

Administratief verantwoordelijk promotor

Prof. dr. Geert Jacobs
Vakgroep Taalkunde

Promotoren

Prof. dr. Tammy Schellens
Vakgroep Onderwijskunde, FPPW

Prof. dr. Torsten Leuschner
Vakgroep Taalkunde

Decaan

Prof. dr. Marc Boone

Rector

Prof. dr. Anne De Paepe

Nederlandse vertaling:

Online technologieën in het academisch schrijfonderwijs in een vreemde taal.
Een onderzoek naar hun potentieel voor individueel en samenwerkend leren.

Cover illustration by Anouk Van Audenhove



Faculteit Letteren & Wijsbegeerte

Carola Strobl

*Affordances of online technologies
for academic writing instruction
in a foreign language*

An exploration in individual and collaborative settings

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de taalkunde

Mei 2015

Dit onderzoek werd gefinancierd door het Onderzoeksfonds van de Hogeschool Gent (HOF)

Abstract

Online pedagogy affords tertiary education fresh scope. The integration of ICTs in university curricula facilitates flexibility in terms of time, space, and individual learning pace. While they allow a broader public access to post-secondary education through open and distance-learning courses, online technologies also open up new instructional perspectives in that they provide different opportunities for feedback, interaction, and process tracking. These opportunities also should to be utilized for language learning. Language learning is primarily associated with face-to-face human interaction. However, human-computer interaction and human interaction facilitated by online technology provide a broad array of additional opportunities for language instruction and uptake. This PhD seeks to unravel the potential of online technologies for the instruction of academic writing in a foreign language, more specifically "writing from sources". This task plays a crucial role both in the academic curriculum (being able to synthesise literature for papers) and in the future professional lives of students majoring in foreign languages. The presented research is therefore situated at the crossroads of two academic disciplines, viz. CALL (Computer-Assisted Language Learning) and Writing Instruction.

To investigate the potential of online learning environments for academic writing, a design-based research cycle of five studies was set up. Two preliminary observational studies based on the commercial online learning environment DUO (Deutsch-Uni Online) served to gain insight into the online learning behaviour of the target group engaged in writing-related tasks. Based on the results, online learning material was created for the three intervention studies in which 162 second- and third-year students majoring in German as a foreign language at two higher education institutions participated. The intervention studies lasted between two and six weeks and aimed at instructing students in two subforms of "writing from sources", viz. summary writing from a spoken source and synthesis writing from several written sources. Beside writing abilities, these tasks require extensive reading or listening skills as well as critical thinking abilities. Information from sources needs to be processed, selected, connected, and organised to build a new text. The writing process places a lot of strain on working

memory capacity because writers need to allocate attention to different mental representations of the emerging text simultaneously. From a pedagogical point of view, mediating this high cognitive load is important to maximize student performance. To explore how online technologies can contribute to this mediating role, several task designs were developed according to different pedagogical approaches, using an array of different online technologies. Consequently, the interventions of this PhD also reflect the development in CALL from the computer-as-tutor, guiding students in the learning process through instructions and feedback, to the computer-as-tool, providing a platform for students-student and student-teacher interaction. The former intends to convey instructivist exercises to build knowledge through pattern drill and cognitive constructivist exercises targeting strategy and skills development, whereas the latter is used in collaborative activities reflecting a socio-constructivist approach to learning.

The theoretical framework for this PhD stems from the Sociocultural Theory of Learning. The concept of *affordances* refers to the inherent characteristics of artefacts (here: online tools) that facilitate, or impede, an activity (here: learning to write from sources in German L2). Those characteristics only have an impact on learning when learners actually perceive, and use, the artefacts in the intended way. To investigate whether this was the case, all intervention studies were evaluated using an Activity Theoretical approach. Activity Theory provides a theoretical-methodological framework for the analysis of the complex interplay of environmental factors in a goal-oriented action that (fail to) enable the affordances of a tool used in and for the action. Therefore, it presents a good basis to discuss innovation through CALL tools in instructional settings. To this aim, the intervention studies were carefully monitored, yielding rich data that were analysed adopting a mixed-methods approach. Pre-post-tests were analysed along text quality parameters to quantitatively measure the impact of the interventions on overall writing quality. Process recordings via keystroke logging and screen-capturing tools were analysed to unravel changes in the writing processes and strategy use. Finally, questionnaires and focus group interviews served to triangulate the observational data with students' attitudes and self-reported learning gain.

Observations of the preliminary research revealed that listening for gist and planning before writing are two important strategies to focus on. Critical success factors in online language learning modules that were singled out are the salience of task instructions and the need to dovetail preparatory language exercises with the corresponding tasks that target skills development by providing content links.

The critical evaluation of the intervention studies culminates in a discussion of affordances of the mediating artefacts used to build online activities according to different design principles. The University's authoring tool *curios* served to construct exercise suites according to instructivist and constructivist principles. For collaborative tasks inspired in the socio-constructivist paradigm, a learning path and a forum

provided within the University's Virtual Learning Environment *Minerva*, and *Google Docs* were used. On the one hand, their potential for effective support of learning-to-write activities in individual and collaborative contexts was revealed in qualitative process analysis, providing evidence of strategy development and of increased awareness of critical text quality features. On the other hand, no statistically significant change in overall text quality could be stated from the pre-post-test comparisons. This might be attributed to the fact that learning-to-write is a long-term process in which changes occur gradually. The analysis also provides evidence that the uptake of the affordances of mediating artefacts highly depends on pedagogically sound design, i.e., the combination of tools and secondary artefacts used within the activity, like a model video and a script, along with other factors that can impose physical (e.g., lack of accessibility), cultural (e.g., learner histories or attitudes), or logical (bugs or unintended interaction with supporting technologies) constraints.

The main sources of contradictions that hampered the uptake of affordances could be linked to institutional rules and students' educational histories. In an overarching learning institution environment that is centred on individual countable learning outcomes, students develop a performance-oriented, rather than a mastery-oriented, mind-set. However, the latter is best combined with computer-mediated activities that transcend a pattern-drill approach, and especially those that rely on collaborative learning principles. This contradiction is reflected in the students' attitudes towards the online learning material. The main findings in this regard showed a stark contrast between students' positive attitudes towards pre-programmed feedback in closed tasks inspired by instructivist learning theories and their mixed attitudes towards feedback in the form of model solutions and reflective questions for self-evaluation. In addition, a positive response to technologies/task design used to support *individual* writing, as opposed to a more critical stance towards the technologies to support *collaborative* writing, was noted. This needs to be qualified by stating that the acceptance of the computer-as-tool for collaboration was higher when it was combined with secondary artefacts intended to steer the collaborative process, like a model or a script.

One of the main conclusions in favour of collaborative settings for pre-writing planning was that content elaboration, which plays a pivotal role in academic writing, was significantly better than in individual settings. However, it was surprising that the higher degree of content elaboration which led to better content selection and coherence did not translate to a higher degree of text appropriation in terms of rephrasing, nor did it lead to better cohesion. Linguistic elaboration remained a stumbling stone both in the individual and in the collaborative writing settings. A pedagogically sound use of Automated Writing Evaluation tools could be an interesting path towards fostering these specific strategies, as is elucidated in the directions for further research.

Samenvatting

Online pedagogiek biedt nieuwe kansen voor het hoger onderwijs door de flexibiliteit in tijd, ruimte en individueel leerritme die online technologieën creëren. Enerzijds wordt daardoor de toegang tot postsecundair afstandsonderwijs voor een breder publiek geopend. Anderzijds kunnen online technologieën ook tot vernieuwing in het onderwijslandschap bijdragen door andere vormen van feedback, interactie, en observatie van leerprocessen mogelijk te maken. Het is van wezenlijk belang dit potentieel ook voor taalstudies te ontginnen. Meestal wordt het leren van een taal aan menselijke interactie gekoppeld. Toch bieden de interactie tussen mens en computer, en de menselijke interactie via een computer een waaier aan bijkomende mogelijkheden om een taal (aan) te leren. Het doel van dit proefschrift is precies om dit potentieel van online technologieën voor het hogere vreemdetalenonderwijs te ontrafelen. De focus ligt hierbij op het academisch schrijven, meer bepaald op het samenvatten en synthetiseren van teksten. Het is voor taalstudenten belangrijk dit soort teksten te kunnen produceren, niet alleen met het oog op het schrijven van wetenschappelijke papers tijdens hun studies, maar ook voor toekomstige taken in hun latere beroepsleven. Dit proefschrift verbindt dus de wetenschappelijke disciplines CALL (computergesteunde taalverwerving) en schrijfpedagogiek.

Om het potentieel van online leeromgevingen voor het academische schrijfonderwijs te onderzoeken werd een reeks van vijf opeenvolgende en op elkaar afgestemde studies opgezet die methodologisch ingebed zijn in het zo genaamde *design-based research*. Twee voorbereidende observatiestudies op basis van de commerciële online leeromgeving DUO (Deutsch-Uni Online) leidden tot essentiële inzichten in het online leergedrag tijdens het uitvoeren van schrijfgerelateerde taken. Voortbouwend op deze inzichten werd online leermateriaal ontwikkeld voor drie interventiestudies waaraan in totaal 162 tweede- en derdejaarsstudenten Duits van twee Vlaamse academische opleidingen deelnamen. In deze interventiestudies die tussen twee en zes weken duurden kregen de studenten online begeleiding bij het aanleren van twee schrijftaken: een samenvatting van een luistertekst en een synthese van meerdere geschreven bronteksten. Deze taken vergen naast schrijfvaardigheid ook luister- en leesvaardigheid en vooral kritisch

denkvermogen. Dit laatste is nodig om de informatie van de bronteksten te doorgronden, de hoofdzaken eruit te selecteren, met elkaar in verband te brengen, en op een coherente manier in een nieuwe tekst te herschikken. Tijdens het schrijven van een dergelijk complexe tekst wordt het werkgeheugen zwaar belast omdat de schrijver voortdurend veranderende mentale representaties van de ontstaande tekst aan elkaar moet toetsen. Het is dus vanuit pedagogisch oogpunt cruciaal om deze cognitief zware taak zodanig te mediëren dat de studenten hun schrijfvaardigheid ten volle kunnen ontplooien.

De basisvraag van dit onderzoek is de vraag naar de rol die online technologieën in dit mediëringsproces kunnen spelen. Daartoe werden gerichte taken ontworpen op basis van verschillende pedagogische benaderingen waarbij diverse online technologieën toegepast werden. Op deze manier weerspiegelen de interventiestudies ook de historische ontwikkeling binnen CALL. De computer wordt enerzijds ingezet als "tutor" in instructivistische leeromgevingen met drill-oefeningen en in cognitief-constructivistische leeromgevingen die vooral het ontwikkelen van strategieën beogen. Anderzijds wordt de computer ook als "tool" gebruikt voor interactie in een sociaal-constructivistisch opgevat leerkader waarin samenwerkend leren en schrijven centraal staat.

De *Sociocultural Theory of Learning* vormt het theoretische kader voor dit proefschrift. Hierin werd het begrip *affordances* ontwikkeld dat verwijst naar de inherente eigenschappen van artefacten (hier: online tools) die een activiteit (hier: leren samenvatten/synthetiseren in Duits als vreemde taal) kunnen bevorderen of hinderen. Deze eigenschappen kunnen slechts dan een positief effect op het leerproces hebben als leerders de artefacten op de bedoelde wijze ervaren en gebruiken. Of dit in de interventiestudies het geval was, wordt in de data-analyse nagegaan. Daarvoor werd gebruik gemaakt van de *Activity Theory*, een theoretisch-methodologisch kader voor het onderzoek van complexe activiteiten, meer bepaald van het samenspel van de verschillende factoren die in een doelgerichte actie spelen. Dit samenspel bepaalt namelijk mee of de *affordances* van de artefacten die ter ondersteuning van de activiteit gebruikt worden al dan niet waargenomen worden. Daarom vormt *Activity Theory* een goede basis om het innovatieve gebruik van CALL technologieën in het onderwijs te onderzoeken. Met dit doel voor ogen werden de studenten tijdens de interventies nauwkeurig opgevolgd en werd er een brede waaier aan data verzameld die zowel kwantitatief als kwalitatief geanalyseerd werden. Pre-post-tests werden op basis van tekstkwaliteitskenmerken kwantitatief geanalyseerd om zo de impact van de interventies op de verbetering van de algemene schrijfvaardigheid van de deelnemers te meten. Opnames van het schrijfproces met behulp van *keystroke-logging* en *screen-capturing* software dienden om kwalitatieve veranderingen in het schrijfproces en in het gebruik van strategieën vast te stellen. Tenslotte werden deze observaties getoetst aan de resultaten van enquêtes en focusgroepgesprekken die de attitudes van de studenten

ten opzichte van de interventies en hun inschatting van hun eigen leerwinst in kaart brachten.

Het voorbereidende onderzoek toonde verschillende struikelblokken aan. Er waren twee strategieën die de studenten niet of nauwelijks toepasten, namelijk selectief luisteren en opstellen van een tekstschema voor het schrijven. Bovendien werd duidelijk dat taakinstructies in een online leeromgeving bijzondere aandacht moeten krijgen, en dat voorbereidende oefeningen die op talige middelen focussen een nauw logisch verband dienen te vertonen met de daarop volgende inhoudsgerichte vaardigheidsoefeningen. Met deze aandachtspunten werd rekening gehouden bij het opstellen van de oefeningenreeksen van de interventiestudies.

De kritische evaluatie van de interventiestudies wordt in het proefschrift afgesloten met een bespreking van de *affordances* van de artefacten (tools) die gebruikt werden om online activiteiten op basis van verschillende pedagogische principes te ontwikkelen. Het UGent-interne auteurstool *curios* diende om oefenreeksen te creëren vanuit zowel een instructivistische als een constructief-constructivistische benadering. Om samenwerkend leren en schrijven volgens het sociaal-constructivistische paradigma mogelijk te maken werden een leerpad en een forum binnen de universitaire leeromgeving en daarnaast ook *Google Docs* gebruikt.

Enerzijds werd vastgesteld dat deze tools inderdaad het potentieel bezitten om niet alleen individuele maar ook samenwerkende schrijfleeractiviteiten effectief te ondersteunen. Dit blijkt uit de kwalitatieve analyses van de schrijfprocessen en de enquêtes die een vooruitgang aantonen met betrekking tot het toepassen van strategieën en het inzicht in de relevante kwaliteitskenmerken van academische teksten. Anderzijds hadden deze positieve veranderingen geen kwantitatief meetbare, statistisch relevante neerslag in een betere algemene kwaliteit van de geschreven teksten. Dit is mogelijk te wijten aan het feit dat leren schrijven een langdurig proces is waarin veranderingen slechts langzaam zichtbaar worden. De analyse op basis van *Activity theoretical* principes toont verder aan dat de *affordances* van de artefacten enkel in pedagogisch weldoordachte leeromgevingen tot hun recht komen, d.w.z., als ze op een zinvolle manier met elkaar en met andere, *secundaire artefacten* gecombineerd worden, zoals een model of een stappenplan voor de schrijftaak. Maar ook andere factoren kunnen interfereren en zo belemmeringen van *fysieke* (vb. toegankelijkheid), *culturele* (vb. leerervaring), of *logische* aard (vb. programmeerfouten of onvoorziene interactie met de omkaderende technologieën), teweegbrengen.

De belangrijkste bron voor *contradicties* die de ontplooiing van *affordances* kunnen belemmeren blijken institutionele regels en individuele leerervaringen van de deelnemers te zijn. In een institutioneel kader dat sterk op individueel meetbare leereffecten gericht is, ontwikkelen de studenten een houding die eerder op leerresultaten (cijfers) dan op leerwinst (inzicht en kennisverbreding) gericht is. Nochtans sluit het laatste beter aan bij de essentie van computergesteunde activiteiten

die een zuivere *pattern-drill* aanpak overstijgen, zeker als er daarbij aandacht is voor samenwerkend leren. Deze tegenstrijdigheid weerspiegelt zich dan ook in de houding van de studenten tegenover het online leermateriaal. De belangrijkste conclusie in dit verband is het sterke contrast tussen de positieve attitudes van de studenten ten opzichte van voorgeprogrammeerde automatische feedback in gesloten taken die door instructivistische leeropvattingen geïnspireerd zijn, en de gemengde attitudes ten opzichte van feedback in de vorm van modeloplossingen en vragen voor zelfreflectie en -evaluatie. Bovendien werden de technologieën die *individuele* activiteiten ondersteunden duidelijk beter onthaald dan de technologieën die samenwerkende activiteiten mogelijk maakten. Hier moet wel een nuance worden toegevoegd: de laatstgenoemde kregen een betere respons als ze gecombineerd werden met *secundaire artefacten* zoals een model of stappenplan.

Een belangrijk resultaat dat het voordeel van samenwerken in de planningsfase van het academisch schrijfproces aantoont, is ten slotte dat de inhoud van de bronteksten in deze setting dieper doorgrond werd dan bij het individuele plannen. Verrassend genoeg vertaalt deze betere *inhoudelijke* verwerking zich niet naar een betere verwerking op *talig* vlak door bijvoorbeeld meer parafrasering en een beter gebruik van cohesiemiddelen. Dit bleef zowel tijdens het individuele als het samenwerkende schrijven een duidelijk struikelblok. Automatische tekstevaluatie, zinvol pedagogisch omkaderd, kan hier wellicht soelaas bieden, zoals in de afsluitende sectie over onderzoeksperspectieven wordt voorgesteld.

Acknowledgements

A PhD candidate has affordances that can only be taken up and lead to the envisaged outcome if the community of the overarching activity systems she is embedded in is supportive. Therefore, I am very grateful to my supportive community, and would like to take this opportunity to mention some of its members by name.

First and foremost, I would like to thank my supervisors and the members of my Doctoral Advisory Committee for their academic support throughout the process. Our annual meetings were always inspiring and encouraging. Geert, with your problem-solving mindset, you inspired me to carry out my first *blended learning* project at Ghent University which sparked my research interest in the area of Computer-Assisted Language Learning. I also learnt from you how to cope with reviewers' comments in a constructive way. Torsten, if it was not for your encouragement, I would not have embarked on this PhD project. Moreover, your reliability as an editor of abstracts with an amazingly short turnaround and 24/7 service is truly remarkable. When do you ever sleep? Tammy, I learnt a lot from your advice on methodology and research design. Pascale, you were always supportive and enthusiastic about my research. Last but not least, Mat: I am ever so glad I took the courage to ask you to join my Committee as an external member, after listening to your inspiring presentation at the EuroCALL conference in Nottingham in 2011. Although we have only met twice in person, it feels like we are long-time acquaintances thanks to the hours of skype meetings in which you patiently shared your expertise. Hurray to online technologies!

In terms of academic support, the research communities of EuroCALL and the EARLI Writing SIG need to be mentioned as well. The critical and encouraging discussions following my presentations at the conferences of these research communities, as well as fellow researchers' presentations have been an important source of inspiration. As it is impossible to list all the people who have contributed to this PhD with their suggestions and thoughts, I would just like to thank the research communities as a whole. Together with the teachers of summer schools and courses I attended, they all contributed to my immense learning curve over the past few years. Special thanks go to Vincent and Sven, teachers at the Linguapolis summer school, who did such a great job in conveying their

fascination for educational research design and data analysis. Sven, if it was not for you, I think I would never have ventured into statistics. I also owe the DUO team, headed by Prof. Jörg Roche at the Ludwig-Maximilians-Universität München, for granting me access to their online learning environment to carry out the preliminary studies.

Beside the academic activity system, the institutional activity system played an equally important role in the genesis of this PhD. I was lucky to join the Department of Translation, Interpreting and Communication for a research assistant position back in 2009, when it was still part of the University College of Ghent. All my colleagues here share a real passion for language and related, "applied" aspects, which makes for a great team spirit and a warm, welcoming working environment. Thank you all for the inspiring research meetings over lunch and the relaxing chat in our cosy staff lunchroom! The special spirit of our institution also rubs off on the students who were very supportive and patient research participants, eager - or at least not resistant - to engage in new and unexplored ways of learning, and very honest in their detailed feedback about their learning experiences. To all 162 participants of the intervention studies, plus the 12 volunteers in the preliminary studies, I owe special thanks. I also am indebted to the techn(olog)ical support staff of the department, the faculty, and the university. Thank you, Peter, Frederic, and the ICTO-team for your willingness to help find solutions to all sorts of problems that inevitably pop up when using technologies in educational interventions. Special thanks go to my direct colleagues in the German Section. Not only did we have lots of fun with our musical acts, at the "Stammtisch", and in occasional getaways, but you also rallied round to free me up from teaching duties in the last semester, which enabled me to dedicate my entire time to finishing this PhD. "Fröhlichste Abteilung", it has been a true privilege to work with you. Hinde, it is a pleasure to share our "doctorandae"-office with such a positive-minded colleague as you are. Thank you for the great time as we watched our "babies" grow :-).

According to the wise tenet "mens sana in corpore sano", the inner human being needs special nurturing in mentally challenging times. In this regard, my thanks go to my Spinning and Zumba teachers. Hannah, Leen, Stefanie, and Freek, thank you for your enthusiasm that kept me moving (on) and made me "sweat to forget", like the fitness club's slogan truly says. Likewise, the KASKcafé team deserves a special thanks for their great wholesome food that, among others, inspired stimulating lunchtime discussions about Second Language Acquisition theories with Kristof.

The tensions that come with a PhD endeavour can only be overcome in a loving and caring environment. The social-emotional activity system - if I might stretch this concept beyond its scientific limits - therefore deserves a special mention here. I would like to warmly thank all the people who surrounded me with their friendship and love over the past few years. Petra, Isola, Ellen, Greetje, Annick, Leen, Mia, and many others: thank you for sharing the good and the bad, the laughter, the singing, and for cheering me up with your enduring psychological support. I am also grateful to my parents who

always encouraged me to follow my vocation, believing in my capacities, and supporting me from a distance with tele-transported mantra's (Mama: "Finde deinen Faden!", Papa: "Wünsche hohen Wirkungsgrad!").

Last but not least, I want to thank my family who has been most directly involved in the process and suffered its consequences. Leo, I cannot imagine a better teammate than you for a PhD candidate: Not only because you have been through this yourself, and so you know all the tricks of the trade, but also for putting up with household and dad's duties in such a natural way, up to the point of taking paternal leave! Anouk and Simon, you had to miss me sometimes, and even if I was there physically, my mind was elsewhere. I am so proud of you coping with your "mentally absent mum" and grateful for not giving up on me. Watching you grow up is such a pleasure, and it helped me put things into perspective on many occasions over the last years. Thank you for being there, and for being you.

Preface

In the beginning there was ... a problem. How to teach practical lessons in German for specific purposes to a class of 75 economics students of all proficiency levels? The key to the solution was a Blended Learning project funded by the University's programme for innovative instructional practices (OPL). This was the third project in online language learning I was involved in, after participating in an e-mail tandem project between Portugal and Germany at Coimbra University and in the development of a Spanish online course for distance education at the Language Centre of Ghent University. The experience built up in these projects together with a passion for teaching and the affordances of technologies to support it shaped my research interest which has resulted in this PhD project. This dissertation reflects my five-year endeavour in the fascinating and fast evolving research field of Computer-Assisted Language Learning, with a specific focus on academic writing in a foreign language.

*Valeu a pena? -
Tudo vale a pena, se a alma não é pequena.
Fernando Pessoa - MAR PORTUGUÊS*

List of Abbreviations

AS	Activity system
AT	Activity theory
AWE	Automated writing evaluation
CAF	Complexity, accuracy, fluency
CALL	Computer-assisted language learning
CEF	Common European framework of reference for languages
CMC	Computer-mediated communication
CSCL	Computer-supported collaborative learning
CSCW	Computer-supported collaborative writing
DBR	Design-based research
DUO	Deutsch-Uni online
EAP	English for academic purposes
EFL	English as a foreign language
F2F	Face-to-face
GAL	Gesellschaft für angewandte Linguistik
GD	Google Docs
GTTR	Giraud's type-token ratio
HCI	Human-computer interaction
HOC	Higher-order concern
ICT	Information and communication technology
IS1	Intervention study 1
IS2	Intervention study 2
IS3	Intervention study 3
LOC	Lower-order concern
LRE	Language-related episode
LSA	Latent semantic analysis
LW	Learning to write
MPL	Mean period length
NLP	Natural language processing
OPL	Open parallele leerwegen

SCT	Sociocultural theory
SILL	Strategy inventory for language learning
SLA	Second language acquisition
TAP	Think-aloud protocol
UBR	Unique bigram ratio
VLE	Virtual learning environment
WLL	Writing to learn language
ZPD	Zone of proximal development

List of Tables

Table 1	Subprocesses for summarising/synthesising, mapped to the three writing stages of the writing process proposed by Flower and Hayes (1981)	18
Table 2	Characteristics of design-based research (Wang & Hannafin, 2005, p. 8)	36
Table 3	Design-based research cycle applied to the studies of this PhD	39
Table 4	Timeline of the studies carried out in the course of the PhD	40
Table 5	Cross-sectional design in Intervention Study 2	122
Table 6	Changes in CAF between pre- and post-tests, IS2 (means for 12 students)	148
Table 7	Changes in cohesive ties use between pre- and post-tests, IS2 (means for 12 students)	148
Table 8	Perceived usefulness of tools for individual and collaborative writing used in IS2	150
Table 9	Variables of quantitative comparative collaboration analysis	252
Table 10	Transformations of the research AS regarding the subject collective	286
Table 11	Transformations of the research AS regarding the subject	287
Table 12	Transformations of the research (and educational) AS regarding the object	289
Table 13	Outcomes of the educational AS' of all interventions	291
Table 14	Outcomes of the research AS' of all interventions	291
Table 15	Teacher presence and student responsibility in the three intervention studies	293
Table 16	Formal rules in the three intervention studies	295
Table 17	Mediating artefacts used in the educational activities	297
Table 18	Affordances of the technologies used in the intervention studies	300
Table 19	Six overarching research questions of this PhD with summarised answers	315

List of Figures

Figure 1	Writing model according to Flower and Hayes (1981)	14
Figure 2	Activity System according to Engeström 1987	27
Figure 3	The six-stage model of educational design (Kirschner et al., 2004)	37
Figure 4	Nested construct of affordances according to Blin et al. (2013)	45
Figure 5	A model of affordances for CALL, based on Kirschner et al.'s (2004) model of affordances for CSCL	46
Figure 6	Activity system of F2F summary teaching prior to IS1, including planned innovations (in italics)	85
Figure 7	Number of differences registered between student's own summary and model	111
Figure 8	Students' stance towards their own summaries regarding content	113
Figure 9	Students' stance towards their own summaries regarding structure	114
Figure 10	Students' stance towards their own summaries regarding language	114
Figure 11	System of the educational activity in IS1	119
Figure 12	Pre-post-test differences in individual writing test, IS2 (all students, n=35)	147
Figure 13	System of the educational activity of IS2	154
Figure 14	System of the research activity of IS2	155
Figure 15	Overview over whole study design and main research questions of IS3	161
Figure 16	Study design of the collaborative activity: crossed conditions	162
Figure 17	Jigsaw principle for the collaboration in triads	165
Figure 18	Workflow quality in IS3	254
Figure 19	Form of individual summaries produced in the model-first group. The numbers in the stacked bars indicate the number of individual summaries produced in the respective form.	255
Figure 20	Form of individual summaries produced in the script-first group	256
Figure 21	Form of collaborative outlines for syntheses in in model-first group. The numbers in the stacked bars indicate the number of collaborative outlines produced in the respective form.	256
Figure 22	Form of collaborative outlines for syntheses in in script-first group	257
Figure 23	Ferret similarity rates in IS3	258
Figure 24	Status of final product of collaboration sessions in model-first group. The numbers in the stacked bars indicate the number of final products in the respective form.	259
Figure 25	Status of final product of collaboration sessions in script-first group	259

Figure 26	Screenshot of an interaction in Google Docs from the model video	261
Figure 27	Overall interaction intensity in IS3	262
Figure 28	Number of chat turns according to focus in IS3	263
Figure 29	Number of comments according to focus in IS3	264
Figure 30	Text-related communication in IS3	265
Figure 31	Content-related communication in IS3	266
Figure 32	Attitudes towards usefulness of the scaffolding instruments	268
Figure 33	System of the research activity in IS3	272
Figure 34	System of the educational activity in IS3	273
Figure 35	An overview of the general research AS of this PhD	281
Figure 36	The research AS embedded in the network of connected AS', and the resulting external contradictions	282

Table of Contents

Introduction	1
"Writing is a complex process" - Why teach it online?	2
How to support academic writing development online	5
Research methodology and thesis outline.....	6
Chapter 1 A sociocultural approach to computer-assisted writing instruction: theoretical and methodological frameworks	9
1.1 Theoretical framework	10
1.1.1 A sociocultural approach to language learning.....	10
1.1.2 Second language writing development.....	12
1.1.3 Academic Writing in a Foreign Language	16
1.1.4 Fostering (academic) L2 writing development: Instructional challenges and potential CALL solutions	20
1.1.5 An Activity Theoretical Approach to Online Pedagogy for Academic Writing	26
1.2 Methodological Framework	33
1.2.1 The evolution of CALL design principles	34
1.2.2 The Educational Design Framework Based on Affordances.....	35
1.2.3 Evaluation of the intervention studies	40
Chapter 2 Preliminary explorative investigation of student-task-tool interaction.....	49
2.1 Preliminary Study 1: Usability of Deutsch-Uni Online.....	51
2.1.1 Background information on the study.....	51
2.1.2 Assessing QuADEM. Preliminary notes on a new method for evaluating online learning courseware (Article in <i>Computer-Assisted Language Learning</i>)	53
2.2 Preliminary study 2: (How) can an online module support students to write a summary from spoken input?	74
2.2.1 Background information on the study.....	74
2.2.2 Results	75
2.3 Conclusions from preliminary usability research	80

Chapter 3	Students' attitudes towards instructivist and constructivist feedback (Intervention Study 1)	83
3.1	Background information on the study	84
3.2	Attitudes towards online feedback on writing: Why students mistrust the learning potential of models (Article accepted for publication in <i>ReCALL</i>).....	88
3.3	Further results and conclusions from the study	110
3.3.1	Evidence of "noticing" in the constructivist part of the module	110
3.3.2	Insight into writing processes: How "planners" and "revisers" elaborate notes for summary writing	115
3.3.3	Attitudes towards PC use in classroom.....	116
3.3.4	Discussion of the study in the light of Activity Theory.....	119
Chapter 4	The effect of pre-programmed feedback and direct peer feedback on process and product (Intervention Study 2)	121
4.1	Background information on the study	122
4.2	Affordances of Web 2.0 technologies for collaborative advanced writing in a foreign language (Article in <i>CALICO Journal</i>).....	124
4.3	Further results and conclusions from the study	146
4.3.1	(Perceived) usefulness of the technologies to support tasks	146
4.3.2	Discussion of the study in the light of Activity Theory.....	154
Chapter 5	Collaborative online writing, supported by a script and a model (Intervention study 3)	159
5.1	Background information on the study	160
5.1.1	Overall study design.....	160
5.1.2	Design of the collaborative activity.....	162
5.2	Learning to think and write together: Collaborative synthesis writing supported by a script and a video-based model. (Chapter in: Guasch, T. & Deane, M. (Eds.), <i>Learning and teaching writing online: Strategies for success</i>)	166
5.3	Assessing the quality of computer-supported collaborative writing in a foreign language (Article submitted to <i>Journal of Writing Research</i>)	193
5.4	Learning to write syntheses: Changes in individual writing after collaboration (Article submitted to <i>Journal of Second Language Writing</i>)	222
5.5	Further results and conclusions from the study	252
5.5.1	Impact of the scaffolding condition on collaboration quality.....	252
5.5.2	Discussion of the study in the light of Activity Theory.....	272
Chapter 6	Discussion of results through an activity-theoretical lens.....	279
6.1	Contradictions involving the research AS	280
6.2	Summarising overview of the three intervention studies according to AS constituent.....	285
6.2.1	Subject collective.....	286
6.2.2	Subject.....	287
6.2.3	Object	288

6.2.4	Outcome	290
6.2.5	Division of labour	292
6.2.6	Rules	294
6.2.7	Mediating artefacts and their affordances	296
Conclusion		315
	Insights into online support in L2 academic writing instruction	315
	The added value of online technologies in the learning process	318
	Limitations of this thesis	321
	Directions for further research	323
Bibliography		325

Introduction

This is the paradox of science: its object is and is not there. (...) The problem in true research is that the researcher doesn't exactly know what he is looking for before he has found it. If he knew it at the beginning, nothing new would be discovered. (Engeström, 1987, p. 118)

An e-pedagogue is a teacher at heart, has a passion for learning, knowledge of learning theories and effective pedagogy, and understands the functionality of e-learning technology. (Mehanna, 2004, p. 290)

Consider the bifocal *work activity* (i.e., research and teaching) of a research assistant in a Higher Education institution in Belgium (*subject*). The *object* of her work is to scaffold students (*subject collective*) in developing their academic writing skills in German as a foreign language through *goal-oriented-activities*. *Instruments* used in this process include such *tools* and *artefacts* as the institution's VLE (Virtual Learning Environment), Google Docs, pedagogical scaffolding mechanisms, software for screen capturing, keystroke logging, and plagiarism detection, as well as popular-scientific source texts. These tools have specific *affordances*, i.e., inherent characteristics that can be enabled through instructional design, to facilitate the *activities*. The intended *outcome* of the *teaching activity* is the increase of the students' writing proficiency that becomes manifest in a better strategy use during the writing process and in a better product quality. In the long term, this should help them to write bachelor and master theses in the foreign language and also prepare them for a future work situation where summarising and/or technological skills may be beneficial. Perhaps they even discover that academic writing can be an enjoyable activity.

