write various academic tasks. Furthermore, this study underscores the need for an explicit—and critical—discussion of expectations and conventions in the genre-based classroom. Students do not write academic texts in experimental settings, they write in different contexts and disciplines, and they may struggle in negotiating the (for them) often ambiguous conventions of the genre they are asked to produce. As Hyland (2011: 22) emphasizes, 'the claim that writing is context-free . . . is now largely discredited . . . texts are a response to particular communicative settings'. Research on metacognition and self-regulation in academic writing cannot ignore the dimensions of power and situatedness tied to academic genres; this study has shown that lack of metacognitive depth and misalignments of criteria affect metacognitive accuracy and possibly performance. An open discussion could raise students *and* teachers' awareness of genre expectations.

Concluding remarks

The calibration paradigm is valuable as it raises the issue of *accuracy*, which has been interpreted here along the dimensions of metacognitive depth and alignment. This adds a new understanding of the qualitative nature of metacognition in academic writing by showing that calibration is negotiated in the internalization of contextual, communicative criteria. This study complements research in L2 writing focused on distribution and frequency of metacognitive behaviors (Ong 2014; Tillema et. al. 2011), and L1/L2 differences (Schoonen et. al 2009, 2011), by exploring metacognition in connection with the development of genre awareness. It extends previous work (Negretti 2012, Negretti and Kuteeva 2011) suggesting that the most effective writers metacognitively balance their task perceptions, genre/rhetorical requirements, and the situated, communicative purpose of the writing task.

The qualitative approach and the use of the calibration paradigm have been useful in showing that 'there are many roads to success as well as to failure' (van Lier 2010:

603). This study has attempted to show the benefits of studying metacognitive processes in an authentic setting, capturing the complex interplay between writer, text and context, as miscalibration brings consequences primarily for learners: 'the question of "what works?" is elliptical for "what works to produce valued education outcomes?"'(Howe 2009: 431).

Acknowledgments

References

Alexander, P. 2013. 'Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration.' *Learning and Instruction* 24: 1-3.

Berkenkotter, C. and T. Huckin 1995. *Genre Knowledge in Disciplinary Communication.* Lawrence Erlbaum.

Breetvelt, I., H. van den Bergh and G. Rijlaarsdam 1994. 'Relations between writing processes and text quality: when and how?' *Cognition and instruction* 12: 103-123.

Brown, A. L. 1987. 'Metacognition, executive control, self-regulation, and other more mysterious mechanisms' in Weinert and Kluwe (eds.).

Charmaz, K. 2006. *Constructing Grounded Theory*. Sage.

Chaudron, C. 2003. 'Data collection in SLA research' in Doughty and Long (eds.).

Cumming, A. 1989. 'Writing expertise and second-language proficiency.' *Language Learning* 39/1: 81-141.

Cumming, A. 2013. 'Multiple dimensions of academic language and literacy development' in Ortega, Cumming and Ellis (eds.).

Dinsmore, D. L. and M. M. Parkinson 2013. 'What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration.' *Learning and Instruction* 24: 4-14.

Dinsmore, D. L., P. A. Alexander and S. M. Loughlin 2008. 'Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning.' *Educational Psychology Review* 20: 391-409.

Dunlosky, J. and K. Thiede 2013. 'Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement.' *Learning and Instruction* 24: 58-61.

Doughty C. and M. Long 2003. *The Handbook of Second Language Acquisition*. Blackwell.

Feak, C. B. and J. M. Swales 2009. *Telling a Research Story: Writing a Literature Review*. Michigan University Press.

Feak, C. B. and J. M. Swales 2011. *Creating Contexts: Writing Introductions across Genres*. University of Michigan Press.

Flavell, J. H. 1979. 'Metacognition and cognitive monitoring: A new era of cognitive-developmental inquiry.' *American Psychologist* 34/10: 906-911.

Flower, L. and J. R. Hayes 1980. 'The cognition of discovery: defining a rhetorical problem.' *College Composition and Communication* 31: 21-32.

Gao, X. 2007. 'Has language learning strategy come to an end? A response to Tseng et. al (2006)' *Applied Linguistics* 28/4: 625-620.

Gentil, G. 2011. 'A biliteracy agenda for genre research.' *Journal of Second Language Writing* 20/1: 6-23.

Graham, S., Harris, K., and Hebert, M. 2011. 'Informing writing: The benefits of formative assessment.' A report from Carnegie Corporation of New York. *Carnegie Corporation of New York*.

Greene, S. and L. Higgins 1994. '"Once upon a time": the use of retrospective accounts in building theory in composition' in Smagorinsky (ed.).

Hacker, D. J., M. C. Keener and J. C. Kircher 2009. 'Writing is applied metacognition' in Hacker, Dunlosky and Graesser (eds.).