The use of new technologies and teaching methodologies as artefacts to support this activity might prove to be a valuable innovation. It is the object of the *research activity* to find out whether this is really the case. By closely monitoring all *constituents* of the activity system, possible *contradictions* can be uncovered that hamper the uptake of the tools' affordances in the activity. For instance, a new technology or a new pedagogical

approach might have a daunting effect on students which could lead them to be dissatisfied with the learning activity or their own perceived progress.

The *community* in which the research activity is embedded consists of the institution's colleagues on the one hand, supportive or not of innovative teaching methods, and of the research communities she meets and is inspired by, on the other hand. *Division of labour* determines the tasks and decision-making powers of the research assistant, of the members of her doctoral advisory committee, of her direct colleagues in the German section, and the Head of Department. Finally, the *rules* regulate such a broad array of features as a six-year time limit for the research and teaching position, institutional regulations for intervention studies regarding timetable, curriculum, experimental design, as well as measurement of learning outcomes, and last but not least, the research standards in terms of validity, reliability, and generalizability.

This *activity system*¹ describes in a nutshell the author's work activity during the last five and a half years. At the same time, it is the main activity system serving as a background against which the intervention studies carried out during my PhD research will be interpreted in this thesis. The writing tasks carried out during these studies by the students form goal-directed individual and group actions that are activity systems in their own right as well as subordinate units of analysis of the overarching research activity system. AT (Activity Theory) is an abstract theoretical-methodological framework that allows to capture the complex interplay of environmental factors in a goal-oriented action that (fail to) enable the affordances of a tool used in and for the action. It has therefore been implemented by CALL (Computer-Assisted Language Learning) and Writing researchers alike, which turns it into a perfect framework for the present PhD that is situated at the crossroads of these two academic disciplines.

"Writing is a complex process" - Why teach it online?

"Writing is complex" is a statement often heard at writing conferences. Indeed, this productive skill does not only require language knowledge, but also involves several cognitive and psycholinguistic activities that have to occur simultaneously and that are difficult to track, to research, and to teach. As Flower and Hayes put it succinctly:

¹ This description is based on two examples, viz. the description of a physician's work activity at a primary care clinic provided on the website of the *Center for Activity Theory and Developmental Work Research* at Helsinki University, <http://www.edu.helsinki.fi/activity/pages/chatanddwr/activitysystem/>, and the adaptation of Blin (Blin, 2005, p. 46) to describe the work activity of a language teacher working in an Irish secondary school.

The writer must exercise a number of skills and meet a number of demands - more or less all at once. As a dynamic process, writing is the act of dealing with an excessive number of simultaneous demands or constraints. Viewed this way, a writer in the act is a thinker on full-time cognitive overload. (Flower & Hayes, 1980, p. 33)

This is even more the case for "writing from sources", an important genre mostly, but not exclusively, used in academic writing. There are basically two different forms of this genre, viz. summary (or *précis*) writing, i.e., producing a text that is substantially shorter than the (single) source text, but covers the same information as the latter, and synthesis writing, i.e., producing a text based on multiple sources that connects the main ideas of those sources based on a critical comparison and selection of their content (Spivey, 1990). Both forms were *objects* of the intervention studies presented in this dissertation. Therefore, "academic writing" is used in this PhD as an umbrella term for these two subgenres. Beside writing abilities, they require extensive reading (writing from written sources) or listening (writing from spoken sources) skills as well as critical thinking abilities. From a pedagogical point of view, mediating this "overwhelming cognitive overload" in writing from sources is important to maximize student performance (Kirkland & Saunders, 1991).

Summary and synthesis writing are intriguing objects for research, because as instructional tasks, they are situated at the intersection between "learning-to-write" and "writing-to-learn". Regarding the latter aspect, summarisation, which is needed in both subgenres, is known to be a main cognitive strategy for understanding and learning. It has been singled out as a highly efficient strategy in classroom education and e-education by Marzano (1998) and Mehanna (2004) in their meta-analyses on the effectiveness of instructional strategies. At the same time, summarisation stands out as highly efficient strategy in learning-to-write, as Graham and Perin have shown in their meta-analysis of writing instruction for adolescent students in their native language (Graham & Perin, 2007a). Curiously enough, although its high value for learning has been recognised, summarisation is seldom explicitly taught in higher education (Doolittle, Hicks, & Triplett, 2006). Students are expected to know how to summarise when they enter university. The lack of summarising strategies becomes an overt problem at the moment of producing bachelor or master theses. Students *patchwrite* their literature reviews, which reflects a lack of elaboration and causes plagiarism (Pecorari, 2013).

Therefore, it is with a good reason that summary writing both in L1 (first language) and in L2 (second language) is integrated as a separate subject in the Applied Languages curriculum at UGent (Ghent University). In this context, it is understood as a "learning-to-write" task rather than a "writing-to-learn" strategy. All too often, though, classroom instruction in L2 summary writing is restricted to a strong "focus on forms" (Long, 1983), i.e., error correction on a grammatical and lexical level, the so-called

"lower-order concerns" (LOCs) in writing. However, the quality of academic writing depends to a large degree on "higher-order concerns" (HOCs) like content selection, elaboration, and coherence building. It is therefore important to investigate how the focus on these text-level concerns can be trained, while problems on word- and sentence level should continue to be addressed (Dodigovic, 2002).

Traditional classroom-based instructional approaches face a severe "teachability"² limit when it comes to academic writing subgenres. As there are different knowledge areas, skills, abilities, and strategies involved, it is impossible to tackle the broad array of individual problems in a one-size-fits-all approach. On the other hand, individual feedback on writing is highly time-consuming and not realistic considering the student-teacher ratio in tertiary education. Furthermore, the focus of writing pedagogy should not be limited to the product, i.e., the final text, but address the writing process in all its phases, i.e. planning, translating, and revising (Hayes & Flower, 1980). The question is, to what extent can the use of online technologies contribute to mediating the cognitive processes in L2 academic writing?

Online pedagogy affords tertiary education fresh scope. The integration of ICTs in university curricula facilitates flexibility in terms of time, space, and individual learning pace. On the one hand, this allows access to post-secondary education for a broader public through open and distance courses. On the other hand, online technologies also open up new instructional perspectives in that they provide different opportunities for feedback, interaction, and process tracking.

These opportunities also need to be exploited for language learning. While language learning is primarily identified with face-to-face human interaction, human-computer interaction and human interaction facilitated by online technology provide a broad array of additional opportunities for language instruction and uptake. Furthermore, "[l]anguage learners are entering a world in which their communicative competence will include electronic literacies, i.e. communication in registers associated with electronic communication" (Chapelle, 2001, p. 2). Therefore, ICT skills need to be integrated as transferable skills in language learning curricula, as Leakey (2011) points out:

Clearly our students welcome the appropriate incorporation of technology into their learning, and are motivated by those platforms, programs and pedagogies that they feel enhance their learning experience, if not their learning outcomes. They are increasingly familiar with a wide range of technologies in their everyday

² I use this term in its broader notion of "ability to be taught", not in the sense of Pienemann's *teachability hypothesis* (Pienemann, 1984) that builds upon the notions of interlanguage and fixed developmental patterns for language acquisition

life, and a curriculum that made no use of technology would potentially deny them essential skills for life (Leakey, 2011, p. 288).

How to support academic writing development online

This PhD project is situated at the crossroads between two major research areas, CALL and Second Language Writing. Research in CALL has highlighted the benefits of technology use for various domains of language learning, especially for knowledge building (vocabulary, grammar), and for skills development at lower proficiency levels (see for instance the most recent meta-analysis of CALL effectiveness studies of Grgurović, Chapelle, & Shelley, 2013). However, research on *advanced* language learning using technologies to date is scarce (Grgurović et al., 2013; Thorne & Reinhardt, 2008). This PhD contributes to our understanding of the affordances that online technologies provide for academic writing, more specifically, writing from sources, exploring different settings and tools, based on different design principles. In this regard, it represents the "post-methodological" multiplicity of language teaching approaches stated by Kumaravadivelu (2001), not favouring a specific approach, but combining the features of different approaches that best suit the task and target group at hand, based on the principles of "particularity, practicality and possibility" (Kumaravadivelu, 2001, 2006). The interventions of this PhD therefore reflect the development in CALL from using the computer-as-tutor (Levy, 1997) in instructivist exercises including pattern-drill and in cognitive constructivist exercises targeting strategy development towards the use of the computer-as-tool in a socio-constructivist approach to collaborative writing.

One could ask whether it makes sense to distinguish between pedagogy with and without online technologies in the classroom nowadays, as ICT (Information and Communication Technology) is embedded in our daily lives. Indeed, CALL-scholars have repeatedly alerted to the fact that it is not the technology that makes the difference, but its pedagogical use (cf. e.g. Pederson, 1987; Warschauer, 1996), and that CALL effectiveness research should be informed by the theories of SLA research and their implications for instruction (Chapelle, 2001, 2009). However, in higher education, especially in the humanities, the use of online technologies for educational purposes is far from reaching the stage of "Integrative CALL" postulated by Bax as early as 2002, where "computers (...) are used every day by language students and teachers as an integral part of every lesson, (...) without fear or inhibition, and equally without an exaggerated respect for what they can do" (Bax, 2003, p. 24). Thus, motivational variables still play a role and have to be taken into account in CALL effectiveness

research. In addition, *pedagogy* is only one of the "three P's" addressed by CALL research, as it is embedded in a *platform* and in a *program*, which distinguishes CALL from general SLA pedagogy. As Egbert (2005) succinctly puts it in her introduction on CALL research methodology: "computer tools, particularly Internet support for computer-mediated communication (CMC), give us different opportunities [for instructed language acquisition] than afforded by other tools, and we need to approach them as something that we do not currently understand" (Egbert, 2005, p. 5).

The main research question of this dissertation is therefore not *whether* online technologies should be used to teach advanced writing skills in higher education, but *how* this can be done in an effective way. As Felix states:

There exists a clear trend away from the comparison studies (...) that wanted to find out whether teaching with computers was better than teaching without them. (...) The most obvious reason (...) is that in an environment where computers have become a natural part of the educational environment (...) the question is no longer as interesting. What remains interesting to investigate is how technologies are impacting learning processes and as a consequence might improve learning outcomes. (Felix, 2005, p. 16)

To answer this question, several task designs were developed according to different pedagogical approaches, using an array of different online technologies. Together, the studies conducted in the course of this PhD can be regarded as a design-based research cycle, as they build upon each other, taking into account the outcome of the previous study each time.

Research methodology and thesis outline

In order to explore the effectiveness of the different pedagogies, platforms, and programs used in the intervention studies for writing development, a socio-cultural theoretical approach is combined with a design-based research methodology. The concept of *affordances* used in the title stems from SCT (socio-cultural theory). It refers to the inherent characteristics of artefacts (here: online tools) that facilitate or impede an activity (here: learning to write from sources in German L2). Those characteristics only have an impact on learning when learners actually perceive and use the artefacts in the intended way. In other words, the uptake of the affordances by the learners plays an important role. Van Lier (Van Lier, 2000) adopted the notion of *affordances* in his *ecological* approach to language learning, in an attempt to shift the attention from *input* to *engagement*:

Learning is not a holus-bolus or piecemeal migration of meanings to the inside of the learner's head, but rather the development of increasingly effective ways of dealing with the world and its meanings. Therefore, to look for learning is to look at the active learner in her environment, not at the contents of her brain (Van Lier, 2000, pp. 246-247)

Consistent with this view on learning, the implementation of online learning environments was carefully monitored in all three intervention studies carried out during the action research period (October 2010 - March 2013). A mixed-methods approach was adopted, i.e., the rich data collected were analysed using both quantitative and qualitative inquiry methods. The quantitatively explored data range from pre-post-tests and measurements for product assessment to questionnaires on learners' individual characteristics like learning style, strategy use, ICT background, and attitudes towards the interventions. Qualitative data analysis mainly served to get a grasp of the complex interactions between learners, tasks, rules, and roles that contribute to their attitude, and ultimately condition their learning. The data sources used for this aim comprise attitude questionnaires as well as post-hoc focus groups.

Chapter 1 presents the theoretical and methodological framework for the empirical research carried out in this PhD. The sociocultural approach to language learning was adopted as a general theoretical underpinning. First, an overview of this approach is presented, with a specific focus on (academic) writing development. Then, the potential of computer-assisted language learning for L2 academic writing instruction is explored from a sociocultural view. Finally, AT (Activity Theory) is presented as a theoretical framework to discuss innovation through CALL tools in instructional settings. The methodological section of the chapter is divided into two parts. In the first part, general design principles for CALL, and principles of DBR (design-based research) are discussed. Subsequently, the DBR cycle overarching the empirical studies carried out in this PhD is elucidated, and the research questions are formulated accordingly. The second part of the methodological section zooms in on the evaluation of CALL interventions. It starts with a short overview of CALL evaluation frameworks within the cognitive paradigm, and continues by outlining the advantages of the AT-based approach to evaluate the affordances of the tools used in the intervention studies and their uptake in the respective activities.

Chapters 2-5 present the empirical research in chronological order. In Chapter 2, the two preliminary studies designed to observe actual students' behaviour in online learning environments are presented. The three subsequent chapters (3-5) are dedicated to the description and discussion of the three intervention studies, each of them implementing a different (combination of) instructional approach(es) for the design of learning material, ranging from instructivism over (moderate) cognitive constructivism to social-constructivism. Chapters 2 through 5 are built around articles or book chapters that resulted from the respective studies and have already been

published (2.1.2, 4.2, 5.2), accepted (3.2), or submitted (5.3 and 5.4) for publication. As those publications in each case only focus on a specific aspect of the respective intervention, I resolved to embed them in a broader chapter. In the first subchapter, in each case a general introduction to the study is provided. The following subchapter(s) contain(s) the publication(s) that resulted from the study. In order to maintain the internal consistency of the publications, they were integrated in their entirety, only adapting the font. This means that the figures, tables, bibliographic references, and appendices of these chapters were left as in the original, not integrating them in the main document. Hereafter, other important results from the study are presented each time. Finally, chapters 3-5 conclude with an evaluation of the respective intervention studies based on the activity-theoretical framework, clarifying the tensions that led to changes in the instructional design of the subsequent study.

Chapter 6 combines the theoretical framework with the empirical findings of the intervention studies, providing a critical summary of the instructional designs and their evolution based on the constituents of the activity-theoretical model. To conclude, the main findings are summarised based on the research questions. In addition, the limitations of the research carried out in the context of this PhD are discussed, and a promising area for future research is outlined.

Chapter 1

A sociocultural approach to computer-assisted writing instruction: theoretical and methodological frameworks

The question is not what specific educational techniques (...) cause. It is rather what they actually afford, which can be referred to as educational affordances of a learning environment (Kirschner et al., 2004, pp. 11-12). [my emphasis]

1.1 Theoretical framework

In the following, the theoretical underpinnings that have driven the research carried out in this PhD are presented, taking a sociocultural approach as overarching framework. In the first subchapter, key elements of this approach to language learning are outlined. The following two subchapters provide an overview of the state-of-the-art in the literature concerning (L2) (academic) writing development from a sociocultural viewpoint. Then, the possible contributions of online pedagogy to the development of academic writing are discussed. Finally, I zoom in on Activity Theory, which is a key concept of the sociocultural theory used in this PhD as evaluation framework. In order to avoid overlap with the literature reviews on specific research areas that are presented in the included articles, the reader is referred to the respective sections where appropriate.

1.1.1 A sociocultural approach to language learning

Johnson (2006, p. 235) described a sociocultural turn in language learning and teaching at the end of the last century. The SCT (sociocultural theory) of knowledge development is rooted in the conviction that higher order mental functioning is mediated by cultural and social factors. Vygotsky (1978), who laid the foundations for SCT with his cultural-historical activity theory, argued that traditional approaches used in psychological research of elementary mental functions, viz. introspection and reaction-time experiments, failed to grasp higher, culturally constructed forms of thinking. Therefore he proposed the so-called "genetic method", emphasising the importance of investigating the learning *process* (instead of solely the outcome), as a result of the interaction of the individual (learner) with all environmental factors and tools mediating learning.

Lantolf (2000), who argued for SCT as a theoretical approach to L2 acquisition labelled this approach *Sociocultural SLA*. This theory differs fundamentally in its perspective from cognitive and linguistic approaches to SLA in that it broadens the focus of analysis from product (i.e., L2 output) towards process (i.e., how this output is shaped in activities). While it has not been accepted without controversy (Ellis, 2008, pp. 547-554), it has certainly contributed to our understanding of instructed language learning (Foster & Ohta, 2005). During the last decade, the sociocultural approach has increasingly been adopted by CALL researchers, especially in the areas of learner autonomy (Blin, 2004, 2005) and (intercultural) CMC (Computer-Mediated Communication) (Blin & Appel, 2011; Thorne, 1999, 2003). In the following, a brief account of the key concepts of SCT and their adoption in SLA and CALL research is presented.

Mediated learning applied to SLA means that an L2 learner engaging in a language related activity needs to resort to an *artefact* or *tool* (both concepts are used interchangeably) in order to successfully reach the *object* of his or her activity. Artefacts can be either concrete (e.g., a dictionary, an communicative online tool, a feedback prompt) or symbolic (e.g., the L2 itself, the L1, a learning strategy), either external (provided by another person than the learner) or internal (i.e., a skill or a strategy to which the learner can resort). As CALL by definition employs technological tools to support the learning activity, the SCT framework provides a fruitful theoretical basis to investigate the mediating role of these artefacts in learning activities (Warschauer, 2005). Warschauer (2000) revealed how different ICT tools afford the appropriate actions in different kinds of language learning activities, depending on institutional contexts and on teachers' and students' beliefs on technology use and its language learning potential. He stated that one of the strengths of adopting SCT is that it

allows us to dialectically link (...) seemingly contradictory perspectives. Yes, technology is just a tool, but, like all tools, it mediates and transforms human activity. Both teachers and researchers need to take into account both how this mediation occurs at the micro level and also how it intersects with, and contributes to, broader social, cultural, historical, and economic trends. By applying the lens of sociocultural theory, we can begin to tackle that challenge (Warschauer, 2005, p. 48).

Language learning from an SCT point of view is primarily based on dialogue. This dialogue can be either external (i.e., human-human or human-computer-interaction), or internal (the so-called "inner" or "private" speech). An advanced L2 learner in a sociocultural sense is someone whose inner speech in the L2 takes over the role of a mediating tool to regulate the own learning and thinking activity in that L2 (Ellis, 2008, p. 525). In order to reach this stage, the learner has to engage in social interaction that is beneficial to language learning. SCT has defined this kind of interaction based on two central concepts: (a) the ZPD (zone of proximal development), i.e., the distance between "the actual developmental level" and "a level of potential development as evidenced in problem solving undertaken with the assistance of (...) an expert or through collaboration with peers" (Ellis, 2008, p. 532) and (b) *scaffolding*, i.e., "the dialogic process by which one speaker assists another in performing a function that he or she cannot perform alone" (Ellis, 2008, p. 527). Scaffolding can take the form of teacher instruction or the guiding assistance of an advanced peer (Driscoll, 2000, p. 251). Even the interaction between peers that are equally proficient in the skills required to achieve the goal of the learning activity can lead to mutual scaffolding (Swain, Brooks, & Tocalli-Beller, 2002). According to Ellis (2008, pp. 527-528), in recent SLA literature the term *scaffolding* is being avoided for human-human interaction due to its association with an object, giving way to the terms *collaborative dialogue* in the case of peer

interaction and *instructional conversation* in the case of expert-novice interaction. In tutorial CALL, scaffolding can also occur via HCI (Human-Computer Interaction), that is, the computer takes over the role of the expert by providing the learner with assistance in the process of shaping L2 output.

SLA and CALL studies have drawn upon these basic concepts of SCT as a theoretical underpinning for research on (a) the effectiveness of feedback, (b) the benefits of collaborative activities, including the use of *metatalk* (both about the task and about L2 language problems) for individual learning, (c) the role of private speech for L2 acquisition, and (d) actual learners' behaviour when engaging in an L2 task, distinguishing between the *task* (i.e., a pre-defined set of actions leading to a pre-defined learning goal) and *activity* (i.e., what learners actually do when engaging in the task, which might lead to different goals and outcomes than the ones originally intended by the teacher who set up the task). For the research presented in this PhD, scholarship in (a), (b), and (d) is relevant and will be discussed in the respective subsections of this chapter.

1.1.2 Second language writing development

[T]he development of writing abilities in second/foreign/additional languages (L2) is a topic of considerable contemporary currency, not least because writing and related literacies play such an important role in how societies function and, by extension, the life prospects of individuals within them. (Norris & Manchón, 2012, p. 221)

What is L2³ writing development, and how can we measure it? In order to define the impact of a writing intervention, these questions first have to be addressed. Norris and Manchón (2012) distinguished between two fundamentally different approaches: L2 writing development can either be defined primarily as linguistic development or as gain in control over textual output (Norris & Manchón, 2012, p. 224). This difference is also reflected in the basic distinction between WLL (writing to learn language), i.e. "the way in which the engagement with L2 writing tasks (...) can contribute to development

³ Terminological note: In this PhD, "second language" and the abbreviation L2 are used as an umbrella term. Actually, the interventions took place in a foreign language learning environment, which strictly speaking differs from second language learning in that the latter refers to a situation where the learner lives in a country in which the L2 is spoken. This distinction is especially important in the US context with its large population of second language learners in secondary and higher education. However, in the European context, L2 is mostly used as an umbrella term for both learning situations (De Bot, Lowie, & Verspoor, 2005), and for *all* languages acquired or learnt after the first one, which is also referred to as "mother tongue".

in (...) language knowledge and skills" (Manchón, 2011, p. 3), and LW (learning to write), i.e. "the manner in which second and foreign (L2) users learn to express themselves in writing" (Manchón, 2011, p. 3). Both perspectives add to a comprehensive theory about L2 writing scholarship, albeit from a different theoretical background and focus. While WLL is mainly being investigated in foreign language learning contexts and typically makes use of cognitive and sociocultural theories of SLA, the LW approach is more frequently used in genre-based writing of L1 and second language learning contexts (Manchón, 2011, p. 3). However, in university foreign language degree programmes, both approaches are "inseparable due to educational and linguistic reasons" (Manchón, 2011, p. 5). Studies in this area therefore can contribute to investigating the close interaction between the two approaches, attending to the "claim regarding the scarcity of studies on complex, meaning-making composition writing tasks in the research on the WLL dimension of writing" (Manchón, 2011, p. 11).

Through a product-oriented linguistic lens, "[p]rogress in acquiring the second language system is seen as manifested by increased fluency and accuracy, and a wider range of syntactic structures, as these reflect expanding knowledge that the learner can draw on automatically" (Foster & Ohta, 2005, p. 403). In writing, this progress in knowledge and its online retrieval manifests itself among others in higher values for CAF (complexity, accuracy, fluency). Hence, the majority of studies on L2 WLL development include quantitative measures for these text features (Housen & Kuiken, 2009; Lambert & Kormos, 2014; Pallotti, 2009; Wolfe-Quintero, Inagaki, & Kim, 1998). In their comprehensive literature review on CAF measurement, Wolfe-Quintero, Inagaki, and Kim (1998) showcased the broad array of different measures for each of these three text quality dimensions.

This product-oriented approach differs fundamentally from a sociocultural approach where "[d]evelopment is visible through microgenetic analyses of episodes of interaction [with other people and with artefacts], as the learner demonstrates increased independence" (Foster & Ohta, 2005, p. 403). As recent technologies allow for more insight into writing processes, this approach is gaining momentum in writing research:

The major methodological advances of recent years have changed, re-focused and significantly influenced writing process research across Europe. The development of modern writing process tools on the one hand and the acquisition of qualitative and quantitative methods originating from neighboring disciplines, adapted to suit the needs of writing process research on the other, have brought about new techniques to track, investigate and value writing processes and products. Key logging, screen capturing, eye tracking and other process tools, as well as accompanying possibilities to obtain, measure, analyze, compare, assess and, as a consequence, teach, enrich the field of writing process research (Grabowski, Becker-Mrotzek, Knopp, Jost, & Weinzierl, 2014, p. 7).

The first cognitive models of the writing process departed from an L1 context. However, the process approach to writing which was propagated by these models also impacted our understanding of L2 writing development and, accordingly, L2 writing instruction (Galbraith, 2009). Flower and Hayes (1981) laid a theoretical basis for writing process research with their cognitive model of the writing process, emphasizing its exploratory and generative nature (see Figure 1).

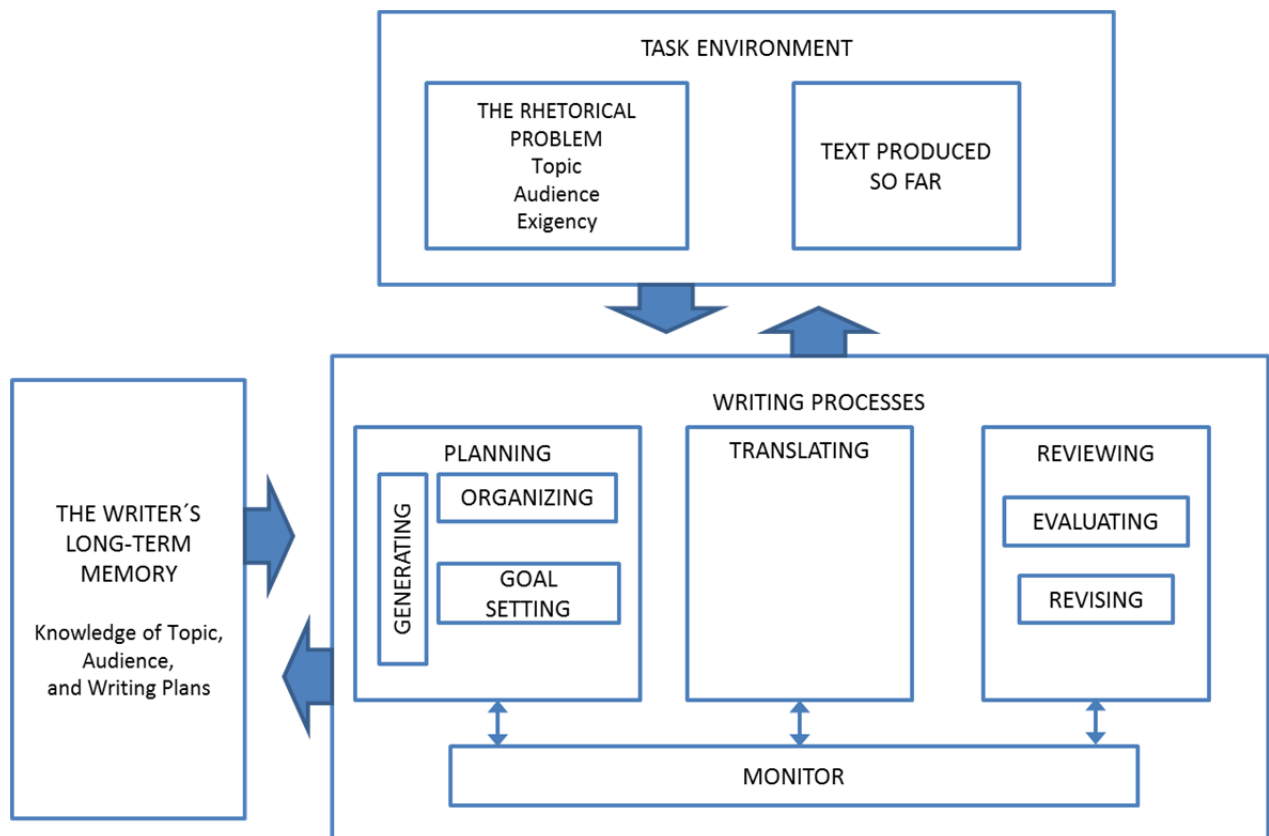


Figure 1 Writing model according to Flower and Hayes (1981)

This model is based on the notion of writing as a three-stage cognitive process, i.e., planning, translating (i.e., transposing the generated and organised ideas and the writing goals to a written text output), and reviewing, which implies evaluating the text produced so far in the light of the intended goals or the task, and revising it accordingly. It is important to note that these stages do not necessarily follow each other in a linear way, but rather are accessed in a non-linear, recursive way by the controlling monitor that regulates the moves between the stages.

One of the central premises of the cognitive process theory presented here is that writers are constantly, instant by instant, orchestrating a battery of cognitive processes as they integrate planning, remembering, writing, and rereading. The multiple arrows (...) are unfortunately only weak indications of the complex and active organization of thinking processes which our work attempts to model. (Flower & Hayes, 1981, p. 387)

Flower and Hayes attributed an important role to the writer's goal setting that highly influences the monitoring process during composition. They found that skilled writers have a more developed network of goals than novice writers, suggesting that goal-setting and the necessary strategies to achieve them should be addressed in writing instruction. Two external factors that influence the cognitive writing process, and vice-versa, are influenced by it, are the writer's long-term memory for knowledge retrieval and the writing task, or, in more general terms, the rhetorical problem that triggers the writing action.

Bereiter and Scardamalia (1987) furthered the understanding of cognitive writing processes by introducing a basic distinction between the processes used by novices, called knowledge-telling model, and those used by skilled writers, called knowledge-transforming model. The former resembles the linear processes used in spontaneous speaking and does not involve a large amount of planning and goal-setting, whereas the latter is needed to produce texts of a more complex nature which require reflective problem-solving thinking, i.e., the generation of goals while writing and the evaluation of the emerging text in the light of those goals. In their model of knowledge-transforming writing, they distinguish between two different sources of knowledge, viz. content knowledge and discourse knowledge, which are both involved in the translation process. Kellogg (2008) expanded the Bereiter and Scardamalia model by adding a third stage he called knowledge-crafting (see section 1.1.3).

Both the Flower and Hayes and the Bereiter and Scardamalia models are writer-centred and depict writing as a more or less solitary act. The first model with an explicit reference to the communicative context is that of Grabe and Kaplan (1996). The authors attributed a new function to goal setting, in that it "generates lenses through which the writer attempts to match external 'context' with internal resources" (Grabe & Kaplan, 1996, p. 227). What makes Grabe and Kaplan's model especially interesting for L2 writing instruction, is that it also includes *language competence* as a component, which is subdivided into linguistic, sociolinguistic, and discourse competences.

A distinctive characteristic of all cognitive models of writing is that they understand the process as being mainly internal, assuming a dichotomous relationship between the writer's cognition and the writing context which is taken into account to a rather limited extent (Lei, 2009). An important change in the understanding of writing occurred with the sociocultural turn in the 1980s. L2 Writing from an SCT standpoint is viewed as inherently social and interactive in nature in that

attention to writers and their products is embedded in local, institutional, and disciplinary contexts, and (...) people, their goals, and their relationship with context matter as much as do grammar, syntax, and discourse, drafting and revising (Lei, 2009, p. 27)

In adopting an SCT approach to writing, intertextuality and reader-orientation became important aspects (Nystrand, 2006). SCT is based on the principle that writing, like all human actions, is socially situated, and therefore, writers are adhering to social conventions. These conventions can be defined by classroom, institutional or cultural settings. Accordingly, the sociocultural turn in writing research also had repercussions on the variables included in the investigation of writing development. Not only has the role of writing strategies been re-assessed in the light of SCT (Donato & MacCormick, 1994; Lei, 2009), but also data collection was expanded including data on writer's goals, motivations, self-efficacy beliefs, and their response to educational affordances and developmental targets in order to grasp the complex and dynamic interplay of individual and social factors:

[W]e have begun to gain insights into the probably major role played by learner volition, goals attitudes and other individual differences, any of which may override the other dimensions of language and writing development (or underpin the lack thereof). These more finely-tuned foci should play a serious part in future avenues of investigation if our objective is to understand real learners and to witness the actual unfolding of L2 writing competence (Norris & Manchón, 2012, p. 241).

Norris and Manchón emphasized the need to investigate students' and teachers' beliefs by relating their experience with an EAP (English for academic purposes) intervention research where "only by looking at what teachers and students believed and did (...) the different pieces of the puzzle were found and could be made to fit" (Norris & Manchón, 2012, p. 238).

In the last decade, an activity theoretical approach has increasingly been adopted to capture the interrelatedness of environmental context factors shaping the writing process, following Hayes' suggestion (Hayes, 2006) (see detailed description of Activity Theory in section 1.1.5). Especially, though not exclusively, accounts of the use of electronic media in writing instruction have been approached through an activity-theoretical lens to investigate, among others, intercultural communities-of-use (Basharina, 2007; Thorne, 1999, 2003), the development of learner autonomy (Blin, 2004; Blin & Appel, 2011), and writer's goals or motivation and their impact on text production (Sasaki, 2009; Yasuda, 2005).

1.1.3 Academic Writing in a Foreign Language

The development of academic writing has been studied from at least four different theoretical-methodological perspectives, viz. the cognitive, the socio-cognitive, the sociocultural, and the "socially-shared" paradigm (Castelló, Bañales, & Vega, 2010). The sociocultural paradigm has given rise to a genre-based approach to writing which plays

an important role in research and instruction of L2 writing in academic contexts (Swales, 1990). Strategy development has been recognised as a key element in academic writing pedagogy (Graham, 2006). According to writing development theories, the genres of academic writing belong to the third and latest evolutionary stage of writing called *knowledge crafting* (Kellogg, 2008) that requires cognitive processes of *knowledge constituting* (Galbraith, 2009), as opposed to *knowledge telling* (first stage) and *knowledge transforming* (second stage). Knowledge crafting implies a threefold mental representation of the emerging text in the writer's working memory, being that of the author's intention, that of the actual text, and that of the reader's interpretation. Allocating attention to these three representations simultaneously puts a high strain on working memory capacity. Therefore, other cognitive processes and strategies like planning and sentence generation need to be automated in order to overcome processing constraints (Torrance & Galbraith, 2006). In an L2 writing context, working memory capacity is additionally strained by lexical and linguistic retrieval. Drawing on Chenoweth & Hayes (2001), Galbraith (2009) stated that using an L2 has repercussions on writing fluency as measured in production burst length. Reporting on experiments that connect production bursts with the on-line generation of ideas, he theoretically concluded that writing in an L2 alters the development of the writer's understanding, because "writing is thinking" (Galbraith, 2009, p. 20). Nevertheless, concerning the specific task of summary writing, Wolfersberger (2007, p. 55) concluded from a short overview of related research that "L1 and L2 summary writing may be very similar".

One of the most distinctive features of academic writing is that it is "heavily intertextual" (Pecorari, 2013), in that the writers have to incorporate external textual sources into their own text. Failure to elaborate and appropriate the textual sources in integrating their content into one's own coherent flow of thoughts leads to excessive textual borrowing which can result in plagiarism (Keck, 2014). Another strategy frequently observed in novice writers while producing academic texts is called *patch-writing*. Drawing on Howard et al. (2010), Pecorari defined patch-writing as a form of textual borrowing that is situated between paraphrase and quotation, "stitching together elements from one text with elements from another and making some superficial changes to the language" (Pecorari, 2013, pp. 70-71). Due to their limited linguistic resources as compared to L1 writers, L2 writers have more difficulties to rephrase ideas from textual sources. Textual misappropriation in L2 writing using sources has been the subject of a lively scholarly debate. The Journal of Second Language has devoted a special issue to this matter (Vol 21, Issue 2: "Textual appropriation and source use in L2 writing"). The theoretical background and pedagogical consequences are discussed more in depth in the article presented in chapter 5.4.

In the present PhD, "academic writing" is used as an umbrella term for two subgenres that focus on writing from sources. This means that the genre of the academic essay or

composition is excluded because it requires different skills and strategies of writers (Wolfersberger, 2007, p. 48). Summary writing and synthesis writing are two closely related, yet different tasks. While a summary is based on *one* source text, intending to reconstruct its gist by paring it down to its major constituents (Rumelhart, 1977), a synthesis is composed based on *multiple* source texts of which content is selected and reorganised in an independent organisational structure (Spivey, 1990). This difference has an impact on the mental representation of the text(s) to be summarised, leading to different composing strategies. Drawing upon Hidi and Anderson (1986), Wolfersberger (2007) stated that while producing summaries, writers have a higher propensity to transfer source text structure to their own text, therefore minimising the process of organising. This process consists of creating a new structure for the emerging text, independent from the organisation of the sources. Together with the processes of selecting (of relevant information from the source) and connecting (of information drawn from the source with background knowledge used for inference), it forms the three-point-model of a synthesis task described by Spivey (1990).

Using a more fine-grained subdivision of the summarising task to understand its complexity, I differentiate between six (for summary writing) or seven (for synthesis writing) subtasks or process steps. The different subprocesses of the task are represented in Table 1, mapped to the three major writing stages of the Flower and Hayes (1981) model of composition writing.

Writing stage	Summarising subprocess
Planning	(1) Establish the macrostructure of source text(s)
	(2) Select main propositions
	(3) Merge the main propositions of the different source texts (only for syntheses)
	(4) Establish an own coherent structure
Translating	(5) Rephrase in writer's own words
	(6) Insert cohesive ties for reader-orientation
Reviewing	(7) Revise the final text on both local and global levels

Table 1 Subprocesses for summarising/synthesising, mapped to the three writing stages of the writing process proposed by Flower and Hayes (1981)

Table 1 clearly visualises that academic writing requires extended planning. A major challenge for novice writers (and their instruction) therefore is to gain awareness of the important role that planning and elaboration play in summarising processes in order to gain deep understanding (Solé, Miras, Castells, Espino, & Minguela, 2013). In order to compose their own text based on the sources, they have to make an appropriate selection of propositions to be integrated and establish a new coherence between these propositions, broadening their focus of linguistic attention from the word and sentence levels towards the text level (Paulus, 1999). This is especially difficult for L2 writers who have a natural propensity to focus on linguistic accuracy on a word and sentence level. Furthermore, they lack the knowledge of "how to create textual cues to guide the reader" (Wolfersberger, 2007, p. 58), i.e., appropriate cohesion strategies in the L2.

To investigate the effect of L2 academic writing instruction, a process approach has been adopted. Inspired by the cognitive paradigm, the development of the students' mental models of the academic writing task (or task representations), and their writing goals have been a focus of research (Manchón, 2012; Nicolás-Conesa, Roca de Larios, & Coyle, 2014; Wolfersberger, 2007). Nicolás-Conesa et al. (2014) found that a process-oriented approach to writing leads students to more problem-solving behaviour, striving for "more sophisticated sub-goals when composing and thus [producing] better texts" (Nicolás-Conesa et al., 2014, p. 16). Based on their findings, they recommended that "[L2] teachers might foster the conception of writing in terms of problem-solving behaviour through instruction and recursive writing practices oriented towards writers' goals" (Nicolás-Conesa et al., 2014, p. 16).

To sum up, L2 writing pedagogy that aims to support students in learning to summarise or synthesise, should address both the process and the product levels. More specifically, it should foster the following strategies and learning outcomes:

1. Process level: Building up writing and summarising strategies to cope with cognitive overload by scaffolding the process, encouraging
 - elaboration of source texts in terms of content (and in linguistic terms)
 - outlining (+ eventually rough drafting)
 - recursive writing
2. Product level: Raising the awareness of
 - audience (leading to reader-orientation)
 - global text quality features, including cohesion strategies
 - correct textual appropriation

In the following subchapter, the potential of computer-based instruction to contribute to the development of the necessary skills is discussed, based on three key elements of instruction according to a sociocultural approach, viz. strategy development, feedback provision, and collaboration.

1.1.4 Fostering (academic) L2 writing development: Instructional challenges and potential CALL solutions

To understand the possible contributions of online technologies to second language writing pedagogy, it is helpful to recall the basic distinction that Levy drew between "the computer as tutor" and "the computer as tool" (Levy, 1997). The first attribute is represented by applications for automated writing evaluation, ranging from simple spell check programs over online assessment tools that allow for the provision of automated, pre-programmed feedback, to sophisticated evaluation software that can handle free text input based on complex algorithms, e.g., LSA (latent semantic analysis). Examples for the "computer as tool" are online writing editors and tools that offer collaboration and interaction facilities, such as forums, blogs, wikis, and diverse applications on Google Drive. In other words, there is a broad range of online technologies available for writing pedagogy. Choices about which technologies to use should be based on general instructional philosophy, specific pedagogical aims, target group, and available facilities. In the following, three aspects in teaching (academic) writing in an L2 are outlined together with a short discussion of technologies that provide the necessary features to foster their development.

1.1.4.1 Strategy development

The ultimate goal of language teaching is to help learners become independent language users. In order to reach this goal, they have to be supported in the process towards independence. Self-regulation plays an important role in writing development, as Zimmerman and Risemberg (1997) have pointed out. They proposed a "social cognitive approach" to understand self-regulation in writing, which they subdivided into "environmental" (creation of effective writing settings and selection of models), "behavioral" (tracking of one's own performance), and "personal" (time planning and goal setting) processes (Zimmerman & Risemberg, 1997, p. 79). They attributed a high value in this process to strategies. Therefore, one important way of scaffolding novice (academic) writers is to help them build up writing strategies. The paramount role of strategies in learning-to-write has long been emphasised by research on writing instruction (De La Paz, 2007; Galbraith & Rijlaarsdam, 1999; Graham, 2006; Kellogg, 2008; Scardamalia, Bereiter, & Steinbach, 1984; Segev-Miller, 2004). From the very beginning of writing strategy research, the extent and the focus of revision have been singled out as factors impacting writing quality (Faigley & Witte, 1981).

Especially in a complex task like writing from sources that requires several skills and techniques, including extensive planning, strategies have to be applied and internalised in order to free working memory capacity. Breaking down the complex task into manageable subtasks is an important strategy learners can be trained through

scaffolding (e.g., scripting or step-by-step instruction). Summarising strategies include strategies that students in higher education are - or should be - already familiar with and implement in a more or less conscious way (like listening/reading for gist) and that just need to be refreshed. Others, like note taking or establishing the macrostructure of a text, are more likely to require explicit training. Other important strategies that have been emphasised in writing instruction research are outlining and rough drafting (de Smet, Broekkamp, Brand-Gruwel, & Kirschner, 2011; Kellogg, 1988; Torrance & Galbraith, 2006).

The focus on strategies in SLA research has first been introduced and advocated by cognitive learning theorists (Oxford, 1990). However, strategies have increasingly been studied from a sociocultural viewpoint, e.g. by Donato and MacCormick (1994), who classified them as "higher mental processes" that can mediate goal-directed actions towards language learning. They studied language learning strategies within an Activity theoretical framework (see section 1.1.5), affirming that

activity theory allows us to *define* strategies more completely than can be achieved with discrete-items lists and static categories. To analyze adequately language learning strategies requires discussion of all three levels of activity e.g. object-oriented learning activity (why the learner is using a particular strategy), goal-directed actions (how the learner is going about this task), and the operational composition of these actions under particular conditions (how the situation shapes, automatizes or de-automatizes strategic actions) (Donato & MacCormick, 1994, p. 455).

Lei (2009) reconceptualised writing strategies from a sociocultural perspective in order to fully explore their role in the writing of skilled and unskilled writers. Within the sociocultural paradigm, strategy development is regarded as highly interactive. It is interesting to note in this context that *observational learning* through coping models has been reported to yield high effects on the development of writing strategies (Couzijn, 1999; Raedts, Rijlaarsdam, van Waes, & Daems, 2007; Van Steendam, Rijlaarsdam, Seru, & Van den Bergh, 2010). In the same vein, both human-computer-interaction ("computer as tutor") and human-human interaction ("computer as facilitating tool") offer potential for writing strategy development. In the case of HCI, an individual learning path is an efficient online medium to train summarisation strategies. In such an environment, students can be guided stepwise via automated feedback through the different phases of summarisation, from identifying and elaborating key information in source texts in the planning phase, over rephrasing in the translation phase, to critically assessing their summaries in the reviewing phase. The advantage of an individual module is that those steps can be carried out, and feedback can be taken in, at an individual pace. Therefore, independently of their previous ability and experience in strategy use, all students within one classroom are challenged according to their level of

development. This instructional form has been implemented in the first intervention study. On the other hand, online groupware can be designed to effectively foster strategy development during human-human interaction.

1.1.4.2 Feedback provision

The lion's share of literature on feedback in L2 writing acquisition discusses the effectiveness of corrective written feedback on product improvement (Bitchener, 2012; Ene & Upton, 2014; Evans, James Hartshorn, & Strong-Krause, 2011; Ferris, 2010; Riazantseva, 2012; Storch, 2010; Van Beuningen, 2010; Wolsey, 2008). The assumption that feedback on grammar and vocabulary, i.e., LOCs, has an impact on writing development is not without controversy (Truscott, 1996; Van Beuningen, 2010). Feedback research drawing upon SCT has contributed to the discussion, pointing out that there is no single best route to stimulate a learner's ZPD because individual learner characteristics and experiences affect to a large extent how (s)he reacts to mediation: "[A]ffordances arise out of the successful tailoring of the interaction to the developmental level of individual learners" (Ellis, 2008, p. 528). Based on this insight, the model of "dynamic assessment" was developed (Davin, 2013; Lantolf & Thorne, 2006; Poehner, 2008; Poehner & Lantolf, 2013). This model suggests to dynamically adapt feedback according to the learner's reaction, mirroring his or her ability to self-regulate the L2 output. The adaptation occurs along a continuum between implicitness and explicitness, triggering to the maximum the learner's capacity of noticing and self-repair. Poehner and Lantolf (2013) and Davin (2013) have demonstrated how principles of *dynamic assessment* can be applied to research in learning-to-write, both in classroom-based and in computerized interaction. The negotiation of feedback provision in direct interaction between both parties involved is an important condition for the success of this approach. Therefore, it takes an experienced human tutor or an intelligent adaptive computer algorithm to implement the principles of dynamic assessment in a fruitful way.

In the context of academic writing in an L2, HOCs play a more important role, although the focus on LOCs should not be abandoned (Zhou, 2009). HOCs are more difficult to address in automated feedback without direct involvement of a teacher. In order to detect problems and provide feedback on the content and coherence levels in an automated way, natural language processing techniques like LSA and readability algorithms have to be applied. This is an interesting research field that goes beyond the scope of this PhD (see section "Directions for further research"). Another way to stimulate learners to focus on HOCs is by explicitly drawing their attention to these kind of textual problems. This can be done using *models* as mediating artefacts. In L2 writing, the successful use of *models*, i.e. a target-like production of writing by a native speaker, has been reported in picture gloss tasks aiming at *noticing* of problems on a local level (vocabulary, grammar) (Hanaoka, 2007; Martínez Esteban & Roca de Larios, 2010). For a

detailed discussion and description about how they can be implemented in advanced writing, please refer to sections 3.2 and 3.3.1.

Apart from a learner's ZPD, there are other individual characteristics that can play a mediating role in the uptake of feedback, as many researchers have outlined for both fields of online learning and learning-to-write (for a review of the relevant research, please refer to the literature review of the article included in section 3.2). Therefore, it is necessary to monitor these characteristics, e.g. learning style, strategy use, motivation, in order to understand their impact on a learner's activity while engaged in a L2 learning task.

How can online technologies provide added value regarding feedback provision? Feedback plays a vital role in online learning and is inextricably linked to design principles. Felix (2005) stated an "about turn" in the provision of feedback in CALL due to the focus on constructivist design principles at the beginning of the millennium. As constructivist pedagogy focuses on the process level and the cognitive involvement of the learner, feedback associated with this approach provides elaborated information about the targeted response. This is in stark contrast to feedback provision according to behaviourist principles: Instructivism seeks to automatize discrete language items like grammar rules and vocabulary retrieval. Therefore, instructivist feedback provides immediate knowledge of result, if necessary, accompanied by the correct response.

For feedback classification in online learning environments, several taxonomies have been suggested based on different characteristics, including timing (immediate vs. delayed feedback), focus (on form vs. on meaning), goal (formative vs. summative feedback), function with regard to the nature of student's learning ((meta)-cognitive or affective-motivational, see for example (Espasa, Guasch, & Alvarez, 2013; Narciss, 2008), origin (teacher, peer, or computer), content (Espasa & Meneses, 2010), degree of personalisation (Gallien & Oomen-Early, 2008), disclosure of result and/or correct response, and its degree of elaboration (Garrett, 1987; Rosselle, Sercu, & Vandepitte, 2009; Shute, 2008; van der Kleij, Eggen, Timmers, & Veldkamp, 2012).

The immediacy is the advantage of electronic feedback most often cited: "The capacity for computers to provide instant and individualized feedback has long been recognized by educators, including foreign language educators" (Zhao, 2003, p. 16). It has been claimed that immediate response to action leads to higher retention of feedback, "because [feedback] may be most efficient when it is provided while the procedural 'knowledge' that led to the error is still active in memory" (Donato, 2000). This seems to be confirmed by a meta-analysis of studies related to feedback timing across different disciplines carried out by Kulik and Kulik (1988), who concluded their analysis by recommending that "[t]eachers who want their quizzes to help students learn should try to arrange conditions so that students receive feedback as quickly as possible after they answer quiz questions" (Kulik & Kulik, 1988, p. 93). However, they also reported that the advantage of immediate feedback over delayed feedback

depended on the learning situation (quiz performance vs. test content), and the nature of the feedback intervention (classroom-based vs. experimental). Along the same lines, the effect sizes for feedback timing reported in Hattie and Timperley's research synthesis (2007) are rather small, compared to other aspects of feedback delivery. The average effect size of five meta-analyses that included a total of 178 studies investigating "delayed vs. immediate" feedback was .34, and the effect size of eight meta-analyses that included 398 studies investigating "immediate vs. delayed" feedback was .24. In general, studies on the impact of feedback type on students' uptake reach different conclusions (Kluger & DeNisi, 1996). Beside feedback characteristics, the importance of the students' attitudes and motivation for uptake has been emphasised (Van der Kleij et al., 2012). Feedback can only be effective when the learner is willing and able to use it. Therefore, it is important to take students' motivation and attitudes into account when examining the effect of feedback. For a detailed discussion on characteristics of (electronic) feedback and students' attitudes towards it, please refer to the article in section 3.2.

1.1.4.3 Collaborative dialogue through Computer-mediated communication

The sociocultural turn in the approach to L2 learning led, among others, to the favouring of collaborative settings for learning. Swain has coined the term *Collaborative dialogue* to describe the interaction between peers engaging in a collaborative language activity (Swain, 1997, 2000). Drawing on the SCT principle of mutual scaffolding, she exemplified in several studies how this dialogue can benefit individual learning. The unit of analysis used in this approach is the LRE (language-related episode). Peers jointly engaged in LREs pool their cognitive and linguistic resources (Storch, 2005), and therefore can rely on a broader array of resources to resolve them correctly than an individual learner. At the same time, the *languaging* (Swain, 2006), i.e., the use of the L2 to negotiate meaning, taking place in this kind of dialogue, can also contribute to individual language learning. From a sociocultural viewpoint, *languaging* represents a socially situated process in which knowledge is co-constructed. This knowledge can then be internalised by the individual participants of the dialogue as new knowledge or reinforcement of existing knowledge about the language (Storch, 2013, p. 156). Based on these assumptions, collaborative writing has a positive bearing on L2 development, because

the outcome of a collaborative writing activity is not just the jointly produced text. It is also collective cognition (...) related to language learning, including, for example, learning new vocabulary, improved ways of expressing ideas, gaining a greater understanding of certain grammatical conventions or greater control over the use of a particular grammatical structure (Storch, 2013, pp. 3-4)

These insights have also informed research on language learning in CMC (computer-mediated communication) settings. Vice-versa, CMC can broaden the understanding of collaborative dialogue in that the mode of communication (written, voice-based, video-based, or a combination) plays a role in its uptake for language learning. As Warschauer put it: "[W]e do not now have a traditional form of writing plus the computer, but rather we have entirely new forms of writing that need to be taught in their own right" (Warschauer, 2005, p. 42). Web 2.0 or CMC-tools provide interesting features for writing research because they facilitate goal-oriented written communication and the tracking of it. Furthermore, via online tools, the classroom can be connected with the world, allowing for intercultural exchanges between native and non-native speakers of a target language, which opens a whole new array of opportunity for language learning.

Research into collaborative writing with and without the use of technology has emphasised the beneficial effects of this setting on writing process and outcome. The effects reported range from accuracy and complexity gains (in comparison with individual writing settings) (De la Colina & García Mayo, 2007; Fernández Dobao, 2012; Glendinning & Howard, 2003; Wigglesworth & Storch, 2009) to better elaboration through content discussion and increased awareness of audience, leading to better reader-orientation (Blin & Appel, 2011; Kuteeva, 2011). CMC has also been reported to promote more equal participation of the participants, leading students, also the anxious ones, to produce more output in the target language (Beauvois, 1998).

However, good quality collaboration that leads to the intended interactions and outcomes, be it focus on form and/or on meaning, languaging, content elaboration, fluency, or mutual scaffolding, does not necessarily come by itself. Designing a collaborative activity requires careful consideration of several factors. Storch (2013), who provided a comprehensive overview of research on collaborative writing in classroom settings, pointed to several key issues that need to be considered, e.g., task type, grading, group size, and proficiency pairing. For computer-supported collaboration, there are even more factors that have to be taken into account. Strijbos et al. (Strijbos, Martens, & Jochems, 2004, p. 418) established a list of 29 questions grouped into six major areas that need to be addressed when designing CSCL (Computer-supported collaborative learning) activities (for a more detailed description, please refer to section 5.1). Especially when the collaborative task is complex or even wicked in nature, it is necessary to support the collaboration process by instructional means (Kirschner, Martens, & Strijbos, 2004). Scripting and modelling are two forms of effective instructional scaffolding that have been well-researched in CSCL settings (Dillenbourg & Jermann, 2007; Hämäläinen & Arvaja, 2009; Rummel & Spada, 2007; Rummel, Spada, & Hauser, 2009) and in writing instruction (Couzijn, 1999; Raedts et al., 2007; Van Steendam et al., 2010). Pre-task modelling was also proposed by Storch (2013) to encourage learners to adopt collaborative patterns in face-to-face interaction.

Research results on the impact of these instructional means on online collaboration quality are discussed in the introduction sections of two articles (5.2, 5.3).

In addition to the benefits that collaboration offers for (language) learning, pedagogic theorists state that it is necessary to introduce collaborative work forms in higher education from a vocational viewpoint.

Higher education suffers from its stress on the individual acquisition of knowledge and skills (competencies) (...) [whereas] [t]he vocationalist view of learning (...) holds that employers want higher education to attend more closely to what they consider they need in the graduates they recruit (...) The generic skills and competencies (...) require the implementation of a different approach to learning in a setting where there are shared responsibilities both for the process of achieving a final product and for the product itself, and where there is mutual trust between the participants such that they are valued for their contributions and their initiative. In other words, this can only be achieved in a collaborative and/or cooperative learning setting (Kirschner, Martens, & Strijbos, 2004, p. 3f).

In the previous subchapters, an overview of the state-of-the-art in the literature concerning (L2) (academic) writing development from a sociocultural viewpoint was provided, including information about the emergence of this approach out of preceding cognitive models, where appropriate. Furthermore, consequences for instruction and the contribution that more specifically CALL can pay to the development of academic writing were identified. The following part of the theoretical chapter presents Activity Theory as a comprehensive approach to investigate online instructional settings.