Hacker, D. J., J. Dunlosky and A. C. Graesser 1998. *Metacognition in Educational Theory and Practice*. Lawrence Erlbaum.

Hacker, D. J., J. Dunlosky and A. C. Graesser 2009. *Handbook of Metacognition in Education*. Routledge.

Harwood, N. 2005. '"We do not seem to have a theory...The theory I present here attempts to fill this gap": Inclusive and exclusive pronouns in academic writing.' *Applied Linguistics* 26/3: 343-375.

Hirvela, A. 2011. 'Writing to learn in content areas: research insights' in Manchón (ed.).

Howe, K. 2009. 'Epistemology, methodology, and education science: Positivist dogmas, rhetoric, and the education science question.' *Educational Researcher* 38/6: 428-440.

Hyland, K. 2007. 'Genre pedagogy: language, literacy and L2 writing instruction.' *Journal of Second Language Writing* 16: 148-164.

Hyland, K. 2011. 'Learning to write: issues in theory, research, and pedagogy' in Manchón (ed.).

Hyland, K. 2012. *Disciplinary identities: Individuality and Community in Academic Discourse*. Ernst Klett Sprachen.

Hyland, K. and P. Tse 2004. 'Metadiscourse in academic writing: A reappraisal.' *Applied Linguistics* 25/2: 156-177.

Hyon, S. 1996. 'Genre in three traditions: implications for ESL.' *TESOL Quarterly* 30: 693–722.

Johns, A. M. 2008. 'Genre awareness for the novice academic student: An ongoing quest.' *Language Teaching* 41/2: 237-252.

Khuder, B. and N. Harwood 2015. L2 Writing in Test and Non-test Situations: Process and Product. *Journal of Writing Research* 6/3: 233-278.

Kobayashi, H. and C. Rinnert 2013. 'L1/L2/L3 writing development: Longitudinal case study of a Japanese multicompetent writer.' *Journal of Second Language Writing* 22/1: 4-33.

Kruger, J. and D. Dunning 1999. 'Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments'. *Journal of Personality and Social Psychology* 77/6: 1121–34. doi:10.1037/0022-3514.77.6.1121

Larsen-Freeman, D. 2006. 'The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English.' *Applied linguistics* 27/4: 590-619.

Manchón, R. M. 2009. *Writing in Foreign Language Contexts: Learning, Teaching, and Research*. Multilingual Matters.

Manchón, R. M. 2011. *Learning-to-write and Writing-to-learn in Additional Language*. John Benjamins.

Manchón, R. M. and J. Roca de Larios 2007. 'On the temporal nature of planning in L1 and L2 composing.' *Language learning* 57/4: 549-593.

Metcalfe, J. 2009. 'Metacognitive judgments and control of study.' *Current Directions in Psychological Science* 18: 159-163.

Negretti, R. 2012. Metacognition in student academic writing: A longitudinal study of metacognitive awareness and its relation to task perception and evaluation of performance. *Written Communication* 29/2: 142-179.

Negretti, R. and M. Kuteeva 2011. Fostering metacognitive genre awareness in L2 academic reading and writing: A case study of pre-service English teachers. *Journal of Second Language Writing* 20: 95–110.

Nietfeld, J., L. Cao and J. W. Osborne 2005. Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education* 74/1: 7–28.

Ofte, I. 2014. English academic writing proficiency in higher education: Facilitating the transition from metalinguistic awareness to metalinguistic competence. *Acta Didactica Norge* 8/2: 1-15.

Ong, J. 2014. 'How do planning time and task conditions affect metacognitive processes of L2 writers?' *Journal of Second Language Writing* 23: 17-30.

Ong, J. and L. J. Zhang 2013. 'Effects of manipulation of cognitive processes on English-as-a-foreign-language (EFL) writers' text quality.' *TESOL Quarterly* 47: 375-390.

Ortega, L. 2012. 'Epilogue: exploring L2 writing-SLA interfaces.' *Journal of Second Language Writing* 21/4: 404-415.

Ortega, L., A. Cumming and N. Ellis 2013. *Agendas for Language Learning Research*. Wiley.

Pecorari, D. and P. Shaw 2012. 'Types of student intertextuality and faculty attitudes.' *Journal of second language writing* 21/2: 149-164.

Pieschl, S. 2009. 'Metacognitive calibration: An extended conceptualization and potential applications.' *Metacognition and Learning* 4/1: 3-31.

Pieschl, S., E. Stahl, T. Murray and R. Bromme 2012. 'Is adaptation to task complexity really beneficial for performance?' *Learning and Instruction* 22: 281-289.

Pressley, M. and E. S. Ghatala 1988. 'Delusions about performance on multiple-choice comprehension tests.' *Reading Research Quarterly* 24/4: 454–464.