1.1.5 An Activity Theoretical Approach to Online Pedagogy for Academic Writing

Fürchte nicht das Chaos, denn im Chaos wird das Neue geboren (C.G. Jung)

1.1.5.1 Evolution of the activity-theoretical framework

To present the research carried out in the course of this PhD in a broader framework, and to evaluate in what way the different pedagogical settings using online technologies supported the learning-to-write process of the students, the Activity theoretical framework is used. The basic principles for this framework stem from Vygotsky's (1978) "genetic method" rooted in a sociocultural approach to learning. Based on this approach, Leontiev (1978) developed the activity theoretical system with a triangular model for learning based on three interacting constituents, viz. the *subject* (an individual or a collective) that carries out actions, the *object* of the actions, and the

mediating tools and artefacts that enable or restrict the interaction between the former two.

Engeström (1987) expanded Leontiev's model in order to describe how technological innovation shapes learning in collective work activities. He integrated the social dimension of human goal-oriented activity through adding the constituents *community*, *rules* and *division of labour* to the model, and expanded the *mediating tools* constituent with *technological tools*. In Engeström's model, all constituents are interrelated and can thus have an influence on each other, which is represented by two-way arrows between them (see Figure 2).

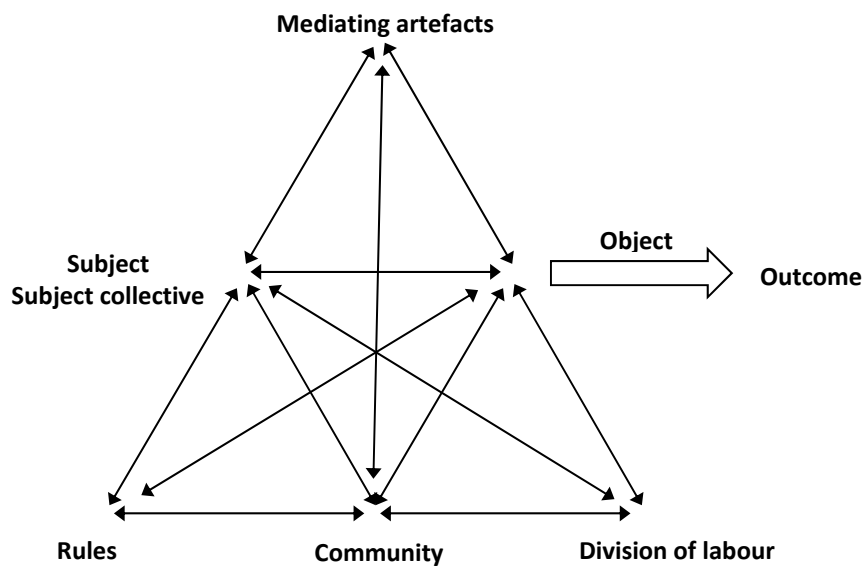


Figure 2 Activity System according to Engeström1987

Engeström (2001) postulated that a contemporary, *third generation* AT includes minimally two interacting AS'. He summarized this concept of AT to underpin his theory of *expansive learning* by stating the following five principles:

1. All goal-oriented *actions* are subsystems of an overarching *activity system* which is the *prime unit of analysis*.

In an educational context, this means that

[a] particular educational context (...) is itself a part of a number of activity systems relevant to participating individuals (for example, their prior experiences with computer technology and online communities or their prior experience with foreign-language learning), as well as the activity system of the semester-long

class (...) Importantly, a single activity system is influenced by multiple other life events and communities (Lantolf & Thorne, 2006, pp. 224-225).

2. The principle of *multi-voicedness* implies that all *subjects* involved in an AS have their own individual histories that shape their relations with the *object* of the activity in different ways. Lantolf and Pavlenko (2001) translate this principle to the educational context as follows:

[I]t is not necessarily the case that all the people in the language classes (...) are (...) engaged in the same activity [although they are engaged in the same overt task-related behaviour]. And this is ultimately what matters, because it is the activity and significance that shape the individual's orientation to learn or not (Lantolf & Pavlenko, 2001, p. 48).

As a consequence of the principle of *multi-voicedness*, individual learner histories, including learning style, self-efficacy beliefs, and attitudes towards the pedagogical approach, need to be integrated in the investigation of a pedagogical intervention, as they can have an impact on the learner's construction of the *object*.

Because of their fundamentally different relationship with regard to the object, the *subjects* involved in a classroom-based activity can be subdivided between the *subject* that formally decides on the object of the educational activity, i.e., the teacher, and a *subject collective* that can fully, partly, or not at all internalise the object proposed by the teacher, i.e., the students. This differentiation has been proposed by Müller-Hartmann and Schocker-v. Ditfurth (Müller-Hartmann & Schocker-v.Ditfurth, 2010). They were inspired⁴ by the concept of a "double subject" that arises from the adaptation of Engeströms AS for the design of a mobile learning project for first aiders proposed by McAndrew, Taylor, and Clow (2006). In their adapted AS, the design activity unfolds in two related spaces, the technological one, and "the more abstract semiotic 'learning-space'" (McAndrew et al., 2006). Accordingly, all AS constituents also unfold in these double spaces, i.e., the subject is at the same time a "technological subject" and a "semiotic subject". Blin and Munro (2008) adopted this new concept in their attempt to unravel why e-learning did not "transform and disrupt teaching practices in higher education" to the anticipated extent. In their proposed AS model, the subject unfolds into the "technological subject" *Moodle user*, and the "semiotic subject" *Lecturer/course designer* (Blin & Munro, 2008, p. 480).

⁴ Source: Personal communication via e-mail

Although this new interpretation of the subject in a double role differs from the original AT concept of the subject as a subject collective, it was found to yield explanatory power to uncover sources of contradictions. It was therefore adopted in this dissertation to reflect on the role of the teacher / researcher, and her relationship with the learners / participants.

3. The principle of *historicity* implies that AS' dynamically change over time. Consequently, "[t]heir problems and potentials can only be understood against their own history." (Engeström, 2001, p. 136). Blin (2005) exemplifies this principle by referring to an L2 teacher shaping her teaching practice according to accumulated experience:

The procedures and tools employed by the teacher have been accumulated over a long period of time, (...) in response to changes and advances in language pedagogy and in educational technologies (Blin, 2005, p. 50-51).

Furthermore, this principle entails the consequence that the roles of constituents are not fixed but can shift: "What initially appears as object may soon be transformed into an outcome, then turned into an instrument, and perhaps later into a rule" ("Center for Activity Theory and Developmental Work Research", 2003-2004). In CALL learning environments, *tools* and *objects* can be interchangeable. A new tool, or rather the attempt to familiarise with it, can be an object of the activity until the subject is sufficiently familiar with it in order to use it as an artefact (Blin, 2005, p. 49). This shift can frequently be observed in the use of technological tools for educational innovation. Blin (2005) applies the principle of historicity also to individual learner abilities whose role in the educational AS can shift between *tool* and *outcome*:

[A]ccording to the sociocultural perspective on Second Language Acquisition, and given the right conditions, the development and use of metacognitive and cognitive strategies can be seen as the *outcome of language use* as well as enabling factors for the development of language proficiency (Blin, 2005, p. 39).

4. *Contradictions* play a central role in furthering the development of an AS, as they can stimulate changes of an AS. They are defined as "historically accumulating structural tensions within and between activity systems" (Engeström, 2001, p. 137) Contradictions can arise through the adoption of new artefacts in an AS where old constituents, e.g., rules or division of labour, have not been changed. From an *expansive learning* viewpoint, contradictions have a positive effect because they trigger the adaptation of an AS to the new element, e.g., a new artefact, and therefore thrive innovation (Engeström, 2001). They can arise at

different levels. Primary contradictions occur within constituents of the AS that have a "double nature as unities of exchange value and use value" (Engeström, 1987, p. 130). Engeström provides the example of an object contradiction for a doctor for whom a patient is a "person to be helped" as well as a "source of revenue" ("Center for Activity Theory and Developmental Work Research", 2003-2004). Blin (2005) translates this example to the educational framework where the student is "a person to be educated" as well as a person "who will pass the state examinations and contribute to the teacher's (and the school's) success statistics" (Blin, 2005, p. 52). Secondary contradictions emerge between two or more constituents of an AS as a consequence of the adoption of new elements. Tertiary contradictions appear between two or more hierarchically interrelated AS', e.g. the educational activity system, and the AS of the teacher-coach-researcher that frames it.

For example, the tools available (e.g. the technology available may be obsolete) or the administrative rules and procedures in place (e.g. face-to-face teacher-learner and learner-learner interactions) may conflict with the new motive triggered by the need to help their students develop new communicative competences, such as synchronous and asynchronous electronic communication (Blin, 2005, p. 53).

Finally, quaternary contradictions can be observed between the AS of a central activity that changes and the AS of its neighbour activities that consistently might need / want to change as well. Blin (2005, p. 54) provides the example of diverging language teaching methodologies of departments within the same institution as a result of innovation in (only) one of the departments.

5. The changes or innovations triggered by the contradictions of an AS are called *expansive transformations*. According to Engeström, those transformations happen in cycles.

Activity systems move through relatively long cycles of qualitative transformations. As the contradictions of an activity system are aggravated, some individual participants begin to question and deviate from its established norms (...) An expansive transformation is accomplished when the object and motive of the activity are reconceptualized to embrace a radically wider horizon of possibilities than in the previous mode of the activity (Engeström, 2001, p. 137)

Based on these five principles, Engeström formulated a new approach to learning which he calls *expansive learning*. It differs from traditional approaches to the phenomenon of learning in that it does not share the presuppositions "that the knowledge or skill to be acquired is itself stable and [that] [t]here is a competent 'teacher' who knows what is to be learned" (Engeström, 2001, p. 137). Instead, learning is seen as a "collective

endeavor" (Engeström, 2001, p. 138) that "produces culturally new patterns of activity" (Engeström, 2001, p. 139).

1.1.5.2 The focus on activity

Activity theoretical systems, although not originally created for institutionalised educational settings, have been used to both describe and analyse situations in which learning takes place with the aim of innovation. One of the aspects that makes AT so interesting to use in educational settings aiming at innovation is its holistic approach to the activity of learning: "Constructing an activity system as a research object involves defining the roles that people, institutions, and artefacts play in moment-to-moment practice, thus eliding the analytic blind spots that teacher-, student-, or technology-centred approaches tend to produce" (Lantolf & Thorne, 2006, p. 224). Adopting AT also means to broaden the focus of research from the task as a "behavioural blue print that researchers employ to elicit linguistic data" (Lantolf & Thorne, 2006, p. 234) to a process-oriented approach to learning that aims to discover "what individuals and groups actually do while engaged in some communicative process" (Lantolf & Thorne, 2006, p. 234), i.e., in a learning task.

The tenet that *task* and *activity* should be treated as separate units of analysis is a basic underlying principle of AT. As Ellis (2008) stated, "task features [can be] seen as predisposing learners to behave in certain ways without predetermining their actual behaviour" (Ellis, 2008, p. 547). In the same vein, Elen coined the concept of "instructional disobedience" to describe unexpected student behaviour that can undermine the instructional goal of a task and even jeopardise instructional research by blurring results (Elen, 2013).

In adopting the *activity* as unit of analysis, the research focus on language learning activities is broadened, because it forces the researcher to clearly identify the role that different actions and artefacts play in a concrete educational setting and for the concrete learner. A number of studies in SLA have exemplified how the shift from task to activity as research construct increases the explanatory potential for observed individual differences in outcome (e.g. Donato, 2000; Roebuck, 2000; Seedhouse, 2005).

Also in CALL research, the dynamic nature of artefacts, and how it affects the potential for language learning, is discussed. Thorne used the AT framework to describe how the heterogeneity of Internet speech communities shaped cross-cultural CMC settings for language learning (Thorne, 1999, 2003).

Blin pointed out the need for research to "interrogate mediational artifacts⁵ and their cultures-of-use" (Blin, 2005, p. 67) because

the educational uses of Internet communication tools are affected by the cultures-of-use evolving from the manner in which these tools are used in everyday communicative practice. Internet communication tools, like all human artifacts, involve processes of acculturation. (...) The result is varying types, qualities, and quantities of participation in foreign language interaction within instructed educational settings with consequences for both the processes and products of language learning (op. cit., p. 66).

1.1.5.3 The concept of "affordances"

In an activity theoretical system, the goal-directed actions of the *subject* can be enabled and enhanced by the *mediating tools*, but they can also be limited by them. The inherent characteristics of an artefact that can enhance or constrain an action are called "affordances". This concept was originally developed by Gibson (1977) in the domain of cognitive psychology. It has been adopted by several disciplines, amongst which, HCI and Educational Psychology. The educational psychologist van Lier (2000) defined "affordances" as follows:

[A] particular property of the environment that is relevant - for good or for ill - to an active, perceiving organism in that environment. An affordance affords further action (but does not cause or trigger it). What becomes an affordance depends on what the organism does, what it wants, and what is useful for it (van Lier, 2000, p. 252)

A prospective user, thus, does not automatically perceive an affordance, because its perception might be hindered by physical, logical, or cultural *constraints* (Kirschner et al., 2004). Therefore, Kirschner et al. (2004), who discuss the term related to CSCL distinguish between *perceived* and *perceptible* affordances which is relevant when evaluating educational design (see discussion in section 1.2.2). In the same vein, it is also possible that certain affordances of an online environment or tool emerge only during learner interaction, without having been consciously engineered by the teacher-designer: "Using mixed methods (...) to trace and interpret the realization (or non-realization) of 'designed' and emergent affordances can assist us in enhancing our designs for successful language learning mediated by technology" (Kirschner et al., 2004, p. 22).

⁵ BE: artefacts, AE: artifacts. As I follow BE spelling conventions in my document, a *tension* between my text and embedded quotes of authors following AE spelling conventions arises.

Another important aspect relates to the *dimension* of affordances, as they can be technological, educational and social in nature. Based on this distinction, Kirschner et al. (2004, p. 16) established a scheme to represent the *usefulness* of an online environment (in their case, for CSCL) as a combination of *utility* and *usability*, drawing upon Nielsen (1993). Utility includes the educational and social functionalities provided through the respective affordances of the environment, while usability arises from the technological affordances.

Also in the domain of CALL, the term "affordances" has been adopted to describe the characteristics of artefacts (tools and learning environments) used to enhance language learning, and their uptake by students in a language learning activity. Blin et al. (2013) alert to the multidimensionality of the concept, pointing out that "[a] theory of affordances for CALL should not be reduced to the technological dimension. Rather, it should relate the latter to educational and social affordances" (Blin et al., 2013, p. 22). Moreover, in CALL, the utility of an artefact is also defined by its language learning potential. Van Lier (2004) added a fourth dimension to the concept of affordances, viz. "language affordances", which he defines as "relations of possibility between language learners [that] can be acted upon to make further linguistic action possible" (van Lier, 2004, p. 95). His view is connected with the two SCT-related concepts of "private speech" and "languaging" mentioned above (see 1.1.1), as mediating artefacts used in the learning process can afford the necessary conditions for those linguistic actions to take place.

1.2 Methodological Framework

In the first part of this subchapter, methodological principles for the design and evaluation of CALL activities are outlined. By way of an introduction, a brief overview of CALL design principles that have inspired the design of the intervention studies of this PhD is presented. In the following, I zoom in on the principles of design based research, and elucidate how the empirical studies carried out in the course of this PhD follow these principles, including the overarching research questions that are formulated within this methodological framework. The second part of the subchapter is dedicated to evaluation frameworks for CALL research. After a brief discussion of proposed evaluation frameworks in this field, I finally present the methodological aspects of the activity-theoretical framework used to evaluate the contribution of each intervention study to the innovation cycle of this PhD.

1.2.1 The evolution of CALL design principles

Design principles for CALL environments have evolved along the same lines as SLA (Second Language Acquisition) instructional paradigms. Speaking in general terms, there has been a shift from instructivist pattern-drill to (social and cognitive) constructivist principles, affording online pedagogy fresh scope (Bax, 2003; Chapelle, 2001; Dalgarno, 2001; Felix, 2002). The constructivist view on learning emphasizes the active role that the learner plays in it. Knowledge is constructed instead of acquired, which implies that instruction needs to support this process of construction (Kirschner, Martens, & Strijbos, 2004, p. 9). It has been widely discussed that the constructivist pedagogical framework challenges instructional design (Jonassen, 1997; Karagiorgi & Symeou, 2005; Windschitl, 2002). As Karagiorgi and Symeou pointed out: "Most of the problems [regarding pre-specification of knowledge, authentic evaluation and learner control] are attributed to the fact that constructivism is a learning theory and not an instructional-design theory" (Karagiorgi & Symeou, 2005, p. 17). As a consequence, teachers who want to implement constructivist pedagogy in the classroom face several dilemmas on "conceptual, pedagogical, cultural, and political planes" (Windschitl, 2002, p. 131). Karagiorgi and Symeou specifically mentioned the problem of measuring learning outcomes: "The conundrum that constructivism poses for instructional designers is that if each individual is responsible for knowledge construction, then designers can not determine and ensure a common set of outcomes for learning" (Karagiorgi & Symeou, 2005, p. 22).

The general trend is, therefore, a sound combination of approaches rather than a strict separation: "Instructional design does not exclude constructivist strategies, but may also choose alternative strategies when they are appropriate. Other theories such as behaviorism and cognitivism also have their strengths" (Karagiorgi & Symeou, 2005, p. 23). "Moderate constructivist" (also called "pragmatic constructivist") learning modules offer tasks with underlying constructivist pedagogical philosophy, translated to instructional design (Karagiorgi & Symeou, 2005). "The plethora of learning theories is so varied, and in some cases overlapping, that it would be naïve to suggest a black and white divide into strictly opposing schools of thought, i.e. constructivist versus instructivist" (Felix, 2005, p. 86). Hubbard & Siskin (2004) also argued convincingly to preserve a role for tutorial CALL. These reflections can be unified in the more general trend to a "post-method approach" advocating a conscious, reflected mix of L2 learning methods based on the three pedagogic parameters of particularity, practicality, and possibility (Kumaravadivelu, 2003 and 2012).

1.2.2 The Educational Design Framework Based on Affordances

"Step back from your natural propensity as designer to believe that you either 'represent the average user' or that you 'know what is best for him/her'." (Kirschner et al., 2004, p. 17)

1.2.2.1 Characteristics of design-based research

As Levy and Stockwell (2006, p. 26) noted, CALL design has become increasingly complex, trying to accommodate different complementary theoretical perspectives and technologies. A CALL designer should aim at the integration of the learner's, teacher's and institution's perspectives which might sometimes be conflicting (Levy & Stockwell, 2006, p. 28). Discussing several premises for CALL design, Levy and Stockwell stress the pivotal role that technology concerns should play in the design process, "especially when we acknowledge that each delivery system or software development tool has its particular limitations as well as its particular strengths" (Levy & Stockwell, 2006, p. 27). In other words, the *affordances* of the environments and programs involved have to be considered. Another important premise for successful CALL design is that the designer needs to attend to the potential users, including "their needs, goals, and characteristics" (Levy & Stockwell, 2006, p. 27). A third interesting aspect of CALL design they mention, building on Shneiderman (1987), is that it is "fundamentally a creative process, which often involves the discovery of new goals". All of the above mentioned statements can be linked to a view on educational design called DBR (design-based research).

Educational research is at the crossroads of educational theory and classroom practice. The former should inform the latter, and vice-versa. In order to bridge the gap between the two, educational scientists in the 1990s have come up with a methodology called design experimentation (Brown, 1992; Collins, 1992) as a general framework for inquiries on how educational innovation works in practice. Applying the same principles to research in CSCL, the term DBR was coined by Hoadley (2002). The general idea of this approach that "is not so much an approach as it is a series of approaches" (Barab & Squire, 2004, p. 2) is to provide a toolkit for the production "of new theories, artifacts, and practices that account for and potentially impact learning and teaching in naturalistic settings" (Barab & Squire, 2004, p. 2). During the last fifteen years, the characteristics of this paradigm have been refined (Barab & Squire, 2004; Design-Based Research Collective, 2003) and DBR has gained momentum in educational research. DBR in itself does not prescribe methodological procedures for data collection and analysis - in fact, preference is given to mixed-methods analysis of so-called "thick" data - but rather comprises a set of required characteristics for research in educational innovation. These five characteristics and a respective explanation are summarised in Table 2. DBR is pragmatic, grounded, iterative, integrative, and contextual. The participants of intervention studies have an important function inasmuch as they are

considered "coparticipants in the design and analysis" (Yutdhana, 2005, p. 171). This is reflected in a thorough monitoring of their attitudes towards and experiences with the pedagogic innovation at stake.

Table 2 □ Characteristics of design-based research.

<i>Characteristics</i>	<i>Explanations</i>
Pragmatic	<ul style="list-style-type: none"> • Design-based research refines both theory and practice. • The value of theory is appraised by the extent to which principles inform and improve practice.
Grounded	<ul style="list-style-type: none"> • Design is theory-driven and grounded in relevant research, theory and practice. • Design is conducted in real-world settings and the design process is embedded in, and studied through, design-based research.
Interactive, iterative, and flexible	<ul style="list-style-type: none"> • Designers are involved in the design processes and work together with participants. • Processes are iterative cycle of analysis, design, implementation, and redesign. • Initial plan is usually insufficiently detailed so that designers can make deliberate changes when necessary.
Integrative	<ul style="list-style-type: none"> • Mixed research methods are used to maximize the credibility of ongoing research. • Methods vary during different phases as new needs and issues emerge and the focus of the research evolves. • Rigor is purposefully maintained and discipline applied appropriate to the development phase.
Contextual	<ul style="list-style-type: none"> • The research process, research findings, and changes from the initial plan are documented. • Research results are connected with the design process and the setting. • The content and depth of generated design principles varies. • Guidance for applying generated principles is needed.

Table 2 Characteristics of design-based research (Wang & Hannafin, 2005, p. 8)

Wang and Hannafin indicated the potential of DBR for educational design using technologies (Wang & Hannafin, 2005) to investigate the impact of the latter on learning. In the same vein, Yutdhana (2005) claimed that DBR is an adequate method for CALL research accompanying the design of new environments. During the last decade, several CALL scholars have advocated and adopted DBR as a methodological framework for the development of computer-assisted learning environments based on an array of subsequent intervention studies (Bush, 2008; Hung, 2011; Jalkanen, 2013; Lund, 2008; Murphy, 2009; Wang, Song, Stone, & Yan, 2009; Wong & Looi, 2010; Yutdhana, 2005). Actually, Hémard's (2003) user-centred approach towards CALL design resembles DBR to a large extent in that it advocates an iterative process based on formative evaluation. Also, Colpaert's (2006) appeal that CALL design should be based on educational engineering principles according to a pedagogy-centred approach points in the same direction. The research consortium CALICO (Computer-Assisted Language Instruction Consortium) dedicated a monograph to the discussion of DBR in CALL research (Rodríguez & Pardo-Ballester, 2013). In his contribution to the volume, Levy (2013) indicated the potential of DBR to contribute to a higher uptake of research results by practitioners, and therefore to the normalisation of CALL in the classroom.

Kirschner (Kirschner, 2002; Kirschner et al., 2004) developed a "six-stage model for educational design" (see Figure 3), based on the tenets of design-based research, for the area of CSCL.

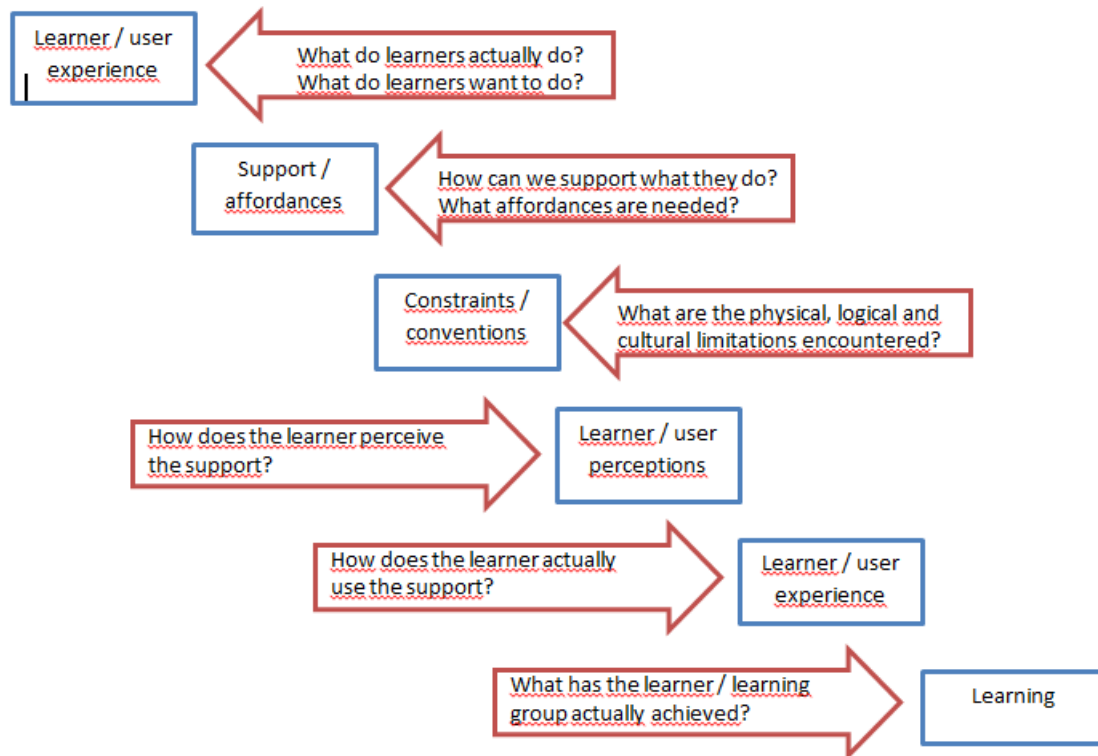


Figure 3 The six-stage model of educational design (Kirschner et al., 2004)

The educational principles underpinning this design model also stem from the sociocultural theory of learning, and therefore, the model is compatible with the activity theoretical approach taken in the description of the intervention studies (see 1.2.3.2). Engeström's model seeks to explain the impact that all constituents have on each other, and ultimately, on learning, which is visualised in the triangles of arrows connecting them. Kirschner's model basically translates the theoretical idea of interconnectedness to a dynamic design framework for research practice, based on concrete questions that build upon each other, focusing on the possible positive or negative impact of constituents, i.e., the educational, technological, and social (in the case of collaborative learning) affordances on learning. According to Kirschner et al. (2004), the educational designer must take into account the difference between the "perceptible" and the "perceived" affordances:

Although every object has specific affordances, what educational researchers and designers are actually dealing with are not the affordances themselves, but rather the combination of the perceptible (Gaver, 1996) or perceived (Norman, 1990; Norman, 1999) affordances, the constraints that are placed upon them, and the conventions regarding the affordance and its use (Kirschner et al., 2004, p. 12).

Translated to the use of technologies in an educational setting, this means that the technological, educational and social affordances of an educational artefact should not be taken for granted because students might not perceive them as such.

What makes Kirschner et al.'s (2004) approach unique is that they take a "probabilistic" view on learning, as opposed to the traditional "causal" design view. Whereas the latter presupposes that task design should lead to measurable skill *improvement*, the probabilistic view also accepts (partial) skill *development*, and even unforeseen learning effects as valuable learning outcomes. (Kirschner et al., 2004, pp. 11-12). Here again, a parallel can be drawn with the AT principle to use the *activity* as research construct, i.e., the focus is on what learners actually do when engaging in a task. Due to its nature that is dynamic and broad at the same time, DBR is suited to inform CALL research: "DBR provides us with a lens for understanding how we can enhance students' language learning through the use of technology" (Yutdhana, 2005, p. 177).

1.2.2.2 The design-based research cycle of the present PhD research

This PhD thesis aims to provide a comprehensive view on advanced writing instruction in technology-enriched environments. To this aim, the contributions and drawbacks of different online platforms, programs and pedagogies - or "the three P's of CALL research", as Chapelle (2001) called them - concerning the learning-to-write process are explained and described. The three intervention studies carried out reflect an attempt to evaluate a variety of different approaches, combining and comparing behaviourist and (socio-)constructivist elements, each building upon the *contradictions* in an activity-theoretical sense encountered in the preceding study. The intervention study cycle was preceded by two preliminary observational studies that allowed the researcher to gain insight into aspects of HCI.

In the following, the five studies carried out in the framework of this PhD research will be described in the light of the six-stage model of educational design proposed by Kirschner (2002). Kirschner explained the six steps that a designer-researcher implementing innovation in a learning environment should take as follows: First, "observe users interacting with software (...), and do this before designing and developing" (Kirschner et al., 2004, p. 17). This is what was done in the two preliminary studies of this PhD research cycle. The preliminary case studies 1 and 2 (P1 and P2) served to observe actual user experience with a commercial online language learning module (a) in general, (b) engaging in tasks related to summary writing. The outcomes of preliminary study 2 led to stage 2: "Determine, based on stage one, what actually needs to be supported/afforded" (Kirschner et al., 2004, p. 17). This stage was addressed at the beginning of the design of all three intervention studies: In IS1 (intervention study 1), the necessary information was drawn from the F2F classroom setting within which summary writing was instructed, and from P2 that provided information on the

strategies students employed while summarising in an online environment. In IS2 (intervention study 2) and IS3 (intervention study 3), the design was informed drawing upon the limitations experienced in the immediately preceding study (stage 3). In other words, the studies build upon each other in an evidence-based design cycle, incrementally refining the PhD project goals and design spaces, thus complying with one of Levy and Stockwell's (2006, p. 38) guidelines for CALL design. Graham also recommended this procedure in his inspiring keynote lecture entitled "How to Conduct High Quality Intervention Studies" at the EARLI (European Association of Research on Learning and Instruction) Writing SIG conference CoWR in Amsterdam 2014, encouraging writing intervention designers to "think like an engineer".

The rest of the stages were addressed in a cyclic way by all of the three intervention studies: in every study, the constraints of the classroom situation, educational conventions, and the technology in use were considered (stage 3), the learner perceptions were investigated by means of questionnaires and/or post-hoc focus groups (stage 4), the actual use of the support provided through technological means was monitored by means of keystroke logging or screen-capturing of the writing process (stage 5), and the learning outcome was measured (stage 6). However, for the data analysis of each study presented in the articles, each time a different specific focus was chosen: In IS1, the focus was on stages 4 and 5 (actual use and perceived effectiveness of the support), in IS2, stages 4 and 6 were addressed, comparing two fundamentally different kinds of learning environment for the same task (individual vs. collaborative online learning paths), and in IS3, stages 5 and 6 were the main scope of the data analysis, investigating the effectiveness of scaffolding mechanisms for a collaborative online writing task.

Table 3 represents the overarching research questions of this PhD based on the questions of Kirschner et al.'s (2004) research-design cycle, adding an activity theoretical stance. In the first column, Kirschner et al.'s formulation of the question is represented, followed by a more specific reformulation adapted to my own research cycle. The last column sums up the studies whose results (mainly) contribute to the answer. Table 4 provides an overview of the timeline of the empirical research in this PhD.

Table 3 Design-based research cycle applied to the studies of this PhD

RQ	Kirschner et al.	This PhD	Studies
1	What do learners actually do?	What are <i>stumbling stones</i> for advanced language learners engaging in online tasks generally, and in summarising tasks specifically?	P1 + P2
2	How can we support the learners?	What are the <i>affordances of different online pedagogies</i> and according <i>technologies</i> for academic writing activities?	IS1, IS2, IS3

3	What are the <i>limitations</i> encountered?	What <i>contradictions</i> arise within the educational activity, and between the educational activity and the overarching research activity?	IS1, IS2, IS3
4	How does the learner <i>perceive</i> the support?	What are the learners' <i>attitudes</i> towards the support using online technologies?	IS1
5	How does the learner actually <i>use</i> the support?	Which affordances of the pedagogies and technologies are <i>taken up</i> by learners? What are the <i>constraints</i> that limit the uptake?	IS1, IS3
6	What has the learner actually <i>achieved</i> ?	What is the <i>learning gain</i> from the different intervention studies? What <i>outcomes</i> did students (not) achieve? Consequently, where should innovation focus on?	IS2, IS3

Table 4 Timeline of the studies carried out in the course of the PhD

	2009-10 sem 2	2010-11 sem 1	2010-11 sem 2	2011-12 sem 1	2011-12 sem 2	2012-13 sem 1	2012-13 sem 2
Preliminary study 1	■						
Preliminary study 2		■					
Intervention study 1			■				
Intervention study 2				■			
Intervention study 3							■

1.2.3 Evaluation of the intervention studies

Naturally, one can go too far in the demand for the application of rigorous conditions to educational research. After all, if we managed to control for every possible confounding variable in an experimental design we would be left with the technology itself as an independent variable, when in today's learning environment this is inextricably linked to the instructional method and the context in which the learning takes place (Felix, 2005a, p. 2f).

1.2.3.1 CALL frameworks for task-based evaluation of language learning software

Throughout the relatively short history of CALL as a research field, the need for a thorough methodology for evaluation has repeatedly been claimed by leading researchers in the field (Chapelle, 2001; Hubbard, 1988; Jamieson & Chapelle, 2010; Leakey, 2011; Levy & Stockwell, 2006). This is a sine qua non to enable comparability of

research results, and to build further on existing knowledge in the field. Different methodologies have been developed for learning software evaluation.

The most common forms used by CALL teachers-designers for the evaluation of self-created materials are checklists and surveys (Levy & Stockwell, 2006, p. 44). Another conceptually different format are "third-party evaluations" (Levy & Stockwell, 2006, p. 45), which are adopted in situations where the evaluator "has had no direct involvement in the object of the evaluation" (Levy & Stockwell, 2006, p. 46). Levy and Stockwell alerted to two main challenges in this specific situation, namely "choosing the appropriate evaluation criteria and really getting to know the software, not only in itself but in judging how best it might be used in different settings with different learners" (Levy & Stockwell, 2006, p. 46). QuADEM (Quality Assessment of Digital Educational Material), the methodology used for the third-party evaluation of DUO (Deutsch-Uni Online) in the exploratory studies of the PhD, attempts to meet those two challenges. Regarding the first one, QuADEM provides a detailed checklist based on well-explained rubrics, allowing the evaluator to pick-and-choose the aspects (s)he considers important for the respective evaluation process. Regarding the second challenge, QuADEM proposes a user-centred approach. The evaluator observes an actual user while interacting with the software, and bases the evaluation on those observations and the user comments. This means that the focus is not on the software itself, but on its usability. On the other hand, QuADEM has its limitations regarding the focus on pedagogy, more specifically regarding the area which I termed "didactic usability" (see section 2.1.2).

Hubbard (1988) and Chapelle (2001) have provided broad frameworks to assess CALL software and CALL studies in all their dimensions. Hubbard's framework is built upon three elements: "teacher fit" (i.e., the underlying pedagogic approach), "learner fit" (i.e., how the design fosters different learner profiles), and "operational description" (i.e., the program surface itself). Chapelle based her task-centred framework on interactionist SLA theory and came up with the "3 P's" to be evaluated, viz. the platform, the program, and the pedagogy of an online task. In her view, online pedagogy, like general SLA pedagogy, should seek to combine focus on meaning with focus on form in order to lead to the acquisition of target-like structures, and therefore, offer the necessary learning potential. In order to measure the latter, she proposed to carry out empirical analyses of learner performance. Leakey (2011) provided a critical overview of all previously proposed CALL and SLA evaluation frameworks by meticulously mapping their criteria for learning effectiveness. Based on this procedure, and drawing on a task-based approach, he came up with a new, comprehensive list of twelve "criteria for CALL enhancement", including Chapelle's (2001) previously established six evaluation criteria. His framework also comprises fine-grained checklists for each criterion as well as a checklist for internal and external validity of the methodological design of CALL studies.

According to Leakey (2011), CALL effectiveness research needs to combine qualitative and quantitative methods. "The challenge to both technology and humanities researchers and pedagogues occurs at the human-computer interface. Is their coming together measurable?" (Leakey, 2001, p. 6). In this regard, he followed other scholars, e.g. Chapelle (2001), who stated that

[t]hese two methods [qualitative and quantitative data analysis] provide different and complementary information both of which are relevant to CALL task evaluation (...). The judgmental analysis should examine characteristics of the software and task in terms of criteria drawn from research on SLA. The empirical analyses address the same criteria but through data gathered to reveal the details of CALL use and learning outcomes (Chapelle, 2001, p. 54).

However, in his own list of validity criteria, Leakey (2011) only addresses quantitative data (with one exception), and therefore does not cover the aspects of a mixed-method or qualitative approach to data analysis. This is in stark contrast with his affirmation that only a mixed-method analysis can reveal a true picture of a potential learning process in CALL. This, in my view, is a missed opportunity to work towards a mixed-methods approach in CALL research.

One possible pitfall of gathering rich data of diverse nature is the inherent risk of creating a data dump, as Leakey (2011) rightly pointed out:

A potential danger of a mixed, or configured, approach might be that of falling between two stools: on the one hand, obtaining insufficiently robust data because of having too much complexity and too many confounding variables to grapple with in a real-life classroom (...), and on the other hand, having insufficient depth of context-specific insight and participant feedback due to a preoccupation with obtaining satisfactory sample sizes, full data sets and non-intrusive observation (Leakey, 2011, p. 12).

Another cautionary note is in place here. Leakey (2011), as most of the preceding CALL software evaluation frameworks, took a cognitive, task-based approach, and therefore reflected a rather "causal" design view. This view was dismantled by Kirschner et al. (2004) as a basis for comparative research studies that "tend to focus on the media used and surface characteristics of the education they provide". Kirschner et al. (2004) warn quite drastically that following a causal design view that focuses exclusively on an intended learning outcome, neglecting learning and interaction processes, can lead to the development of "mathemathantic"⁶ (Clark, 1989) learning material, i.e., material causing the "death of learning" (Clark, 1989, p. 10). By adopting an activity-theoretical

⁶ from Greek *mathema*=learning + *thanatos*=death

lens for the evaluation of the interventions of this PhD, the view on learning is broadened towards the whole spectrum of surrounding factors that can have an impact on learning.

1.2.3.2 Activity theory as basis for an activity-based evaluation of pedagogic innovation

Basic principles and graphical conventions

Engeström (1987) proposed a "methodological cycle of expansive developmental research" (Engeström, 1987, p. 323) that matches fairly well with the six steps for design-based research proposed by Kirschner et al. (2004). However, Engeström places more emphasis on the formation of new instruments as thriving force for innovation which he calls the "most dramatic [step] in the expansive methodology" (Engeström, 1987).

The new instruments can only be implemented in selected strategic tasks. Such tasks represent the points of probable breakthroughs into the qualitatively more advanced form of practice. In carrying out these tasks with the help of the new instruments, the participants of the activity system face intense conflicts between the old and the given new ways of doing and thinking - the tertiary contradiction (Engeström, 1987, p. 335).

Engeström further acknowledged that "[f]or the researchers, this step of expansive research is the most difficult and the most rewarding one" (Engeström, 1987, p. 335). He pinpointed two difficulties for the researchers, which are actually both related to patience. AT based expansive developmental research requires "patient on-site data collection" (Engeström, 1987, p. 335) on the one hand, and perseverance or even tolerance of ambiguity on the other, as the researchers might be confronted with unexpected outcomes:

The researchers face the fact that all their skillful efforts to make the participants acquire and apply the culturally more advanced models according to a plan have been partially futile. A genuine expansive cycle inevitably produces not only civilization but also an ingredient of wilderness. To get a theoretical grasp of this wilderness, to find and understand something unexpected as a piece of the history of the future is the reward (Engeström, 1987, p. 335).

Here again, a parallel can be drawn to Kirschner et al.'s *probabilistic design view* (2004) which implies that design-based research also produces "unforeseen" learning effects next to predefined goals (Kirschner et al., 2004, p. 11).

Due to its focus on innovation through mediating artefacts, AT has been gaining momentum as a theoretical-methodological framework in CALL research over the last decade. Blin (2005) adopted the principles of AT in order to underpin her research on language learner autonomy. For this purpose, she expanded the framework established by Chapelle (2001) for evaluation of CALL software and tasks. Chapelle's framework addresses CALL as situation-specific, evaluating it under the two perspectives of judgemental software and task analysis and empirical analysis of learner performance. Blin interpreted CALL as activity-specific, adding the collective activity layer of AT to the framework:

While Chapelle acknowledges the importance of situating the evaluation of the CALL task or software in its context (...), her set of principles and criteria do not extend to the collective activity (...) Chapelle's framework can therefore be interpreted, in activity theoretical terms, as primarily concerned with individual or group actions oriented towards specific language learning goals and mediated by a CALL task or artefact (...) An activity theoretical approach to CALL evaluation requires locating Chapelle's "sub-activity" (...) in its overall context, which is determined by the collective activity of which it is a constituent part (Blin, 2005, p. 59).

Blin (2005) therefore expanded the judgemental analysis to *all* constituents of the learning AS, including the potential contradictions which may manifest themselves through disturbances or conflicts. Furthermore, she advocated an empirical analysis of both process *and* outcomes of the activity, including the identification of the contradictions actually occurring while the activity is being carried out by learners, holding that "[s]uccessful integration of computers in language learning requires a holistic approach in order to scientifically understand what learners do when working with technology" (Blin, 2005, p. 22). As she further exemplified, applying the AT model to study a learning process is not an easy task because the researcher herself has to determine the minimal unit of analysis and delineate the boundaries of the system accordingly in order to identify the underlying contradictions (Blin, 2005, p. 72), as "[t]he most well-planned and streamlined actions involve failures, disruptions and unexpected innovations" (Engeström, Miettinen, & Punamäki, 1999, p. 32) that can only be explained at the level of the activity system. To graphically represent those contradictions in an AS or between adjacent AS' that trigger further development and innovation, a two-sided arrow or a lightening symbol is used.

Lantolf and Thorne (2006), who adopted AT to discuss several examples of qualitative classroom research regarding L2 development represent "outcomes that would be furthered through changes in the system" (Lantolf & Thorne, 2006, p. 259) in italics, and mark "new outcomes that were only hinted at in the initial (...) activity" (Lantolf & Thorne, 2006, p. 259) by a preceding asterisk. In this way, desirable future innovations can be integrated in the AT model of an activity under analysis. Describing an action

research example of peer-to-peer scaffolding in L2 essay writing, Lantolf and Thorne (2006) used these graphical conventions to exemplify how an AT-based approach can help understanding shifts in motivation that are a result of contradictions and can lead to considerable changes in outcome: "A number of possible outcomes (...) would require (...) certain changes in the rules - division of labor, and mediating artifacts of the activity system - to see their complete fruition." (Lantolf & Thorne, 2006, p. 258).

Investigating the affordances of mediated artefacts

As mediating artefacts are at the very heart of CALL research, due consideration should be given to the investigation of their impact as constituents of both the educational and the overarching research-related AS. A good starting point for this investigation are their affordances for a specific language learning activity. Blin et al. (2013) proposed to differentiate between four dimensions of affordances in CALL research, and represented them as a nested construct (see Figure 4).

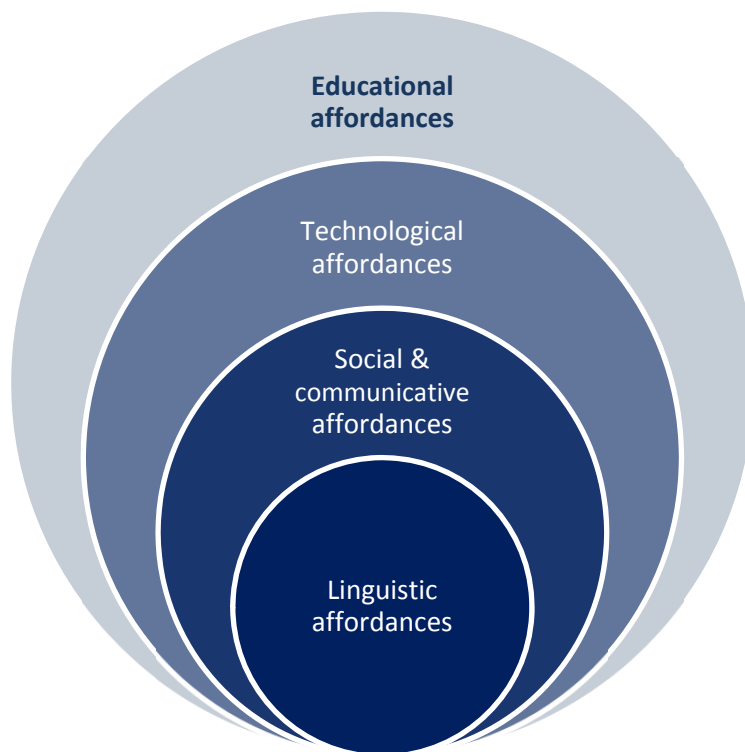


Figure 4 Nested construct of affordances according to Blin et al. (2013)

For the purpose of evaluation, however, it seems more appropriate to represent the dimensions of affordances as separate entities that can be mapped onto the concepts of *utility* and *usability*. Therefore, I propose the following scheme, expanding Kirschner et al.'s (2004, p. 16) model of affordances with *linguistic affordances* as important contributor to the educational functionality of artefacts in CALL (see Figure 5):

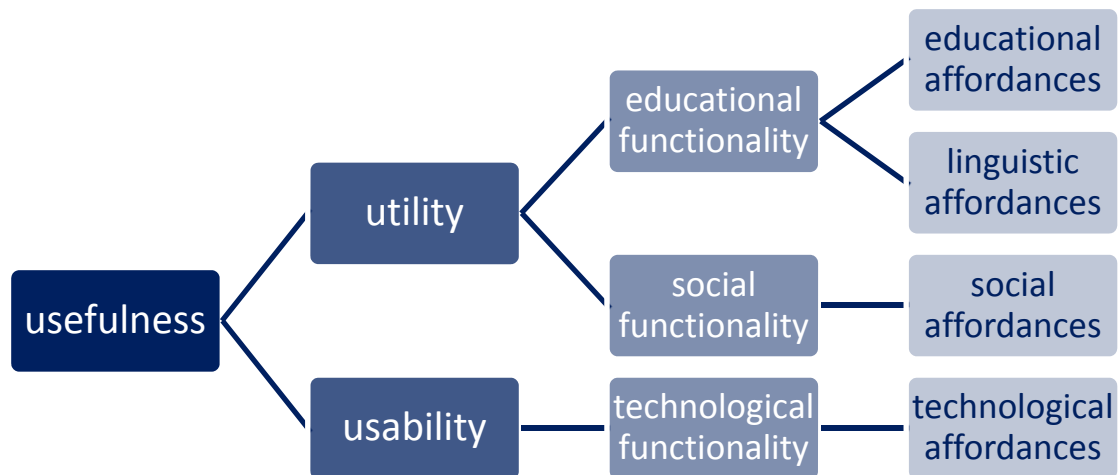


Figure 5 A model of affordances for CALL, based on Kirschner et al.'s (2004) model of affordances for CSCL

However important affordances of mediating artefacts might be in usability studies involving technologies, like is the case for CALL, they are only part of the whole picture to be investigated, as the AT framework suggests. Tools have affordances that can or cannot fit specific activities in which specific subjects are involved to reach a specific outcome, under specific circumstances (rules, communities, division of labour). In this vein, Colpaert⁷ proposed a formula to define the *added value* of a language learning technology as a function depending on the fit between two variables, viz. the requirements of the learning environment and the affordances of the technology:

$$\begin{aligned} & \text{Added_value (technologyX)} \\ & = \text{FIT (requirements (optimally_designed_learning_environment), affordances} \\ & \quad \text{(technology))} \end{aligned}$$

Drawing upon his earlier work (Colpaert, 2006), the learning environment is required to be "optimally designed" according to educational engineering principles of language learning. In other words, technologies in and by themselves do not have an inherent added value for language learning. It is only in the interplay with a matching learning environment that their affordances can add to learning.

⁷ Source: Personal communication via e-mail, January 2015

A broad perspective on evaluating CALL activities

To conclude, it can be stated that evaluation frameworks based on SCT and cognitive learning theories, respectively, actually address two basically different questions. Whereas the evaluation of learning material from a cognitive approach mainly answers the question "Did it work?", SCT-based evaluation moves beyond this focus on outcome to seek answers to the question "Why did it (not) work?". An AT approach can help to describe and understand the dynamics of CALL processes as well as the causes for and the role of innovation. Taking into account Kirschner et al.'s (2004) *probabilistic* view on learning, a third question can be added: "In what other, unforeseen, ways did it work?". These questions will be answered in the Conclusion section, in which the findings of all intervention studies carried out in the course of this PhD will be summarised, taking into account the results of their evaluation through an activity-theoretical lens presented in Chapter 6.

The two preliminary studies and three intervention studies carried out in the course of this PhD are understood as "expansive research" in the sense of Engeström (1987). Therefore, I followed his basic rule for evaluating and reporting my research:

Reporting and assessing outcomes of expansive research is not easy. (...) There is a simple rule for such reporting. One should apply the historico-genetic method also in the presentation of the research findings. In other words, one should reproduce the actual course of the expansive transition, following its basic temporal structure (Engeström, 1987, pp. 335-336).

Consequently, each of the three chapters that describe and discuss an intervention study (3-5) is concluded with an AT based reflection on the contradictions encountered during its assessment, indicating the consequences for the design of the following study each time. Like this, the development of the expansive research cycle is presented applying the "historico-genetic method".

Chapter 2

Preliminary explorative investigation of student-task-tool interaction

To gain insight into HCI (human-computer interaction) processes at the outset of the PhD research cycle, two preliminary studies were carried out targeting the usability of an existing language learning environment for advanced German L2. The environment used in this preliminary stage is the commercial online environment DUO (Deutsch-Uni Online) developed at the Ludwig-Maximilians-Universität München⁸ that aims at preparing students for, or accompany students during, a study abroad in Germany. DUO was deliberately chosen for this study due to its state-of-the-art pedagogic approach following a "moderate constructivist" design. The language modules present a sound mixture of (mainly) constructivist tasks that involve the learners in their own learning process, and instructivist-like pattern drill where appropriate. This learning software thus illustrates well the coexistence of multiple teaching and learning theories in one package, proposed by Kumaravadivelu as "post-method approach" (Kumaravadivelu, 2006), and supported by several scholars that have investigated the effectiveness of e-learning environments (e.g. Mehanna, 2004).

According to Kirschner et al. (2004), a pivotal aspect of educational design using technology is "locus of control", i.e., the extent to which the learner can modify specific aspects of the task to his/her perceived needs. The basic assumption, based on "conventional wisdom" is that the more control a learner has over the task or tool, the more rewarding the learning experience is (Kirschner et al., 2004, p. 24). This is backed by some studies evidencing a higher intrinsic motivation when the locus of control in technology-enhanced learning environments is shifted towards the learner (e.g. Kinzie, Sullivan, & Berdel, 1988). Along these lines, commercial language learning environments, like DUO, allow users to adapt the software to their learning preferences to some degree. In DUO, the learner can choose the access to the exercises via module, subject, or a predefined learning path, and the form of metalinguistic explanations (presented either inductively or deductively). In this way, DUO partly compensates for the lack of knowledge about its users' needs and characteristics, an inherent problem of commercial products that aim at a broader public (Levy & Stockwell, 2006, p. 36).

Following Kirschner's (2002) six-stage model of educational design, usability tests are a good starting point for the enterprise of developing online learning material. This is reiterated by Blin et al. (2013): "Conducting usability tests in the context of CALL can serve as an effective method to elicit information about learner behaviour, shed light on the process of the learner-task-tool interaction, and ultimately better scaffold this process" (Blin et al., 2013, p. 25).

⁸ in cooperation with the *Gesellschaft für Akademische Studienvorbereitung und Testentwicklung* (g.a.s.t.) e. V. <http://www.deutsch-uni.com/gast/duo/info/index.do?do=index>

2.1 Preliminary Study 1: Usability of Deutsch-Uni Online

2.1.1 Background information on the study

This first exploratory study was carried out in February-March 2010 and involved two volunteering female students of the Applied Languages study programme at HoGent (University College of Ghent). Thick qualitative data of their interaction with the software that lasted about two hours each were gathered for a usability analysis, viz. screen-capturing including a TAP (think-aloud protocol) and a video recording of the participants' faces with the screen-capture software *Morae*⁹. To elucidate specific aspects of their interaction, highlight videos were produced from the screencast recordings serving for stimulated recall in a semi-structured post-hoc interview. The six tasks selected for the usability study are all related to information about studying in Germany, but at the same time target specific language issues (see Appendix 1¹⁰). The two students were selected according to their common interest in, but different perspective on this topic. One of them was a Ba3-student who had just finished her study abroad in Germany, the other one was a Ba2-student who was preparing her study abroad in Germany. The Ba3-student also wrote a bachelor thesis about her experience with the DUO software, critically assessing the informational value of the tasks she completed in the light of her study experience abroad in Germany (Cologne). For the data elicitation and analysis, I used the quality assessment method QuADEM (Quality Assessment of Digital Educational Material) as a framework. This method was developed in 2008-9 at the Universities of Antwerp and Ghent, funded by the European *Minerva* programme.

The two main research questions of this qualitative study were:

1. How appropriate is the QuADEM framework for a reliable usability analysis of an online language learning module?
2. What are stumbling stones in online educational material to be avoided when creating online language learning modules?

The results of the study were presented at the Antwerp 2010 CALL conference (title of the presentation: "Quality assessment of an online language learning module

⁹ Please refer to section 6.2.7.2 for a detailed description of the tool

¹⁰ All appendices are provided in electronic form on the CD-ROM that accompanies this book

preparing a study semester abroad"), and the results concerning RQ1 were elaborated in an article for the international peer-reviewed journal listed in the Web of Science (impact factor 0.880) *Computer-Assisted Language Learning* that was published in co-authorship with Geert Jacobs (vol. 24 nr. 5, pp. 334-349). He assisted me in the preparation of the presentation, proofread the article and steered me through the reviewing process.

Concerning RQ2, I found two important factors that caused motivational loss even with my two highly motivated participants, viz. a) an unclear learning objective or a task or exercise design that does not (seem to) fit the identified learning objective and b) overwhelming negative feedback in conjunction with a missing indication of the potential error source, i.e., the students were required to compare the - sometimes very subtle - differences between their own solution and a model solution.

Based on the observations, I proposed to subsume feedback provision and presentation of learning objective under the denominator "didactic usability", which should be treated as a separate and core unit in the assessment of digital educational material (see article in the following section).

2.1.2 Assessing QuADEM. Preliminary notes on a new method for evaluating online learning courseware (Article in *Computer-Assisted Language Learning*)¹¹

In this paper we set out to assess QuADEM, one of the latest methods for evaluating online language learning courseware. What is special about QuADEM is that the evaluation is based on observing the actual usage of the online courseware and that, from a checklist of twelve different components, the evaluator is free to pick and choose one or more. In particular, we focus on the QuADEM evaluation of a module of the digital environment DUO (Deutsch-Uni Online) that aims at preparing B1/B2 students for a study semester in Germany. DUO is meant for self-study supported by an online tutor. For our assessment, we observed two respondents during their activities in the online learning module, using think-aloud protocol, video registration and keystroke logging, and we conducted semi-structured post-intervention interviews with them. Zooming in on usability, we found that this QuADEM component lacks assessment criteria regarding feedback and task design, both of which turned out to play an important motivational role in our assessment. While both could be added to the QuADEM usability dimension under the denomination “didactic usability”, we suggest that it might be worth reconsidering QuADEM’s pick and choose approach.

Keywords: quality assessment of digital courseware; usability; feedback; task design; didactic usability

Introduction

In view of the ever-increasing amount of digital courseware for language learning a great variety of evaluation instruments has been designed, ranging from complex frameworks (e.g. Hubbard, 1988) over rubrics (e.g. Thomé, 1989, Garrido & Geissler, 1998) and surveys (e.g. Hémard and Cushion, 2001) to checklists (e.g. Odell, 1986; Baumgartner, 2002; Niehoff, 2003; Roche, 2008). The choice for a specific evaluation tool should be informed by various interrelated motives, such as the assessment scope (formative or summative), the assessor’s involvement with the software development (“teacher-designer” vs. “third-party-evaluation”) and the intended outcome (decision-driven evaluation vs. research) (cf. Levy, 2006, p. 40ff).

¹¹ Strobl, C. & Jacobs, G. (2011): Assessing QuADEM: preliminary notes on a new method for evaluating online language learning courseware. *Computer Assisted Language Learning* 24(5), 433-449.

An interesting newcomer to the wide range of evaluation tools is the QuADEM method (Quality Assessment of Digital Educational Material), which has been recently developed at the Universities of Antwerp and Ghent and which was funded by the European *Minerva* programme. As the QuADEM assessment method is based on checklists, some fundamental caveats in the literature regarding checklist-based assessment (for a good overview cf. Susser, 2001) should be considered. We will refer to these later when we describe QuADEM. Following the core idea of the QuADEM approach, the evaluation is based on observing the actual usage of the online learning module by users belonging to the target group. In this way, one of the major problems of courseware assessment in general can partly be overcome: the tendency towards subjective and biased teachers' evaluation as observed e.g. by Reiser and Dick (1990) is less likely to occur because the assessor's judgement is based on a second viewpoint, viz. the learner's. The QuADEM methodology complies with the principle stipulated by Chapelle (2001) that "CALL should be evaluated through two perspectives: judgemental analysis of software and planned tasks, and empirical analyses of learners' performance." (Chapelle, 2001, p. 52).

While QuADEM and many other tools have been created to meet the growing demand for evaluating the quality of online learning materials, there is as yet very little work available that, in turn, assesses the utility of such tools. Therefore, our study had a clear objective: present a preliminary investigation into QuADEM. Put simply, we intend to find out to what extent QuADEM proves to be an adequate methodological framework to analyse an online language learning environment.

Extensive data collection, including video registration, think aloud protocols, and keystroke and screen logging of the intervention formed the input for our qualitative analysis. The main findings from the data evaluation together with video excerpts served as the basis for a post-hoc semi-structured interview, which aimed at better understanding the observed actions and reactions of the respondents.

In this paper, we first set out to describe QuADEM in the light of current discussion on evaluation methods and instruments. We then provide a short description of the learning software DUO that was used for the assessment, followed by an elucidation of the assessment procedure and the data collection. The core part of the article consists of the presentation and analysis of the main results. The discussion of those results in terms of completeness will finally lead us to a concluding appreciation of the QuADEM method and some general observations on the evaluation of language learning software, including the suggestion of a new evaluation category.

The QuADEM method in the light of current discussion about online courseware evaluation

The QuADEM method was specifically developed for the assessment of digital educational material in the field of academic/professional writing, but, according to the authors, it can also be used as a toolbox for the assessment of digital educational material in general (Opdenacker et al., 2009, p.9). Furthermore, the authors claim that the method “is especially well-fitted to support quality assessments that wish to apply a broad scope, going beyond a focus on the ‘ease of use’ to include an assessment of the accessibility and didactical effectiveness of the digital educational material under revision” (Opdenacker et al., 2009, p.11). It therefore constitutes a good example of the “move from a focus on technical usability to an approach that is centred increasingly on the user experience” as observed by Kukulska-Hulme and Shield (2004).