Reiff, M. J. and A. Bawarshi 2011. 'Tracing discursive resources: how students use prior genre knowledge to negotiate new writing contexts in first-year composition.' *Written Communication* 28/3: 312-337.

Sasaki, M. 2011. 'Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A longitudinal study.' *TESOL Quarterly* 45/1: 81-105.

Scardamalia, M. and P. Paris 1985. 'The function of explicit discourse knowledge in the development of text representations and composing strategies.' *Cognition and Instruction* 2/1: 1-39.

Schoonen, R., P. Snellings, M. Stevenson and A. van Gelderen 2009. 'Towards a blueprint of the foreign language writer: the linguistic and cognitive demands of foreign language writing' in Manchón (ed.).

Schoonen, R., A. Van Gelderen, R. D. Stoel, J. Hulstijin and K. Glopper 2011. 'Modeling the development of L1 and EFL writing proficiency of secondary school students.' *Language Learning* 61/1: 31-79.

Schraw, G. 1998. 'Promoting general metacognitive awareness.' *Instructional Science* 26: 113-125.

Schraw, G. 2009. 'A conceptual analysis of five measures of metacognitive monitoring.' *Metacognition and Learning* 4/1: 33-45.

Schraw, G. and R. S. Dennison 1994. 'Assessing metacognitive awareness.' *Contemporary Educational Psychology* 19: 460-475.

Silverman, D. 2001. *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*. Sage.

Smagorinsky, P. 1994. *Speaking about Writing*. Sage.

Smagorinsky, P. 2008. 'The method section as conceptual epicenter in constructing social science research reports.' *Written Communication* 25/3: 389-411.

Spolsky, B. and F. M. Hult 2010. *Handbook of Educational Linguistics*. Wiley. com.

Swales, J. M. 1990. *Genre Analysis: English in Academic and Research Settings*. CUP.

Swales, J. M. and C. B. Feak 2012. *Academic Writing for Graduate Students*. Michigan University Press

Tardy, C. 2009. *Building Genre Knowledge*. Parlor Press.

Tillema M., H. van den Bergh, G. Rijlaarsdam and T. Sanders 2011. 'Relating self reports of writing behaviour and online task execution using a temporal model.' *Metacognition and Learning* 6/3: 229-253.

Thompson, G. 2001. 'Interaction in academic writing: Learning to argue with the reader.' *Applied Linguistics* 22/1: 58-78.

van den Bergh, H. and G. Rijlaarsdam 2001. 'Changes in cognitive activities during the writing process and relations with text quality.' *Educational Psychology* 21: 373-385.

van Lier, L. 2010. 'Ecological-semiotic perspectives on educational linguistics' in Spolsky and Hult (eds.).

Vandergrift, L. 2005. 'Relationships among motivation orientations, metacognitive awareness and proficiency in L2 listening.' *Applied Linguistics* 26: 70–89.

Veenman, M., B. Van Hout-Wolters and P. Afflerbach 2006. 'Metacognition and learning: conceptual and methodological considerations.' *Metacognition and Learning* 1: 3-14.

Victori, M. 1999. 'An analysis of writing knowledge in EFL composing: a case study of two effective and two less effective writers.' *System* 27: 537-555.

Weinert, F. and R. Kluwe 1987. *Handbook of Child Psychology: Vol. 3 Cognitive Development*.Wiley.

Wenden, A. 1987. 'Metacognition: an expanded view on the cognitive abilities of L2 learners.' *Language Learning* 37/4: 573-597.

Wenden, A. 1998. 'Metacognitive knowledge and language learning.' *Applied Linguistics* 19: 515–37.

Winne, P. H. and A. F. Hadwin 1998. 'Studying as self-regulated learning' in Hacker, Dunlosky and Graesser (eds.).

Winne, P. H. and D. Jamieson-Noel 2002. Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology* 27/4: 551–572.

Zimmerman, B. J., and A. Bandura 1994. 'Impact of self-regulatory influences on
writing course attainment.' *American Educational Research Journal* 31/4: 845-
862.

---

i   Considerations of space do not allow for a detailed justification of each criteria: the reader may wish to consult the specific
sections in the literature. For Macro criterion (a), see Feak and Swales (2011: 10 and 2009), and Paltridge and Starfield (2007, cited
in Swales and Feak 2011: 21); for Macro (b), see Feak and Swales (2011: 59); for Macro (c), see Smagorinsky (2008: 48), and
Swales and Feak (2012: 289). For Micro criterion (a), see Feak and Swales (2011 77-80); for (b) see Feak and Swales (2011: 59-
60); for Micro (c), see discussion in Swales and Feak (2012: 14).