Since the method employs a broad interpretation of quality, it also runs the risk of becoming laborious and cumbersome. To maintain purposiveness and efficiency, the QuADEM method is devised in such a way that users can *pick and choose* exactly what they need. In this, it is comparable to other evaluation rubrics that list “[i]ssues to consider when developing a personalized software evaluation checklist” (Garrido & Geissler, 1998, p. 5). Therefore it “supports quality assessments of varying scale, ranging from focused evaluations aimed at very specific points of concern to extensive investigations” (Opdenacker et al., 2009, p.12) covering twelve different components that can determine the quality of digital learning modules: (1) blended learning, (2) learning objectives, (3) content, (4) style and language, (5) intercultural aspects, (6) usability, (7) learning styles, (8) writing styles, (9) testing, (10) examples, (11) multimedia, and (12) questionnaires.

For each of those components called “assessment units”, the QuADEM manual not only provides the assessor with checklists bundling analysis criteria (for an example of a checklist, refer to the “Results” section of this article), but it also gives advice on adequate assessment methods and provides short rationales on the underlying approach and state-of-the-art theories. In other words, apart from the question of *what* to assess, QuADEM also gives an idea of *how* and *on what grounds* the evaluation should take place. This makes the method accessible even for non-experienced courseware evaluators. For each of the checklist criteria, an “explanation” (how should it ideally be?) is given in the manual section of each unit. QuADEM thus helps the assessor not only to identify key issues, but also to indicate how a problem could be solved, filling a gap that is typical of checklist-based evaluation (Oliver, 2000). To illustrate how the QuADEM manual supports the assessor, we include the screenshot of one item out of the “usability unit” checklist, i.e. “The layout of the digital learning module is appealing”. In the right “explanation” column, a short rationale on the relevance of the item is given,

followed by a list of some concrete and specific points of interest, and an advice on the most appropriate research method.

Usability		
	Criteria	Explanation
1	The layout of the digital learning module is appealing.	<p>If users consider the module not visually appealing it will negatively influence their perception of the module. To make the module more appealing:</p> <ul style="list-style-type: none"> ▪ Use a limited number of colours. ▪ Use bright colours sparsely. ▪ Do not overuse pictures. ▪ Do not avert attention from the main content. <p>The idea of what is appealing is subjective and is ideally verified in a focus group.</p>

Figure 1. Screenshot of the QuADEM manual, p. 52.

It has been suggested that evaluation frameworks should not be biased toward a specific pedagogical approach (Hubbard, 1988). This is not the case for QuADEM, as the authors clearly state that their “view on what constitutes ‘good’ educational material is strongly influenced by the theories of cognitive constructivism and social constructivism” (Opdenacker et al., 2009, p.11). Obviously, this turns QuADEM into a less appropriate framework for the analysis of educational material that is underpinned by instructivist learning theories. However, this bias should not jeopardize the evaluation procedure when assessing courseware that complies with the pedagogical approach of the framework. It is important, though, that the underlying pedagogical assumptions – which will always be present in a more or less transparent form in any evaluation framework taking the educational aspect seriously – are explicitly stated (cf. also Susser, 2001, p. 270).

As a result of the inherent flexibility, one crucial step in the QuADEM assessment procedure consists in selecting the relevant evaluation units. The manual stipulates that “[t]he selection of the assessment units will depend on the scope and focus of the assessment, but also on the nature of the digital educational material under review” (Opdenacker et al., 2009, p.15). This freedom of choice results in a high responsibility for the assessors because their choice will strongly influence the evaluation outcome. Therefore, it is one of our key questions whether this freedom of choice might jeopardize a reliable assessment result. In other words, we wondered whether it might

be dangerous to stimulate partial assessment of educational software without proposing any kind of hierarchy or “key elements” that should always be considered.

The online learning environment used for the assessment experiment: Deutsch- Uni Online (DUO)

DUO was primarily set up as a language learning environment that specifically meets the needs of foreign students preparing for a study period in Germany. Its development history started ten years ago with a research project at the German FL department of the Ludwig-Maximilian-Universität in Munich, Germany. According to the authors, state-of-the-art pedagogical and didactic insights were considered when developing DUO, with a clear focus on content and pedagogy rather than on technology (Wegele, 2006). This developmental focus could be seen as an effort to bridge “the gap between technology and language pedagogy”, which was identified as the main obstacle for the breakthrough of tutorial CALL by Colpaert (Colpaert, 2006, p. 480). The pedagogical foundations of DUO can be classified as “moderate constructivism” (Roche, 2008).

The language learning modules in DUO are subdivided into independent *tasks* that can be accessed via two different menus, ordered by skill or by topic. Each task consists of several *steps* that guide the user through a range of *exercises* scaffolding the acquisition process. DUO clearly chooses a multi-method approach, both on a macro- and a micro-level. Multi-method on a macro-level means that users can choose between creating their own learning path by selecting tasks and following a linear path through the tasks. Exercise types range from multiple choice over drag-and-drop and fill-the-gap exercises to open-end questions with sample solutions for self-correcting, which complies with the cognitive constructivist approach. They also include questions to be answered and discussed in a tutorial forum or chat, stimulating social interaction in the target language. On a micro-level, we find a multi-method approach in the grammar presentations, where the user can choose between two animated films, corresponding respectively to the inductive and the deductive approach. With this mix DUO addresses different learning styles, which turns it into an interesting environment for user-centred research: we expect to observe divergent users’ reactions to its various elements, and thus be able to draw conclusions on learner preferences and learning styles.

Besides the tasks DUO provides the user with additional resources, personal spaces, and communication channels. These can all be accessed via intuitive icons in the upper navigation frame of the website. There are six information resources: an overview of all the contents in the module, a grammar section, a lookup dictionary, a reference guide to the efficient use of the different exercise types and to language learning strategies in general, a list with external links of general interest to exchange students, and an extensive help section. Two instruments invite the users to individually manage their

learning process: a notebook and a personal administrative section where the users' homework and corrections by the tutor are stored. Four communication channels support social interaction: forum and chat (a general one accessible to all DUO users and a class-specific one accessible to users who choose the subscription with tutorial aid), and e-mail and virtual whiteboard (linked to a virtual class).



Figure 2. Screenshot of the main page of one module in DUO.

Design of the assessment study and data collection

As mentioned above, QuADEM offers 12 assessment categories for a *pick and choose* approach. For our preliminary assessment of QuADEM we have decided to focus on *usability* since it is a key aspect for the quality of interactive online material and forms part of most evaluation instruments, albeit under different denominations (e.g. “technical aspects” combined with aspects of “learner fit” in the CALICO software review procedure). As mentioned above, our aim was to investigate to what extent it was possible to achieve reliable and satisfying results for our analysis of DUO based on just a single aspect of the materials QuADEM provides.

While the term *usability* embraces a broad range of quite diverse elements in website design, the QuADEM method is based on a relatively narrow interpretation of this term. The checklist items cover only the two layers of *technical usability* and *general usability*

established in website usability testing. We will elaborate on this idea later in the discussion section.

1	The layout of the digital learning module is appealing.
2	The interface of the digital learning module functions efficiently.
3	First time users find it easy to use the digital learning module.
4	The learning module allows the user to find specific information quickly.
5	The digital learning module meets the expectations of the user.
6	All multimedia (video, audio, hyperlinks, pdf) within the digital learning module functions as it should.

Table 1. QuADEM checklist for usability assessment.

For the evaluation several tasks were selected out of two different DUO modules, “uni-deutsch sprachkurs” at CEF-level B1/2 focusing on general and academic language acquisition, and “uni-deutsch studienpraxis” focusing on practical information about studying in Germany. The selection was informed by two criteria: 1) relevant content for students planning a semester in Germany, and 2) a broad range of exercise types and targeted skills.

We used the “mediated concurrent verbal protocol” methodology, which is an ethnographic observation procedure proposed in the QuADEM methodological compendium (Opdenacker et al., 2009, p. 113-116). This means we stimulated the respondents to think aloud, using their mother tongue (Dutch) or the target language (German). The researcher was present while the respondents carried out the tasks and occasionally interfered, e.g. asking the respondents to utter their thoughts when they remained silent for too long, and sometimes stimulating them to use a specific tool or feature when they did not do so spontaneously. Together with the verbal protocol, we recorded a video of the respondents’ facial expressions, and tracked keystrokes and mouse-clicks. This was done using Morae recording software. With the rich data that is captured simultaneously, this software is an ideal monitoring tool for ethnographic observation of learners’ behaviour in a CALL environment.

Our respondents were two female translation bachelor students of the University College Ghent, one 3rd-year bachelor student having recently completed a *Socrates* semester abroad in Germany (referred to as S1 in the following), and one 2nd-year bachelor student expecting to spend a semester abroad in Germany in the near future (S2). First, S1 worked on six pre-selected tasks, of which she then chose the two that she

considered the most representative and interesting ones – based on her former experience as a *Socrates* student – for S2. The two selected tasks focus on study-specific, interculturally interesting topics. S1 selected these two tasks not only because of the content but also because of the usability-related difficulties she encountered while working on them. The analysis presented in this paper is based on those two tasks as we can compare the two respondents’ visions and experiences.

A few weeks after the recordings, a semi-structured interview was conducted with both respondents at the same time in Dutch (L1), following a methodological suggestion in the QuADEM manual. During this interview, the respondents were invited to watch excerpts of the recording and to comment on general statements concerning their DUO experience and on concrete instances of the experiment.

It has to be noted that, for a true usability test, at least four participants would have been needed in the experiment, even when dealing with such a homogeneous group of respondents (cf. Opdenacker et al., 2009, p. 124). Nevertheless, this is beyond the scope of this study. We intended to perform a preliminary test on the strengths and weaknesses of the QuADEM approach. The rich data we gathered using the two methodologies described above were found to be sufficient for our evaluations in that the results provided enough insight in order to answer the research questions spelled out above.

Data description: analysis of DUO based on the QuADEM usability unit

In this section, we will provide a selection of the experimental data gathered through the concurrent verbal protocols and the post-hoc-interview. Based on these data, we drew our conclusions regarding the QuADEM assessment method that will be presented below. For a better cohesion of the following presentation of our data, we opted not to follow strictly the order of the QuADEM checklist items (cf. table 1). Instead we grouped the most important findings according to the different elements of the online learning module.

Whereas the sober and clearly arranged interface and overall navigation facilities were judged positive by both respondents, they did not agree concerning the appeal of the layout, to the extent that even the basic colours were recalled differently during the post-hoc-interview (referred to as INT in the following). S1 speaks about the “complementary colours blue and orange”, perceived as functional:

Ik vond ook de layout hoe het eruit zag aangenaam omdat – ja, ik heb er op het moment zelf niet zo veel aandacht aan besteed, dus ik denk dat dat positief is want dat heeft me niet te hard afgeleid. (INT)

Translation by the authors (in italics after the original, as all following citations):

I also found the layout, the way it looked, pleasant because –yes, I did not pay a lot of attention to it at the moment, so I think this is positive because it did not distract me too much.

S2 recalls them as “grey and orange” – which is correct – and perceived the layout as rather dull:

Ik vond hem gewoon vrij sober, het is zeker niet dat het storend was maar het kon denk ik iets meer tot de verbeelding spreken of iets meer aantrekkelijk zijn dan dat het was. (INT)

I just found it pretty sober; it’s certainly not that it would have disturbed me but I think it could be a bit more imaginative or attractive than it was.

The layout at the *task level* was perceived as less transparent. The numbering of the task steps is not distinctive enough (no indentation or font difference from the running text), which sometimes caused orientation loss when the respondents returned to the task page after completing an exercise or reading a text that opened in new windows.

The screenshot shows a task page from the DUO website. The page is titled "Aufgaben: 1 2 3 4 Studienleistungen". It contains two tasks. Task 1 asks about different types of performance certificates in Germany. Task 2 describes a seminar by Jürgens on "Entwicklungspsychologie des mittleren und höheren Erwachsenenalters". The page includes a navigation menu on the left with options like "Wortschatz", "Leseverstehen", "Hörverstehen", "Schriftlicher Ausdruck", "Mündlicher Ausdruck", and "Grammatik". A top navigation bar contains icons for "INHALT", "GRAMMATIK", "LEKIK", "TIPPS", "LINKS", "CHAT", "FORUM", "EMAIL", "NOTIZEN", "HILFE", and "ADMIN". A sidebar on the right has a "Gehen Sie jetzt zu den Übungen." link. At the bottom, there are icons for a printer and a book.

Figure 3. Screenshot of one task.

At the *exercise level*, the respondents were unhappy with time-consuming window hopping in exercises that require several windows to be open as these cannot be viewed simultaneously (e.g. the task window for background information plus the exercise window for instructions and input plus a window with the reading text plus (potentially) the dictionary window):

Daar was dat heel ambetant, ik herinner dat nog, dat je heen en weer moest springen. (S2, INT, commenting on a recording excerpt)

There it was really annoying, I still remember, you had to jump to and fro.

Nog een laatste negatieve punt dat er verschillende vensters vaak openstonden wat ook ingewikkeld was en extra tijd vraagt (S1, in the brainstorming phase of the interview).

One last negative thing to mention was that often there were several windows open simultaneously which is complicated and time-consuming.

S1 even hesitated the first time she dealt with an exercise that required window-hopping:

Maar als ik nu naar de oefening ga, dan verdwijnt die andere informatie [*referring to the background information text that opened in a separate window*] misschien opnieuw. [*she tries it out*] – ah, ok, nee ik heb alles nog. Moet dan wel constant verspringen van venster. (concurrent verbal protocol of S1 (referred to as VP1 in the following))

But if I go to the exercise now, then the other information might disappear – ah, ok, I still have got everything. But then I constantly have to hop between windows.

The exercise window itself consists of three separate frames: the instruction frame at the top, the input frame in the middle, and the answer / feedback frame appearing at the bottom when requested. The grey font is quite small, and was perceived by S2 as old-fashioned:

Van de oefeningen vond ik echt wel – die layout vond ik zelfs een beetje ouderwets, allez ook het lettertype en zo – [*hesitating*] ja, dat kan sober zijn en toch modern maar dat was eigenlijk, ja ... dat kan zeker beter (...) en het was ook zo in het grijs. Moest er een kleur bij gezeten hebben zou dat anders – totaal iets anders zijn. (INT)

The exercises, well I found that layout even a bit old-fashioned, also the font etcetera – yes, it can be sober and modern at the same time, but actually here it was... this could definitely be improved (...) It was also in grey. A colour would have made for a different – a totally different impression.

A navigation bar situated at the bottom gives access to feedback, restart and save functions, help, and submission (for exercises corrected by the tutor), and it displays straightforward symbols that were located intuitively by our respondents. When two or more exercises follow in a row, they can be accessed by arrows pointing forward / backward in the middle of the bar, accompanied by the text “X out of Y”. Although intuitively placed, this function was overlooked by both respondents the first time they needed to use it, and thus both had to be prompted to use it.



Figure 4. Screenshot of one exercise.

The *feedback* colours are green and red, and they did not arouse confusion. In some cases, though, they provoked an emotional reaction because the respondents seemed to physically “feel” at one glance they had failed or done a (surprisingly) good job. However, this emotional reaction was not directly related to the *layout* of the feedback. We will refer back to this important observation in the discussion section. The simultaneous use of both colours for the feedback on one single answer item showed to be counterproductive. It indicates that, although correct, there is still a missing part in the answer, e.g. in exercises with multiple correct answers or many-to-one relationships. Both respondents misinterpreted this mix of colours in an exercise where the many-to-one relationship was not explicitly mentioned in the instruction and tried to “correct” the assumed error, finally giving up on finding the correct solution for the exercise.

As both respondents were first-time users of the environment, our specific concern was with self-explanatory usage. DUO offers extensive explanation about the built-in functionalities in the “hints” and “help” sections accessible via the respective links in the upper navigation frame. In general, though, website and software users do not have the patience to read pages of information before actually starting to use the product. The same applies to users of a virtual language learning environment, who expect to find their way through it by using their intuition, only occasionally referring to the help section when needed. However, S1 decided to prepare her first experience browsing the online manual. In the post-hoc-interview she stated that otherwise she possibly would have got lost in the navigation structure:

De verschillende balken, links, rechts, bovenaan... ik was blij dat ik op voorhand de handleiding had gelezen, omdat ik dan wist waar moet ik naartoe als ik hulp nodig heb of zo – ik weet niet of dat anders zo intuïtief gegaan zou zijn. (INT)

The different bars, left, right, on top, ... Luckily I had read the manual beforehand, because like this I knew where to look for help and so on – I don't know whether otherwise it would have worked so intuitively.

S2 adds to this:

Het is inderdaad niet zo duidelijk als je iets zoekt waar je het precies gaat vinden. (INT)

Indeed it's not very clear where you can find the things you are looking for.

Nevertheless, S1 concludes that the overall navigation structure is good:

Er zit wel een duidelijk systeem in, een logica, maar als je er de eerste keer mee werkt dan dringt dat niet door. (INT)

There is a clear system behind it, following a logic, but when you use it the first time, you just don't realize it.

DUO responds to the attitude of a spontaneous first-time user providing screen capture videos that show the functioning of the various exercise types. Those can be accessed directly from the exercise by clicking a “help” button at the lower left corner of the window and proved to be an effective and efficient help that both respondents located intuitively (though they only started looking for it when prompted). The video was especially helpful to S2 in her first drag-and-drop exercise. After she had tried repeatedly to drop an element without success, she was prompted to use the help-function:

*[reading aloud] 'Animation' – ja, dat is altijd wel plezant natuurlijk [starts to watch the video] Ah, ik moet misschien **wachten** tot dat ik zo een ballonnetje heb! (VP2)*

*Animation - always a fun thing, of course. - Ah, I just might have to **wait** until I get such a little balloon!*

Dropping requires great accuracy in DUO, which is helped by means of a parachute symbol. Both respondents needed some time to understand how to exactly drop the elements. The same applies to gap filling. The gap turns blue when a specific spot is clicked and ready to be filled / manipulated. S1 had trouble using this functionality, often unintentionally erasing whole gaps instead of one character or leaving the spots that indicate the gap position, which caused negative feedback. S2 did not work on any exercise that required gap filling.

Some functionalities, however, could not be found intuitively, such as the access to the forum, which both respondents expected to find in the running text of the task – in analogy with the access to exercises and additional resources – instead of the upper navigation frame. Another issue worth mentioning is the input of the special character “ß”. Unlike most language learning software programs, DUO does not provide a shortcut for the quick input of special German characters. The respondents did not think of looking for help in the respective section (“Hilfe” in the upper navigation bar, where the key combinations are listed); instead they searched for a shortcut in the editor menu (appearing in open question exercises). S1 eventually decided to replace the “ß” by double-s, receiving negative feedback as a result.

The last of the ancillary components of DUO we want to mention is the built-in lookup dictionary. It is linked to both an internal word list and an external online dictionary, DWDS. This renowned and reliable free online dictionary for German not only provides extensive definitions but also context examples and frequent collocations. The amount of information clearly caused cognitive overload to our respondents, who were expecting a quick translation or a synonym in order to be able to continue the task or reading process:

Je moet echt veel zelf zoeken. Het zou beter zijn moest je een korte uitleg krijgen in plaats van al zo een gans [moving her hand upwards down] – go ja, dat is wel normaal, het is een woordenboek, maar ik weet niet, het is niet zo duidelijk. (VP2)

You really need to search a lot yourself. A short explanation would be better than such a whole – well, ok, this is normal, it’s a dictionary, but I don’t know, it’s not so clear.

S2 even was reluctant to use the dictionary a second time when she was prompted to do so in the description of a specific exercise:

*Ah ja, maar dat is weer met die woordenboek, ik ga dat gewoon **zo** proberen. [without the dictionary] (VP2)
Ah yes, it’s that dictionary again; I’ll just try it **like that**.*

This emphasizes the importance of tracking students' behaviour in order to shed light on the "true instructional effectiveness" of ancillary components (Fischer, 2007).

Regarding *technical usability* several instability issues had to be dealt with during the experiments. The problems ranged from poor visualisation over failed / incorrect loading of exercises or tasks to a sudden blockage of the DUO environment that could only be helped by restarting the computer. Naturally, those events caused indignant reactions. S2 even engaged in a spontaneous dispute with the computer. After receiving a "loading failed" message she decided to re-login and was puzzled by the fact that she was not redirected immediately to the login-page after logging out:

Zo **raar!** Daarnet ging dat toch! Zo vreemd...Moet ik anders nog eens proberen uitloggen? (...) Is het hier? 'Abmelden', ja - (...) Ja, maar dan moet ik wel terug aanmelden, hé jongens! (VP2)

*But this is so **strange!** A minute ago it did work! That's weird... Should I try to logout again, perhaps? (...) Is it here? 'Abmelden' [German for logout], yes. (...) Yes, but then I have to login again, hey, guys!*

At the end of the post-hoc-interview the respondents were asked to rate a statement concerning the overall usability of DUO ("Ik vind de leermodule gebruiksvriendelijk."¹²) on a 5-point Likert scale, and to give a short explanation on their judgement. The average is 2.5 (S1: 3, and S2: 2). Both agreed that the need to look for usage information when dealing with content was demotivating:

Ik moest nog een beetje mijn weg zoeken. Alles ging nog niet heel vlot. (S1 in INT)

I had to try and find my way round; it did not go all that smoothly.

This is a frequently mentioned issue in courseware evaluation literature, cf. e.g. Murray & Barnes: "Despite the apparently high levels of ICT skills amongst many pupils, learners must be *learning the language* when they are using the software, not struggling with how to use the software." (Murray & Barnes, 1998, p. 255).

Having browsed our data in view of the corresponding QuADEM *usability* checklist items, the question arises whether we found our results to suffice for a thorough analysis of the usability of DUO. An answer to this question will be given in the following section.

¹² I find the module user-friendly

Discussion of the analysis results regarding completeness

QuADEM has some qualities that make the method accessible also to non-experienced third-party-evaluators, amongst which the clear, short and yet exhaustive rationale behind each checklist criterion, and the methodological compendium. The rich observational data that was revealed in the results section using QuADEM's *usability* checklist also indicate that the six criteria of this list cover some crucial factors. Nevertheless, after the evaluation of the DUO module implementing the QuADEM *usability* criteria, we were left with two remaining elements that had drawn our attention but that we could not place in the analysis framework: *poor feedback* and *poor task design*.

In QuADEM *feedback* and *task design* are listed in the *testing* unit (meant to assess the methods which the digital learning module uses to evaluate learner performance), not in the *usability* unit. Nevertheless, in our analysis they appeared to be the stumbling blocks in exercises that functioned well on a *general* and *technical* usability level, but still irritated the respondents, sometimes even to the extent of abandoning the exercise.

The QuADEM quality criteria for tasks are “authenticity” and “relevance” and they are described in the “explanation” sections as follows:

Authenticity is a rather subjective concept. Being “real” does not automatically constitute an authentic task. Tasks should be as situationally authentic as possible, which means that the situation referred to in the task should have meaning or should seem authentic to the individual user. (...)Tasks should also be as interactionally authentic as possible, which means that the interaction should reflect a real-life interaction as closely as possible. (Opdenacker et al., 2009, p.69).

For a task to be motivating it needs to be relevant within the rest of the content. There should be a clear link between the test and the module/course content. (ibid.)

Regarding feedback, the only QuADEM suggestion is that it should be “given in a non-threatening way” (Opdenacker et al., 2009, p. 70):

If feedback is limited to a numerical score, it might be perceived as threatening or impersonal. Personalised feedback and suggestions for improvement can make feedback less threatening and more personal. After some training students (peers) have been seen to be able to rate and give peer feedback. Peer rating is generally perceived as less daunting than tutor rating. (ibid.)

Although important, we found these criteria unsatisfactory for our analysis. As we will show in the following, we encountered poor task design and poor feedback leading to considerable loss of motivation even when dealing with “authentic” and “relevant” tasks from a content perspective. Regarding feedback, our findings show that its

potentially threatening character plays indeed an important role; but this question goes far beyond the main factor explicated in the QuADEM manual, viz. numerical score vs. personalised feedback. The following observations illustrate our findings.

In several instances, both respondents were negatively surprised by feedback. This occurred when they had actually given correct answers that showed only slight alterations with respect to the pre-programmed solutions, such as remaining dots in a gap, a typo, differently ordered (though equally possible) answer elements, or even a mistake in the sample solution. As it is always the whole answer that is coloured red, and not only the erroneous parts, the first reaction was total disappointment. Only when comparing their answers with the sample answer did they find out that minor changes provoked the negative feedback. Our respondents reacted emotionally to this feature, apparently perceiving the feedback as offensive, which is reflected in an indignant dispute with the imaginative “correctors”:

OEI! Ja, maar ze zetten er niet bij wat ik er **wel** moet invullen (...) Had ik dat er nu niet staan? Is dat omdat er die puntjes daar nog stonden? Maar dat is wel streng hoor! (...) Ja, dat vind ik nu wel, dat ze daar wat minder streng op moeten zijn (VP2)

*Whoops! Yes but they don't tell me what I **should** write (...) Was this not what I actually wrote? Is this because of the dots that I left there? But this is very strict! (...) Yes, I really think they should be less strict about this.*

AH, ik was maar die '1' hier vergeten [*finding the not language-related detail that caused negative feedback*]... Ik vind dit sowieso niet zo'n nuttige oefening, eigenlijk (...) Ik denk dat er hier verschillende goede oplossingen zijn [*correct observation*] maar het systeem keurt er maar één enkele goed. Ik weet niet of dat zo een goede benadering is, didactisch gezien, dan. (VP1)

*Oh, I just forgot to put the “1” here... I don't find this a very useful exercise, anyway (...) I think there are several options here [*correct*], but the system accepts only one... I don't know whether this is a good approach, didactically speaking.*

Although this is an inherent problem with automatic feedback, there are ways to avoid user demotivation by programming the feedback in a more flexible way, adjusting the failure tolerance according to input requirements, or highlighting only the *wrong* part of the answer in order to make it appear less threatening. Another issue regarding feedback was the lack of explanation on potential error sources. The users are left alone to compare their answers with the model answers, which can be time consuming and counterproductive, especially when minor differences cause the negative feedback (cf. also main findings from user walkthroughs in Hémard, 2004).

Het werd gewoon groen of rood en dan moest je zelf maar zien wat je ermee zou doen. In de tekst was het dan vaak wel zoeken 'wat heb ik fout gedaan' en uiteindelijk kon ik toch zelf meestal wel afleiden 'waar ligt het aan', maar het was leuk geweest als – bv. als je toch zo een tekstje moet schrijven, dat er dan niet zo – één vaste oplossing komt en dan met jouw tekst niets gedaan wordt. Al die moeite die je er dan instopt gaat dan verloren. (S1, INT)

It just turned green or red, and then you had to find out by yourself how to deal with it. In the text, this meant you had to look for what you did wrong. In the end, I was usually able to detect the problem myself; but it would have been nice if – for instance, if you have to write a text, that there is not one fixed answer, and nothing is done with your text. All the effort you invest in it gets lost then.

With *poor task design* we refer to exercises whose structure seemed counterproductive to the intended learning objective or whose learning objective did not become clear to the respondents at all. We observed reactions showing bewildering to this respect on several occasions. The utterance “Ik vind dit raar/vreemd”¹³ referring to a task element (but not to the content) was counted eight times in SV2 and four times in SV1. S2 even stated clearly in her concluding remark:

Soms waren er wel rare oefeningen bij” (INT).

There were some strange exercises in it.

To illustrate this with one representative example we will describe S2’s reaction to the final exercise of a task on the German study cycle where the user has to substitute certain words with synonyms in an informative text. No hints are given as to the required synonyms (some of them belonging to an elaborate register), and there is no obvious connection with the vocabulary newly acquired during the task or with the previously read texts. S2 spent a lot of time browsing the latter, convinced that she would find the synonyms there.

Wat? Ik moet dus gewoon synoniemen **verzinnen** voor die woorden? ... En geen lijst om te kiezen? ... En dan nog **zó’n** tekst! [*referring to the difficulty level, raising her eyebrow with indignation*] (...) Wat is **daar** nu het nut van? (...) [*answering her own question*] Ik denk dat je de woordenschat leert kennen die vaak voorkomt in die context. Maar ja, er zijn er zo veel mogelijkheden, en je kan ze **perfect** uit de context begrijpen, ookal ken je er geen synoniem... (VP2)

¹³ I find this strange

*What? I just have to **invent** synonyms for these words? ... I cannot **choose** from any list? ... And then **such kind of a text!** (...) What's the use of **this?** (...) I guess you get acquainted with the vocabulary that is frequently used in this context, but yes, there are so many possibilities and you can understand them **perfectly** from the context, even if you don't know a synonym.*

She finally abandoned this exercise, totally demotivated. There might be a good didactic reason for the structure of this exercise, but the crucial point is that the user needs to *understand* it in order to be motivated enough to work on it. Instructions play a key role here. A lack of clear instruction can lead to misinterpretation, motivation loss, and it can even encourage the learner to abandon the exercise. It is remarkable in this respect that the statement “Ik vind de taakbeschrijvingen duidelijk.”¹⁴ was the only one that received a unanimously low rating (two times 2 out of 5) in the post-hoc interview.

It is not by chance that feedback and task design are the two key issues we identified as critical in our observations. Both seem to be intrinsically correlated: “feedback (...) appears to have the most impact when goals are specific and challenging but task complexity is low.” (Hattie and Timperley, 2007).

Conclusion

To sum up, we would like to get back to the research question of this preliminary analysis: is QuADEM a reliable assessment tool for an online language learning environment? We aimed to investigate this question by picking one out of the twelve assessment categories (units) proposed by QuADEM, i.e. *usability*, as an exemplary showcase.

As we have elucidated in the tool description section of this article, QuADEM has very useful features to facilitate a thorough evaluation, including the rationales and methodological suggestions it provides. The *usability* checklist that was used for our exemplary analysis of DUO was found to cover a number of crucial factors. Nevertheless, our discussion of other critical instances in our observations regarding usability shows that these six criteria are not exhaustive.

As mentioned above, in QuADEM *feedback* and *task design* are listed in the testing unit. In the light of task-based learning, which has gained momentum in SLA and CALL practice, we suggest that feedback and task design should not be subparts of a separate *testing* unit but figure crucially in the *usability* unit, adding a new layer besides technical and general usability. As both are important elements in (language) didactics, we would suggest calling this new layer *didactic usability*. With this broader interpretation of usability, we build upon the statement made by Kukulska-Hulme and Shield that “there

¹⁴ I find the task descriptions clear.

are areas of usability relating to language learning websites that are currently under-researched (...) issues that may (...) be discipline-specific” (2004). They propose the term *pedagogical usability* for issues related to language use (first vs. target language) and to the organisation of content and communication in a multi-modal structure. *Pedagogical usability* would thus be situated on a structurally higher level than the aspects we propose to subsume under the denomination *didactic usability*. Good didactic usability would then be the outcome of a cognitive approach to design as postulated by Plass (1998), who says it is the only approach “that puts both the user and the learning task in the center of the design process” (Plass, 1998, p. 43). Put differently, we identified a compatibility problem between the evaluation method and the assessed courseware, a problem described by Decoo (1984): “questions (...) which are of utmost importance for a specific courseware have not been foreseen in the evaluation form”. However, in this case, the questions are foreseen, but do not get the attention they deserve in our view, due to their placement in the *testing* unit. Furthermore, the criteria descriptions of *task design* and *feedback* are not exhaustive enough to embrace all related problems that were identified in our evaluation.

This leads us to conclude that task design plays an important, perhaps even a crucial role regarding motivation in online learning modules. We therefore suggest that any assessment of online learning materials that takes motivation seriously should not overlook nor let the assessor decide of the relative importance of feedback and task design, since these are essential components in online learning. With regard to QuADEM, we would therefore suggest to reconsider the *pick and choose* approach, and to include at least some indications as to crucial assessment units that should not be left out in a thorough analysis of digital online material.

Biographical notes on contributors:

Carola Strobl is Research Assistant at the Department of Translation Studies of University College Ghent. She has a teaching career as a German lecturer at universities in Germany, Italy, Portugal, and Belgium. In combination with experience in language technology, this has led to her interest in Computer-Assisted Language Learning. She is currently writing a dissertation on a comparative study of instructivist vs. constructivist design of online language learning modules.

Geert Jacobs is Associate Professor at the Department of Linguistics of Ghent University. His research focuses on the study of professional discourse in a linguistic pragmatic perspective. In addition, he has co-ordinated several national and international projects in the field of foreign language e-learning.

References

- Baumgartner, P. (2002). Didaktische Anforderungen an (multimediale) Lernsoftware. [didactic requirements for (multimedia) learning software]. In Issing, L.J. & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (pp. 427-442). Weinheim: Psychologie Verlags Union.
- Chapelle, C.A. (2001). *Computer Applications in Second Language Acquisition. Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.
- Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal* 23, 477-497.
- Decoo, W. (1984). An application of didactic criteria to courseware evaluation. *CALICO Journal* 2, 42-46
- Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *CALL* 20, 409-442
- Garrido, P., & Geissler, C. (1997). *A methodology for software evaluation*. Retrieved from ERIC database (ED 421 140)
- Hattie, J., and Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77, 81-112.
- Hémard, D., & Cushion, S. (2001). Evaluation of a Web-based language learning environment: the importance of a user-centred design approach for CALL. *ReCALL* 13, 15-31.
- Hémard, D. (2004). Enhancing online CALL design: The case for evaluation. *ReCALL* 16, 502-519.
- Hubbard, P. (1988). An Integrated Framework for CALL Courseware Evaluation. *CALICO Journal*, 6, 51-72.
- Jamieson, J., Chapelle, C.A., & Preiss, S. (2005). CALL evaluation by developers, a teacher, and students. *CALICO Journal* 23, 93-138.
- Kukulska-Hulme, A., & Shield, L. (2004). *Usability and Pedagogical Design: Are language learning websites special?* Paper presented at ED-MEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications. June 22-26. Lugano, Switzerland. Retrieved from <http://halshs.archives-ouvertes.fr/docs/00/19/03/40/PDF/KukulskaHulme-Agnes-2004.pdf>
- Levy, M., & Stockwell, G. (2006). *CALL dimensions: options and issues in computer assisted language learning*. NY, London: Routledge
- Murray, L., & Barnes, A. (1998). Beyond the "wow" factor – evaluating multimedia language learning software from a pedagogical viewpoint. *System*, 26, 249-259.
- Niehoff, M. (2003). *Fremdsprachenlernen mit Multimedia: Anforderungen aus Sicht der NutzerInnen. Eine qualitative Untersuchung zum selbstorganisierten Lernen*. [Learning a FL with multimedia: Requirements from the users' viewpoint. A qualitative study on self-organised learning]. Frankfurt am Main: Peter Lang.
- Odell, A. (1986): Evaluating CALL software. In G. Leech and C. Candlin (Eds.), *Computers in English Language Teaching and Research* (pp. 61-77). London: Longman.

- Oliver, M. (2000): An introduction to the Evaluation of Learning Technology. *Educational Technology & Society* 3(4), 20-30. Retrieved from http://www.ifets.info/journals/3_4/intro.html
- Opdenacker, L., Vaes, S., Van Waes, L., & Jacobs, G. (Eds.) (2009). *QuADEM - Manual for the Quality Assessment of Digital Educational Material*. Retrieved from http://www.scribani.org/quadem/material_Q/handbook/Handbook.pdf
- Plass, J. (1998). Design and evaluation of the user interface of foreign language multimedia software: a cognitive approach. *Language Learning & Technology* 2(1), 40-53.
- Reiser, R.A. & Dick, W. (1990). Evaluating instructional software. *Educational Technology, Research and Development*, 38(3), 43-50.
- Roche, J. (2008). *Handbuch Mediendidaktik - Fremdsprachen*. [Handbook of Media didactics – Foreign Languages]. Ismaning: Hueber.
- Susser, B. (2001). A defense of checklists for courseware evaluation. *ReCALL*, 13, 261-276.
- Thomé, D. (1989). *Kriterien zur Bewertung von Lernsoftware. Mit einer exemplarischen Beurteilung von Deutsch-Lernprogrammen*. [Evaluation criteria for educational software. Including an exemplary assessment of language learning software for German]. Heidelberg: Hüthig.
- Wegele, E. (2006). Tutorielle Betreuung beim Online-Sprachprogramm uni-deutsch.de. Erste Erfahrungen aus der Praxis. [Tutoring in the online language learning environment uni-deutsch.de. An early practical experience report.] *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 11 (2). Retrieved from <http://zif.spz.tu-darmstadt.de/jg-11-2/beitrag/Wegele1.htm>

2.2 Preliminary study 2: (How) can an online module support students to write a summary from spoken input?

2.2.1 Background information on the study

This second preliminary study was carried out in March-April of 2010 and involved ten volunteering students from the second and third years of the Applied Languages study programme at HoGent with German L2. Like the first study, it was conceived as an exploratory study based on existing commercial software in order to prepare my own intervention studies using self-developed learning modules. For the second study, I chose two tasks (see appendix 2) that together cover different strategies and skills required to produce summaries from spoken input, i.e., the task that was going to be targeted in the first intervention study. In the first task, students trained listening strategies, and in the second, they wrote a short summary from written input. This task included preliminary exercises to build up vocabulary and strategies for cohesion. The participants took in total approximately 1 hour to complete the tasks, and another 30 minutes to answer pre- and post-hoc questionnaires, to train thinking aloud while engaging in an online activity, and to get acquainted with the learning environment. They were compensated for their participation with a cinema voucher. The pre-hoc questionnaire (see appendix 3) investigated their language learning strategy use with questions based on Oxford's (1990) SILL (strategy inventory for language learning), as well as their pc use in general and for language learning. For the short post-hoc questionnaire, a built-in feature of the screen-capturing software *Morae* was used, consisting of a list of ten questions to be answered on a Likert-scale. It served to elicit spontaneous reactions about their experience with the learning software directly after finishing the tasks. All interactions with the software were recorded for subsequent analysis with *Morae* software, including screen-capture, video and think-aloud protocol of the participants. The main research questions for this preliminary study were:

1. How can this complex task be broken up into instructional units to effectively support students in an online learning environment?
2. How do students go about the subtasks, using the online support? Which strategies do they (not) employ?
3. What are students' attitudes about the online support they received for this task?

The results were presented at the LASIG (Learner Autonomy Special Interest Group of the German Association for Applied Linguistics) conference in Essen and at the 3rd ICT4LL conference in Florence, which both took place in November of 2010, and a short

paper was published in the proceedings of the latter conference (http://conference.pixel-online.net/ICT4LL2010/common/download/Proceedings_pdf/IBL32-Strobl.pdf).

2.2.2 Results

2.2.2.1 Features of the module to support targeted skills and strategies

The listening task of the DUO module used for this study is subdivided into three steps. First, the students listened to the introduction of the feature in order to identify the main topic. After that, they listened to the full text twice, each listening session targeting a different type of listening strategy (skim vs. scan) with appropriate instructions. The answers to the open-ended questions were written in a separate exercise window which had to be opened by clicking on a link. Finally, a sample solution could be requested for comparison with the user's own solution. During listening, students were able to use pause, stop, and rewind functions that are straightforward to access and use.

The writing task includes a suite of preparatory exercises to train cohesive strategies and related vocabulary. However, this drill-like training is decontextualized from the writing task as they are not linked at all regarding content. Understanding of the source text, which in this case consisted of a bulleted list of key phrases, was ensured by a couple of preliminary questions. No subtasks are present that target planning of the text before starting to write or revision. The preparation for writing production is thus limited to the linguistic level, and an exercise on structuring input at a content level is missing.

2.2.2.2 Students' self-reported and observed strategy use

As strategy use is of paramount importance for the execution of a complex task and varies considerably between individual students, it was one of the main concerns in this study. Prior to engaging in the online tasks, the participants were asked to indicate on a Likert-scale ranging from 1 (never) to 5 ((almost) always) the extent to which they use listening, writing, and general language learning strategies. In the following, their self-reported strategy use (see appendix 3) will be compared with the *observed* strategy use. The mean of all scores for the self-declared strategy use in the pre-hoc questionnaire is 3.45 on a Likert scale of 5. Most individual mean scores cluster around the general average, but there are two students with a considerably lower mean score (both 3.14/5), and one student with a considerably higher score (4.07/5). The observed (online) strategy use varied considerably regarding the following aspects:

- a. Manipulation of an audio file (stop / rewind)
- b. Note-taking: Amount and level of detail
- c. Reading and following instructions
- d. Transform input into new text: Cut-and-paste vs. retype
- e. Drag-and-drop lists: follow a linear vs. a priority-oriented approach
- f. Reaction to feedback: Dealing with emotions on automated negative feedback and taking a critical stance towards one's own solution when comparing it with a sample solution

In the following, the listed strategies will be discussed in detail. An interesting difference in the use of listening strategies between individuals is related to the usage of the pause and rewind functions (list item a). The students are used to the classroom setting where the teacher alone controls those functions. In class, they listen to a whole input text twice. It is interesting to note that most students followed the same pattern when working individually in the online module, although they were explicitly told before starting the exercise that they could use the player functions according to their needs, i.e. stop, rewind, repeat at will when necessary. One interesting exception was a student who, being the only one who listened three times, used the pause and rewind buttons excessively, stopping after nearly every sentence during his third listening session. This was surprising given that he had rated his strategy use of skim-listening in the questionnaire high (4/5). On the other hand, he is one of the students who had a low mean score for overall strategy use and rated low (2/5) his use of the emotional strategy "When I fail to understand something I don't panic but rather concentrate on the things I do understand", an attitude that might account for his obsessive attempt to grasp every detail during the last listening session.

The listening task also revealed another interesting detail of the students' behaviour in the online learning environment: very few students actually *read* the specific instructions prior to each listening session (list item c). This attitude resulted in a note-taking behaviour that did not fit the task requirements. After the first (gist) listening session, they had to answer only one question concerning the main idea of the text. Nevertheless, most of the participants immediately started to write down as many details as they could grasp. In view of the second listening session, students were provided with keywords representing the main topics of the text and were instructed to write down information on these topics in the input window. Only two out of ten students wrote down the keywords on their note sheets first, leaving some space to fill in the remaining information, before listening for the second time. The lack of organisational strategies of the majority of participants may in part be due to the task structure: the task window containing the audio player and the instructions on the main steps of the task informs them only that they should listen for a second time and then proceed to the exercise by clicking a link. Only the two students who opened the exercise window *before* listening saw the keyword list there and were thus properly

prepared for the task (list item b). The metacognitive strategy "I clearly identify the purpose of the language activity; for instance, in a listening task I might need to listen for the general idea or for specific facts" of Oxford's SILL reflects this behaviour. Unfortunately, this strategy statement was not included in our questionnaire, but replaced by three rather detailed statements concerning the listening process, viz. "I try to grasp the main topic from the first few sentences" (mean score: 4 on a scale of 5) – which corresponds to the first step of the task –, "While listening for the first time, I try to identify the most important keywords without focussing on details" (3.2 / 5) – which in this case was done for the students, as the keywords were provided by the program –, and "While listening for the second time, I try to write down as many details as possible around the keywords" (4.3 / 5). It is interesting to note that the third statement has a significantly higher average score than the second one, thus matching the actual observations.

Concerning writing strategies, "planners" and "revisers" were distinguished, as described in scholarship on student writing (Torrance, Thomas, & Robinson, 1994). As mentioned above, no specific task step was provided for text planning. Accordingly, most participants did not show external signs of conscious planning, e.g., drawing an outline or thinking aloud about the text structure, before engaging in writing. This does not mean, of course, that they did not plan at all, but rather that planning exclusively occurred as a mental activity. To recognise mental planning activities, the writing flow was observed which revealed big differences between individuals. Whereas some students seemed to write in one flow, others paused a lot in order to think about the emerging text. Thus, the overall time spent on the task minus the time spent actually typing could serve as a rough indicator of planning time. However, another strategy that is specific to the online setting has to be taken into account, viz. cut-and-pasting of the source (list item d), which in this case consisted of key phrases. The extent to which students made use of this strategy varied considerably: some cut-and-pasted the whole list of key phrases before even starting to write the first sentence, others started writing their own text and "imported" the respective key phrases one after the other when appropriate, a third group (the minority) did not cut-and-paste at all but typed the whole text, including the key phrases. One student even retyped the key phrases in the input window, fully relying on her memory.

Another interesting observation related to the pre-structured input for written text production is that only three out of ten participants changed the original main structure of the key phrase lists (advantages vs. disadvantages of biofuel). Restructuring improved text flow considerably since most of the advantages and disadvantages could be combined by topic, and repetition of arguments and of linguistic connectors could thus be avoided. Concerning linguistic elaboration with regard to cohesion strategies, students reported in the questionnaire that they catered for variation in cohesive ties ("I pay attention to variation in linking words when writing": 3.5 / 5), but actually, their

final texts revealed little variation in linking words and co-reference strategies, despite the immediately preceding exercises in that respect.

Three statements about revision behaviour were included in the pre-hoc questionnaire: "When writing in the FL, I reread (I) / rewrite (II) what I have written" and (III) "I notice my own mistakes and try to correct them". While students report rereading their own text to a large extent (4/5), rewriting is a less commonly self-reported strategy (3/5). The mean score for mistake correction is 3.4 on a scale of 5. These figures coincide with the tendencies observed in the screencast recordings. Most students took time to reread their emerging texts during production but corrected only local errors, mostly of morphological or typographical nature. Only four out of ten students actually attended to the global textual level, restructuring their texts during the writing process. Surprisingly, one of the two students who reported overall low strategy use was a major exception in that respect. Not only did he take notes of the source text input on a sheet of paper, but he was also the only one to think aloud during the structuring process, while reshuffling text chunks. Moreover, he was the fastest to finish the task (10.5 min, the maximum being 26 min and the median 19 min).

2.2.2.3 Discussion and conclusions

There are two kinds of conclusions that can be drawn from the reported study. The first conclusion concerns the degree to which students' self-perceived strategy use coincides with the observed behaviour in the online module. It was found that most students seemed to have a well-developed awareness of their own strategy use, although there were some interesting mismatches. One student who had reported low strategy use in the pre-test questionnaire actually turned out to rely extensively on cognitive and memorisation strategies throughout the experiment. On the other hand, the student who self-reported a remarkably high strategy use seemed to easily lose control in the online learning module. Further research, e.g. using in-depth interviews based on stimulated recall, would be needed in order to clarify reasons for these kinds of mismatches.

Other mismatches observed are those between targeted strategy use as apparent in the instructions of the online module, and actual strategy use as displayed by the participants. This was especially noticeable in the listening task, where skim- and scan-listening were distributed in different sub-tasks. However, due to a lack of knowledge of task objective on the participants' side, their actual behaviour did not coincide with this intention. In a similar vein, the guidance provided concerning linguistic elaboration of cohesion did not yield the intended effect. Students did not transfer the linguistic means that were trained in the pattern-drill exercises to their writing production.

From these observations, further conclusions regarding the affordance of the online module for the targeted skills can be drawn. The case study illustrates the importance of clear and easily identifiable objectives in online learning modules. Task instructions

must be visible at first glance in the surface structure and should not be hidden in a clickable window. Only then can they scaffold students in the development of the targeted strategies and skills. Regarding the writing task, the lack of guidance regarding text planning led to the absence of the students' planning strategies on a content level. Most students failed to notice that restructuring the bulleted list would improve text coherence. This means that elaboration of source content needs to be explicitly trained in an online module. Finally, a lack of transfer between the pattern-drill exercises regarding cohesive ties and the immediately ensuing production task was observed. This indicates that the connection between exercises with a focus on *forms* (i.e., language accuracy) and production exercises where focus on *form* happens during meaning-centred activities (Long, 1983) needs to be clearly established through content overlap in order to facilitate transfer. To conclude, it is worth mentioning that most of our respondents seemed satisfied with their online learning experience and felt sufficiently supported in both the listening (3.9 / 5) and the writing (4.1 / 5) tasks, despite the shortcomings in the design of the tasks that were uncovered in the analysis.

The observed strategy mismatch in the listening task can also be analysed through an activity-theoretical lens. From this perspective, a contradiction between two AS' in which the participants are involved, caused the unexpected behaviour. Within the AS of the online activity, the intended objective is to scaffold the subjects in their strategy use according to subtasks and their individual needs. Accordingly, the labour is divided between the online module hinting at the strategies, and the subject collective (i.e., the learners), who have control over the audio-device (rules) in order to shape the activity according to their individual needs. The learners also control the quality assessment (rules) of their outcome by comparing it with a model. In contrast, within the AS of the classroom activity where students engage in the same task, the subject (i.e., the teacher) has an important role in the division of labour, dictating the rules of the activity to ensure the comparability of the external conditions, and in extension, of the individual outcomes, for assessment. Most participants were bound by the rules of the (classroom) AS they were acquainted with, and did not adapt to the new rules of the online AS. When transferring a classroom task to an online AS, it is therefore of paramount importance to raise students' awareness of the changed rules and conditions for division of labour, so that they can fully benefit from the advantages of a more individually-oriented task setting.

To conclude, it can be stated that strategy use varied greatly between the respondents. In an online environment, strategy use is also highly influenced by the level of online and pc literacy. Even in this relatively homogeneous group of "digital natives" (Prensky, 2001), considerable differences at this level were found. The individual differences in handling the online tasks that were observed in the present study underline the importance of diversifying the pedagogical approach in an individual online setting, as frequently suggested in the literature (Opdenacker, Stassen,

Vaes, Van Waes, & Jacobs, 2009). Students should be allowed to choose between different approaches to the task and degrees of scaffolding intensity. In order to make an informed choice, they first need to be sensitised to their own (online) learning behaviour and preferences. The importance of strategy awareness for the effectiveness of the FL acquisition process as claimed by Oxford (1990) holds true especially of individual online settings.

2.3 Conclusions from preliminary usability research

In order to investigate the online learning behaviour of students representing the target group, i.e., students majoring in German L2, of the planned intervention studies before actually engaging in setting up interventions, two preliminary studies were carried out. The online learning environment that was chosen for this preliminary research, *Deutsch-Uni Online*, is built upon "moderate constructivist" principles, and therefore covers a broad variety of different exercise types, trying to map instructivist as well as constructivist learning theories to online instructional design. Furthermore, it offers modules that specifically focus on language skills and knowledge for academic purposes. At this point, the first research question of our design-based research cycle should be recalled:

RQ1: What are *stumbling stones* for advanced language learners engaging in online tasks generally, and in summarising tasks specifically?

The answer can be given based on the analysis of these two qualitative usability studies. They revealed interesting insights in the online learning behaviour of this target group on the one hand and in design-related problems that can turn into *stumbling stones*, i.e., hamper the intended objective or learning satisfaction, on the other. With regard to the latter, the outcomes of both studies highlighted the importance of certain features of the online environment that I subsumed under the term *didactic usability*, i.e.,

- a clear task design that is transparent for the purpose of the learning activity,
- feedback provision in a clear and exhaustive, but not face-threatening way,
- avoidance of window-hopping during one and the same activity (instructions that are provided in a different window than the task input were not even read)
- close integration of preparatory exercises to build up language knowledge with subsequent content-focused tasks that require recall of this knowledge.

P2 furthermore provided insights into listening, note-taking, and writing strategies of the participants when carrying out respective online activities. It was interesting to note that students followed the patterns of classroom-based language learning activities, even when the online environment suggested other strategies, because they did not even seem to *read* the given instructions. Concerning writing strategies, different types of writers were distinguished. The lack of an exercise targeting pre-writing planning led to the production of texts with poor coherence.

The main conclusions drawn from this preliminary research in view of the design of my own planned learning environments were that

1. Students need structured support in order to build up favourable listening and writing strategies, especially regarding scan-listening and outlining, and
2. Preliminary exercises that aim to build up knowledge for further use in subsequent content-focused exercises need to be clearly connected with the latter in structured exercise suites based on content similarity.

Chapter 3

Students' attitudes towards instructivist and constructivist feedback (Intervention Study 1)

3.1 Background information on the study

The first intervention study was carried out in the framework of a language proficiency course in the second year of the bachelor in Applied Languages at the HoGent, specifically aiming at learning to summarise spoken input in the foreign language. The "traditional" approach to summarisation implemented in the course before the innovation through the use of online technologies was F2F (face-to-face) classroom teaching, highlighting problems in example summaries of selected students. The AS (Activity System) of this approach is depicted in Figure 6. The feedback process occurred via the mediation of an overhead projector or the blackboard, and a guided classroom-based dialogue. It was characterised by a strong focus on LOC problems like frequently occurring grammatical and lexical errors and content problems caused by linguistic or cultural misunderstandings. This means that the feedback provision in the course actually followed the rules of a generic language acquisition course targeting listening and general writing skills. This is also reflected in the scoring system of the exam which was based to a large extent on accuracy rate. HOCs were not an issue in the classroom sessions nor an evaluation criterion. As a result (outcome), the students were able to produce a sufficiently correct account of the source content, but their texts lacked structure and cohesion.

Thus, a contradiction can be found between the object of the course (learning to summarise) and its rules (scoring) and mediating tools (feedback provision). Apart from the fact that summarisation strategies were not explicitly taught, students also did not get enough writing practice and personalised feedback (each week, only 2-3 students handed in an example summary) to raise their self-efficacy beliefs with regard to the multiple challenges included in the task at hand. As a consequence, they suffered from an extremely high exam anxiety. There was a clear need to reshape the perception of their role, i.e., the division of labour in the writing activity. Students needed to become aware of the importance of their active participation in the writing and feedback process to improve their writing, instead of passively taking in the weekly feedback provided by the teacher on the basis of examples.

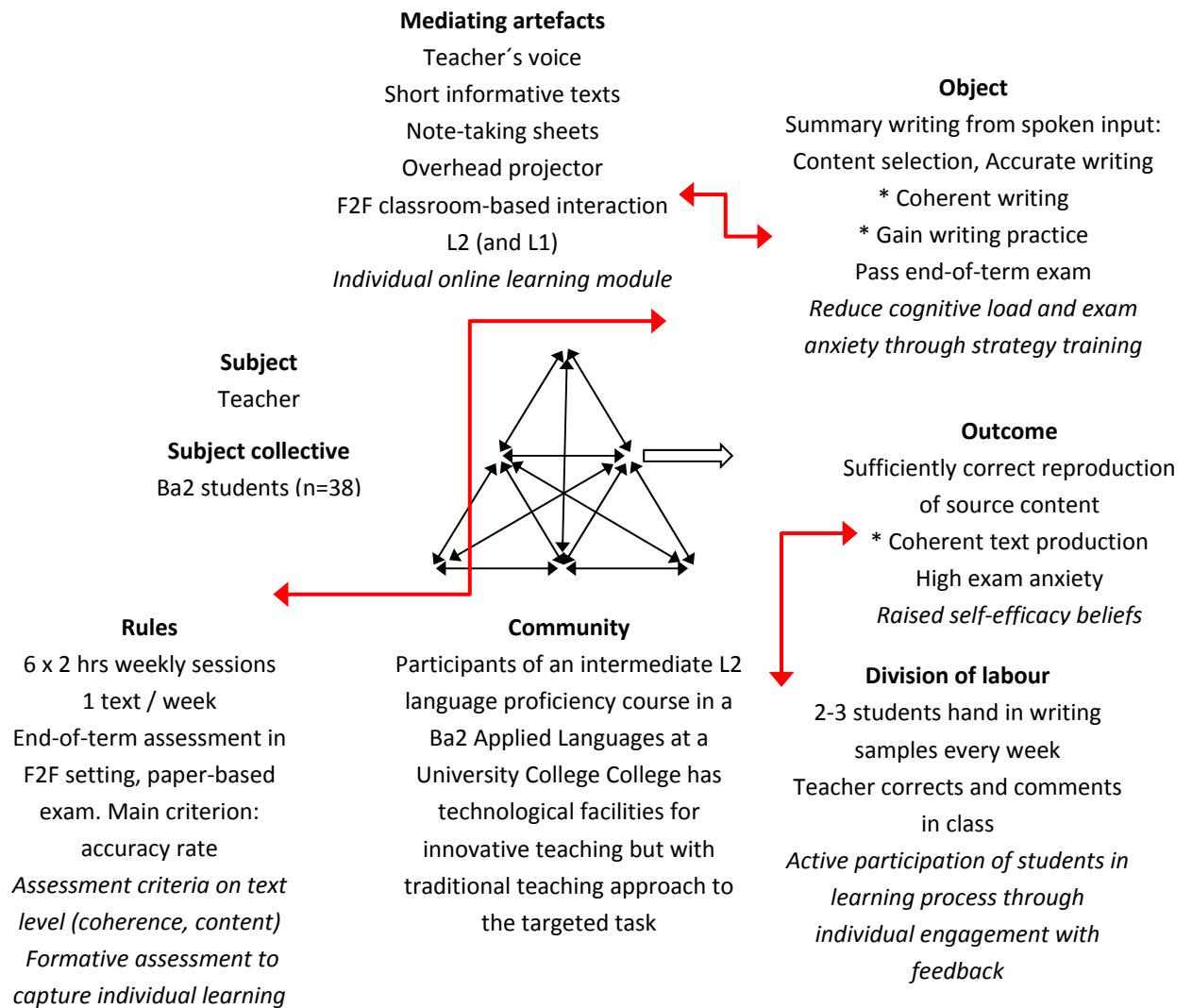


Figure 6 Activity system of F2F summary teaching prior to IS1, including planned innovations (in italics)

The mentioned contradictions are represented in the figure by red two-sided arrows. Following Lantolf and Thorne's (2006) graphic representation for innovation in an AS, outcomes that would be furthered through changes in the system, i.e., implementing an individual online module, are marked in italics. In the same vein, necessary changes in the learning object, the rules and the division of labour, are italicised. New prospective objects and outcomes that were only hinted at in the F2F classroom activity are marked by an asterisk.

The main research questions of this first intervention study were the following:

1. Is it possible to create a better learner fit, and consequently, better learning outcomes, for the instruction of this complex task through an individual online module (as compared to classroom-based F2F instruction)?

2. How do students perceive the learning gain through the online module, more specifically, the closed exercises with automated corrective feedback used in the preparatory phase of the task, and the open-answer exercises with model solutions used in the task execution phase?
3. Does this coincide with their actual learning gain as measured in their "noticing" and in their overall performance in summary writing?

To investigate these questions, an intervention study was set up (n=38) for which four learning modules were created in *curios*, an authoring system for the creation of online tests and exercises that was developed at UGent and is accessible through the institution's VLE (see appendices 4-7). In the creation of the modules, I implemented some of the principles observed in the DUO learning environment (moderate constructivist approach combining different task and feedback types, building up tasks as a suite of interconnected exercises), at the same time trying to avoid the observed pitfalls (lack of clarity regarding task intention, lack of contextualisation of exercises targeting focus on forms, face-threatening negative feedback). Both listening and writing skills were targeted in the exercises, breaking up the complex task into small, digestible units in order to manage the "overwhelming cognitive load" (Kirkland & Saunders, 1991), and to reduce in this way students' anxiety when engaging in this task. The skills training (listening strategies, note taking, content selection for summary) was combined with contextualised closed exercises focusing on form (linking words, linguistic chunks for summarisation). Furthermore, students were prepared for the listening task by content-related exercises in the form of webquests or similar activities that required their active engagement. This was deemed necessary in order to stimulate their interest in and build up knowledge about the broad array of topics that were covered in the online modules. These topics reflected discussions with a broad media coverage in German-speaking countries at the time of the intervention (entertainment industry, biomass energy, social media architecture, Swiss democracy system).

Regarding data collection for analysis, a mixed-methods approach was adopted. Quantitative data were gathered to measure learning gain (pre- and post-tests on cohesive strategies, logged data of all activities in the online modules including scores achieved in closed exercises), changes in strategy use (pre- and post-hoc questionnaires) and attitudes (three questionnaires administered during the six-week intervention). Qualitative data consisted of pre- and post-tests on summary writing, answers on open-ended questions (self-evaluation of open-answer exercises, questionnaires on attitudes towards the exercises completed, post-hoc focus groups on attitudes towards the online module).

The results obtained from the data analysis provided the following insights:

1. As the qualitative analysis of individual pre- and post-test revealed, students had a clear learning gain regarding higher-order concerns (content selection, structuring the summary, coherence, rephrasing, avoid lexical recurrence).
2. As the qualitative analysis of their (guided) reflection in the open-answer exercises with model-solution showed, the quality of noticing is accordingly high.
3. However, the quantitative and qualitative analyses of their attitudes towards the exercise types revealed that they prefer clear-cut corrective feedback to model solutions.
4. The latter is due to a mismatch between their perceived role as a learner (passively consuming corrective feedback) and the necessary change in this role when receiving feedback in form of model answers.

The results of IS1 were presented in two papers at the Junior Research Meeting 2011 of the GAL (Gesellschaft für Angewandte Linguistik) in Essen, Germany, and at the EuroCALL 2012 conference in Gothenburg, Sweden, and in an article for the international double blind peer-reviewed journal *ReCALL* which is listed in the Web-of-Science (impact factor: 1.226) (see following section).

3.2 Attitudes towards online feedback on writing: Why students mistrust the learning potential of models (Article accepted for publication in *ReCALL*)¹⁵

This exploratory study sheds new light on students' perceptions towards online feedback types for a complex writing task, summary writing from spoken input in a foreign language (L2), and investigates how these correlate with their actual learning to write. Students tend to favour clear-cut, instructivist feedback over constructivist, guided self-evaluation through model solutions in online learning environments. However, the former type is too limited to tackle all dimensions of advanced writing. Constructivist feedback in the form of guided modelling allows to address the higher-order concerns involved in summary writing. In addition, it is widely acknowledged that activating the zone of proximal development (ZPD) through cognitive involvement is beneficial to learning. To investigate students' learning from both types of feedback, a one-group pre-post-test intervention study was set up. Students attending a course on summary writing in L2 of a bachelor program in Applied Languages (n=38) followed an individual online learning module containing both instructivist fill-the-gap exercises and model solutions with constructivist guiding questions for self-assessment. Their actual learning gain was measured through pre- and post-tests, and compared with students' perceived learning gain, as expressed in self-evaluation. The comparison reveals a dichotomy between the students' promising learning curve and an underestimation of their own progress. This dichotomy was found to originate in a mismatch between their expectations towards the online learning module and the characteristics of the constructivist feedback conveyed. This mismatch can be attributed to three key factors: (1) evaluation, (2) linguistic focus, and (3) learner motivation.

Keywords: Constructivist feedback, Learner attitudes, Summary writing, Modelling, Post-secondary education

Introduction

Feedback plays a vital role in online learning and is inextricably linked to CALL design principles (Felix, 2005; Heift, 2010). Felix (2005) states an "about turn" in the provision of feedback in CALL due to the focus on constructivist design principles at the beginning

¹⁵ Strobl, C. (accepted): Attitudes towards online feedback on writing: Why students mistrust the learning potential of models. *ReCALL* [expected print publication in issue 27(3), Sept. 2015]

of the millennium. However, she holds that behaviourist pattern-drill should not be abandoned altogether, mainly on account of its time-efficiency. According to Kirschner's model of educational design, decisions about the appropriate design of an online intervention should primarily be informed by knowledge about the learners' background and by the learning goals (Kirschner, 2002). An online module that aims to foster summary writing from spoken input in a foreign language needs to combine different instruction techniques and, accordingly, forms of feedback. The task is complex and challenges learners on several levels, as it involves different skills (listening and writing), techniques (note-taking), linguistic and cognitive demands. Moreover, feedback on summary writing needs to address both lower-order concerns (LOCs), i.e., problems on word- and sentence-level, like accurate choice of cohesive ties, and higher-order concerns (HOCs), i.e., problems on text level, like content selection and coherence. The latter demanding cognitive ability cannot be trained by means of pre-formatted instructivist feedback.

Following the Cognitive Mediation Framework (Doyle, 1977), no direct relationship can be assumed between instruction, including the provision of feedback, and actual learning results. Several factors play a mediating role, among which the learners' perception of the instruction provided. In the same vein, Norris and Manchón (2012) claim that writing development is mediated by different individual factors like learners' goals, beliefs, attitudes, and learning histories. Therefore, in order to assess the effectiveness of an intervention in learning-to-write, it is important to triangulate data related to process, product, and students' attitudes in order to ascertain the role of feedback in the learning process (*op. cit.*:238).

Below is a brief review of literature related to feedback in online learning and feedback on L2 writing. Due consideration is also given to previous investigation of learners' attitudes towards feedback.

Feedback in online learning

In the area of CALL, research into the effectiveness of different forms of feedback has come to diverging conclusions (Felix, 2002; Heift, 2005; Robinson, 1991; Rosselle, Sercu & Vandepitte, 2009). Robinson (1991) concludes from her literature review that internal detection of errors and the learner providing the answers are more effective than the external provision of answers through program disclosure. Felix (2002), however, holds that pattern drill exercises are more appropriate and time-efficient to foster accuracy. Heift (2005) and Rosselle *et al.* (2009) found that students' uptake, i.e., the response to corrective feedback, for grammar and vocabulary-related exercises, was higher when the feedback contained metalinguistic information than in the case of recasts or error localisation only. The only CALL study to my knowledge that includes a focus on HOCs for summary writing is Kintsch *et al.* (2000). The authors discuss the affordances of educational software based on latent semantic analysis for this instructional goal. They

report only partial success, pointing to the fact that, apart from feedback, students need explicit strategy training and modelling. So, a combination of pre-task strategy training and constructivist feedback in form of model provision could be a promising route.

Feedback on content plays an important role in learning to summarise, but has not been a focus of CALL studies until now. Therefore, this brief literature review on feedback effectiveness also considers relevant studies in general online pedagogy. Research on the impact of online feedback on students' learning in other areas than CALL also leads to different conclusions (Mandernach, 2005; van der Kleij, Eggen, Timmers & Veldkamp, 2012). According to van der Kleij *et al.* (2012), an important factor in explaining the different findings is the level of intended outcomes. They propose a comprehensive framework for online feedback classification based on Hattie & Timperley (2007) and Shute (2008), integrating the feature "targeted level of outcome" with the traditional description levels "content" and "timing". Regarding timing, feedback can either be delayed or immediate. On a content level, they differentiate between "knowledge of results" (KR), "knowledge of correct response" (KCR), and "elaborated feedback" (EF). The targeted outcome of feedback can be at the "self" level (e.g., learner characteristics), "task" level (e.g., knowledge building), "process" level (e.g., a worked-out example), and "regulation" level (e.g. self-assessment). As constructivist pedagogy focuses on the learner's process level and cognitive involvement, feedback associated with this approach provides elaborated information about the targeted response. Instructivist pedagogy, on the other hand, seeks to automate discrete items of a task, and therefore provides immediate KR without elaborating on the correct response. In their experiment on the effect of feedback in online learning in a content-related course, van der Kleij *et al.* found that, contrary to the initial hypothesis, the students in the KCR+EF condition, which aimed at task and process levels, did *not* outperform the students in the KR only condition. Feedback timing turned out to play an important role in their study, as students paid more attention to KCR+EF feedback delivered immediately after each response than to delayed feedback. Mandernach (2005), who also found that the various types of computer-based feedback implemented in his study (KR, KCR, elaborated feedback in the form of topic-contingent and response-contingent information) did not impact learning to a significant degree, concludes that

“there is no clear-cut ‘best’ type of feedback in computer-based instruction for all learners and learning outcomes (...) [W]hile computer-based feedback may help to clarify simple, definition-based errors, it may be less effective in correcting more complex errors in understanding. In addition, research indicates that student understanding is enhanced more through the application of relevant examples than through repetition of basic information” (*op. cit.*: 3)

The use of feedback in the present study reflects this insight: Computer-based (automated) feedback was used to create and broaden students' knowledge base concerning discrete language items, whereas models were provided as "relevant examples" for more complex problems like selecting content and creating a coherent text.

A potential problem concerning KCR+EF feedback stated by CALL and non-CALL scholars is the observation that students spend no or little time examining it (Pujolà, 2001; Heift, 2010; Mandernach, 2005). Heift and Pujolà relate this behaviour to the observation that students are satisfied to know their result or the correct response, and do not see the need for delving into the elaborate response. Van der Kleij *et al.* (2012) point out the importance of attitudes and motivation for the time spent reading the feedback, stating that feedback can only be effective when "the learner is willing and able to use [it]" (van der Kleij *et al.*, 2012:265). As Mandernach states, "[t]his failure of students to have utilized the computer-based feedback likely accounts for the lack of learning gains in response to the various forms of feedback-elaboration" (*op. cit.*: 9).

Feedback on (L2) writing: The role of models

A major concern of research on written corrective feedback is the question whether error correction has a positive impact on writing performance, and to what extent specific feedback features, like degree of explicitness and focus, can play a role in the development of accuracy (see van Beuningen, 2010, for an overview). However, for the present study, these questions are of less importance. Error correction only occurred in the form of predefined computer-provided feedback in exercises targeting restricted language use. Instead, feedback on the written summaries was provided in form of a model solution, being the "most extreme form of indirect feedback" (van Beuningen, 2010:12). After all, the main learning goal was to raise students' awareness of HOCs in writing, i.e. content selection, rephrasing, and coherent text structure, being the main challenges in summary writing.

Models as cognitive-constructivist form of feedback can help stimulate the learners' ZPD through noticing the gap between their own performance and a target-like performance. The theoretical foundations for the use of models as feedback on L2 writing can thus be laid in sociocultural theory in combination with the concept of "noticing" (Schmidt, 1990). "Noticing" refers to learners' awareness of "a mismatch or gap between what they can produce and what they need to produce, as well as between what they produce and what target language speakers produce" (Schmidt, 2001:6). That is, the first step of "noticing" takes place during production, and the second step during comparison of the own production with a model. Martínez & Roca de Larios (2010) and Hanaoka (2007) report on the use of models as feedback for a narrative writing task based on picture stimuli. According to the task type and relatively low proficiency level of the learners involved, they found that noticing occurred mainly in the field of

lexicon, and to some extent in the field of ideas and expression. Hanaoka also reports that noticing depends largely on proficiency. This is in line with Aljaafreh & Lantolf (1994), who state that the potential relevance of feedback depends on “where in the learner’s ZPD a particular property of the L2 is situated” (*op. cit.*:480).

To date, research on the use of models in L2 writing is scarce, and the focus has been restricted to LOCs. To the best of my knowledge, there has been no CALL study that reports on the use of models as feedback. In this respect, the present study contributes to both fields of research. On the one hand, it broadens the focus of L2 writing research on the use of models, and on the other, it introduces the use of models in constructivist CALL pedagogy for advanced writing.

Learners’ attitudes towards feedback

While there is a broad range of quantitative studies on uptake and effectiveness of feedback types in computer-based learning environments, few studies have investigated the user perspective qualitatively, taking students’ attitudes into consideration (Cotos, 2011). This is in stark contrast with the impact that attitudes have on the uptake, given that “feedback is not necessarily a reinforcer, because [it] can be accepted, modified, or rejected” (Hattie & Timperley, 2007:82, drawing on Kulhavy, 1977). As early as in 1977, Kulhavy related the importance of research on students’ perceptions of feedback to the variety of feedback provision made possible in computer-assisted learning: “[B]ecause computerized instruction allows such a wide range of strategies for each response, the question of how one most effectively matches feedback parameters with response characteristics is indeed an important one.” (Kulhavy, 1977:224). Kulhavy singles out learners’ confidence in the response as an important factor impacting on the attention spent on, and the effect of, feedback.

Most CALL studies attending to feedback attitudes deal with grammar or vocabulary instruction. Nagata (1993) found that students’ satisfaction with feedback depended on its degree of sophistication. She implemented a CALL program for the training of a complex Japanese grammatical construction, providing both pre-programmed feedback based on pattern matching, and feedback produced by an intelligent tutoring system using natural language processing. The participants preferred the latter, because it explicitly guided them towards finding the correct answer, providing a lot more contextualised EF. In the same vein, van der Kleij *et al.* (2012), who investigated how students perceive the usefulness of formative computer-based assessment found that the attitudes were significantly more positive when students received elaborated feedback instead of corrective feedback (KR) only. This also coincides with findings of Rosselle *et al.* (2009), who studied the effect of five different feedback types in an L1-L2 translation activity. Her learners rated the usefulness of feedback that offered both diagnosis and guidance higher than simple KCR. However, feedback indicating the location and providing a meta-linguistic clue *without* revealing KCR did yield a lower

score, because students felt uncertain whether they had interpreted their errors correctly.

Also in the field of L2 writing, several researchers (Hyland, 2003; Norris & Manchón, 2012; Storch, 2010) assert that it is important to focus on learners' attitudes along with the effectiveness of the feedback provided. The findings of Hyland's (2003) case study confirm that L2 students value form-focused feedback. Diab (2005) reports the same result from a questionnaire study in which university ESL students "overwhelmingly (90%) agreed (...) that it is important to them to have as few errors as possible in their written work" (*op. cit.*: 30). Radecki and Swales (1988) added to the discussion by developing a typology of post-secondary students' behaviour in ESL writing instruction, depending on their perception of the teacher's role. Differentiating between "Receptors", "Semi-resistors", and "Resistors", they found that students' expectations regarding the focus of feedback on writing (surface error correction vs. rhetoric comment) largely depended on that perception, as well as on the course attended (English language classes vs. courses in a study discipline where the language plays a subservient role). Enginarlar (1993) investigated students' attitudes towards multiple-trait formative feedback, including indication of error location plus metalinguistic gloss and a summative evaluation on paragraph as well as text level, but not providing KCR. Next to stating a positive overall evaluation of this feedback system, Enginarlar also found that "when feedback (...) is provided in a problem-solving manner, students seem to regard revision work as a collaborative type of learning where responsibility is shared by the two parties [i.e., learner and teacher]" (*op. cit.*: 203).

To sum up, there seems to be a trend that learners' acceptance of feedback in L2 production, whether provided by pre-defined algorithms or by a human evaluator, increases with the degree of focus on forms and the contextualisation of the information provided. However, we need to know more about "how students actually engage with feedback and how feedback shapes their writing processes, revising practices and their self-evaluation capacities" (Hyland, 2010: 179). The present study responds to this need by investigating attitudes towards online feedback on writing, focusing on their perception of the usefulness of models. It sets out to broaden our understanding of the complex mechanisms at play by triangulating actually observed learning with students' perceived learning gain, and their reported attitudes towards the feedback received.

More specifically, the following research questions are addressed:

1. How did students' summary writing change after following the online module?
2. How did students themselves perceive their learning gain through the online module?

3. What are the students' attitudes towards the different feedback types provided, and is it possible to pinpoint their mediating role between perceived and actual learning gain?

Study design

The exploratory intervention study from which the data for this research were drawn comprised a six-week period of weekly classes in summary writing from spoken input. The 38 participants were 2nd year bachelor students of Applied Language Studies at a higher education college in Belgium. Their writing proficiency in the target language, German, was at level B2 of the Common European Framework. For 76% of the participants, this was their first experience with an online learning module. Prior to the online learning phase, the students had received face-to-face (F2F) instruction on the same task. In the first and sixth weeks of the study, pre- and post-tests in summary writing and a fill-the-gap test on cohesive ties were administered. The actual intervention consisted of four consecutive weekly sessions of individual online learning in class. Every week, the students wrote a summary of a radio feature on a news item of about 3-5 minutes. They were guided through three different task phases by an online learning path following a "moderate constructivist" design (Karagiorgi & Symeou, 2005) in which diverse exercise and feedback types were combined (see figure 1). The rationale for the distribution of instructivist and constructivist task types in the module is that "pre-determined, constrained, sequential, criterion-referenced instructional design is most suitable for introductory learning while constructivist approaches are more appropriate for advanced knowledge acquisition" (*op. cit.*: 23). This rationale was followed in order to design the online module: It includes different task types, ranging from multiple-choice over fill-the-gap to open answer questions, and provides different types of feedback, ranging from corrective feedback to models. This complex architecture was deemed necessary in order to meet the different learning goals involved in summary writing: (1) Build up listening strategies in order to reduce anxiety and foster scan-listening for main content items; (2) Build up writing strategies to produce a new coherent text based on a different input genre (here: radio features); (3) Provide the necessary linguistic knowledge in the L2 to write a concise and coherent summary, i.e., typical chunks and phrases for the genre and cohesive strategies. According to insights from scholarship on feedback in online learning on the one hand, and in (L2) writing, on the other, feedback types were adjusted to the different goals.

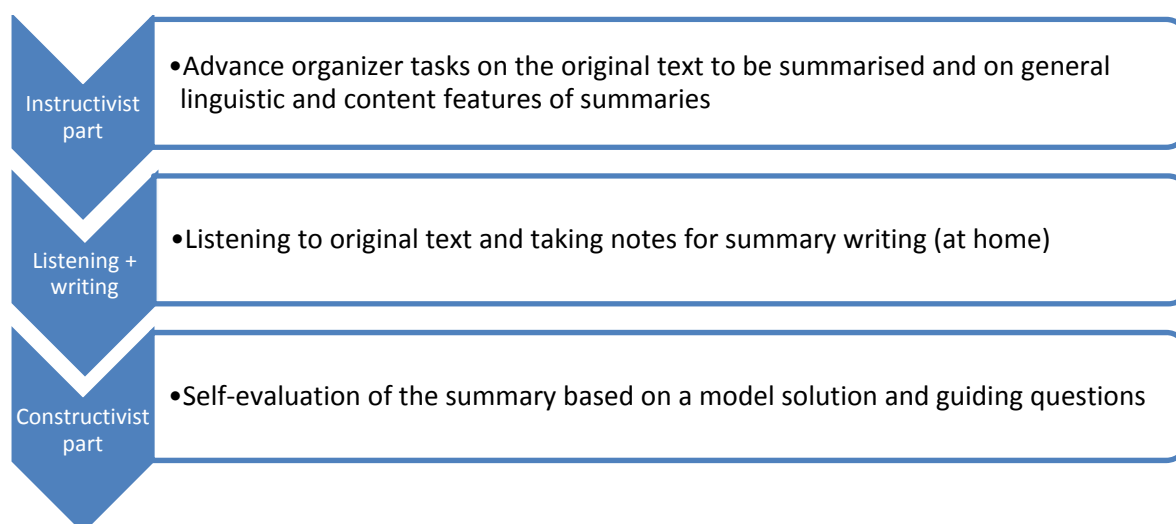


Fig. 1 Flowchart of the design of the individual online learning module

The actual writing phase started in class, and was finished at home. Therefore, the self-evaluation of the previous week’s summary was actually the first step in each online module. Table 1 provides an overview of the different exercise types, the instructional focus, and the feedback type, according to van der Kleij *et al.*’s (2012) classification.

Table 1 Features of the task types in the online module

Task type	Instructional focus	Feedback classification
Guided self-evaluation of the previous week’s summary based on a model solution, incl. questions for self-reflection on strategy improvement	Advanced skill and strategy acquisition through a complex task relying on constructivist principles	<ul style="list-style-type: none"> • Delayed • no Knowledge of Result (KR) • Knowledge of correct response (KCR) + Elaborated feedback (EF) • Focuses on process level (content selection and structure)
Advance organisers on new topic, listening fragments, and language-related exercises in the form of multiple choice and fill-the-gap exercises with automated feedback	Introductory learning through instructional cognitivist design: pre-determined, constrained, sequential	<ul style="list-style-type: none"> • Immediate • KR • KCR (+ EF) • Focuses on task level

In the instructivist part, which contains preparatory exercises for the actual listening task, feedback was delivered immediately after the completion of each exercise in the

form of (1) knowledge of result (correct / wrong), plus (2) knowledge of correct response in case of a wrong answer and/or (3) elaborated feedback, i.e., background information on applicable rules and possible sources of error. In order to avoid that students would neglect the feedback, a problem frequently mentioned in studies on online feedback which can even lead to inconclusive results (*cf. supra*), some elements were built in to make students read the feedback: exercises were conceived as a suite of closely related items, building up on the knowledge provided in the KCR and (eventual) EF of immediately preceding exercises¹⁶.

In the constructivist part that focuses on self-evaluation of the writing process and product, feedback was delayed (i.e., in the class session following the elaboration of the summary) and provided no knowledge of result. Instead, knowledge of correct response was made available by means of a model solution, and students were prompted to compare their own solution with the model to stimulate noticing (Schmidt, 1990). They were guided in this process by reflective questions that directed their attention towards specific features of the model summary, e.g. information provided on the source text, answers given to main questions in the introduction.¹⁷

It is important to note that naturalistic classroom data was analysed. All students followed the same learning path and, as such, received the same support. Therefore, this study has no intention of making claims about differential effects of feedback types. Instead, rich data was collected in order to explore students' overall learning gain, to monitor their attitudes before, during, and after the intervention in an online environment, which was a new experience for the learners, and to assess their self-reliance after having followed the online module.

The data sources for the analysis are represented in figure 2, which depicts their triangulation for interpretation.

¹⁶ and ² see examples in the supplementary materials provided online

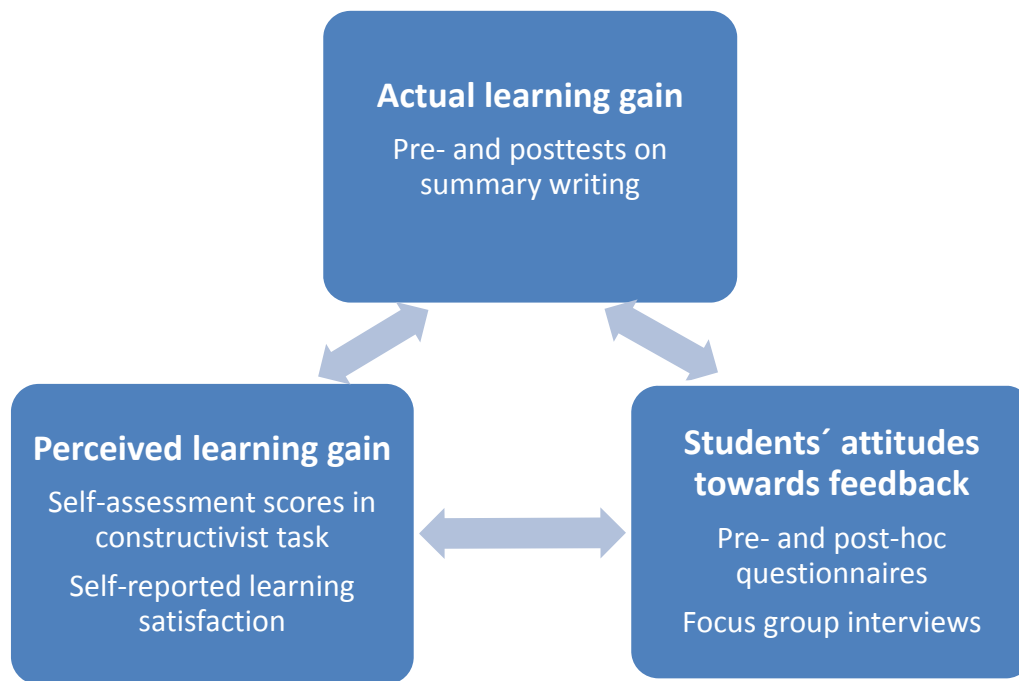


Fig. 2 Data triangulation for the analysis of perceived and actual learning gain

In the next section, data analysis and results will be presented for each of the three subsets.

Data analysis and results

Actual learning gain: Pre-post-test on summary writing

In order to investigate a possible learning gain, students wrote a pre-test in summary writing immediately before the intervention and six weeks later a post-test, immediately after the last online module. The tests were compared with regard to features that address HOCs, that is, concerning the text level (as opposed to word and sentence level) (for an overview, see table 2):

1. Degree of content elaboration through restructuring information. The original radio features to be summarized were interviews. High content elaboration was seen as (a) abandonment of the question-and-answer structure in the summary, and (b) changes in the original proposition order that favoured a coherent summary structure.
2. Degree of linguistic elaboration through rephrasing content (as opposed to a *verbatim* copy of the original text), and through variation in co-reference

strategies. As a unit of analysis for the latter, the references to the interviewee were analysed. A high degree of linguistic elaboration coincided with a broad range of different reference strategies (like pronouns, synonyms, hypernyms, etc.) as opposed to the recurrence of the interviewee's name.

Table 2 Overview of the quality features assessed in the pre- and post-tests on summary writing

Target	Feature	Pre-test	Post-test
Content elaboration	Q&A-structure	retained	abandoned
	Proposition order	retained	changed
Linguistic elaboration	Reproduction	quasi- <i>verbatim</i> (w/o quotes)	in own words
	Co-reference (interviewee)	recurrence (name)	synonyms

For each individual student, a holistic evaluation of the learning gain concerning these quality features was made (progress / status quo / deterioration). For 48% of the students, a clear progress was stated, for 33.5% , the post-test showed a status quo, and 18.5% performed worse in the post-test.¹⁸

Perceived learning gain

In order to measure students' perceptions about their own learning through the online intervention, two units of analysis were defined: (a) the students' self-reported overall learning gain, and (b) the students' self-assessment scores on their summaries based on the comparison with the model solution.

Self-reported overall learning gain

As a unit of analysis for students' self-reported learning, their ratings for overall progress in summary writing after the F2F lessons, that is, before the online intervention, and those after the online learning sequence, are compared (see table 3). In anonymous pre- and post-intervention questionnaires, the students rated the

¹⁸ see example of a commented pre- and post-test exemplifying a case study in the supplementary materials provided online

statement "I feel I have made progress in summary writing" on a five-points Likert scale. On a positive note, the number of students who felt they learnt "a lot" increased considerably after the online class (from 2.7% to 20%), while there is a noticeable decrease in the number of participants who declared they learnt "a little" (from 65% to 34%). While none of the students declared that they had not learnt anything at all, there are students (11%) who were rather pessimistic about their progress ("not much") after the F2F lessons, and this percentage even increased to 14.5% after the online module. The most noticeable result is that considerably more students declared that they felt **insecure** about their progress after the online class ("don't know": from 21.5% to 31.5%).

Table 3 Self-estimated progress after F2F lessons and online module on a 5-point-Likert scale

Rating	After F2F lessons	After online module
"I made (...) progress"		
a lot of	2.7%	< 20%
a little	65%	> 34%
don't know	21.5%	< 31.5%
not much	11%	< 14.5%
none at all	0%	= 0%

Self-assessment with model solution

As part of the task in the online module, the participants were asked three times to evaluate their summaries. They were assisted in the evaluation process by a model solution and reflective questions¹⁹. The self-assessment consisted of two elements: (a) a score on a scale of 10, (b) a rationale for the score.

There was noticeably low variance in the three scores the students attributed to their own summaries, both at individual and at group level. The average score over the whole group and the three assessment sessions was 6.2 /10 (5.9; 6.5; 6.0), with a spread ranging from 3 to 8. An increase of 0.6 in the second self-assessment is followed by a decrease of 0.5 in the third and final assessment. Therefore, there is hardly any difference between

¹⁹ see example in the supplementary materials provided online

the averages of the first and last assessment scores. The individual variance between the three scores did not exceed 2 scales (/10).

For the sake of completeness, it has to be mentioned that the above-stated insecurity with respect to the self-assessment does not hold for all participants. Some of the students' rationales reflect a high degree of awareness of relevant summary quality features, and firm self-efficacy beliefs:

"7[/10], because all important propositions are there, and I tried to reformulate specific words."

"8: My summary has a different structure [than the model], but I actually think it's also a good one."

"8: I actually like my summary, though I failed to mention some figures. I don't consider them to be important."

The overall low variance in self-assessment scores suggests that either the students did not feel comfortable with this task, or they did not experience stable progress in their own summary writing. But how can this be reconciled with the stated actual progress in summary writing? In order to understand the assumptions underlying this apparent dichotomy, it is important to explore the students' attitudes.

Students' attitudes towards the feedback received

Answers in questionnaires

In the course of the intervention, students were asked twice in an anonymous online questionnaire to express their attitudes towards the exercises and the feedback received in the online module, the first time immediately after the second online session (i.e., the first self-assessment with a model solution)(Q1), and the second time after the last online session (Q2). The question asked was: "Which of the exercises in the online learning module did you (a) like / (b) dislike? Please explain". The answers to the first part of the question were processed quantitatively. The answers to the second part of the question were coded according to main characteristics mentioned (e.g., "difficult", "tedious", "fun to do", "helpful"). As a full report of the qualitative analysis is beyond the scope of this article, selected statements are used as illustrative examples, enriching the picture emerging from the quantitative results.

The students were united in their positive attitude towards the preparatory listening exercises in the instructivist part of the online module (only positive mentions, Q1:12; Q2:16). The opinions towards instructivist grammar exercises on cohesive ties were diverging: While they received 13 positive mentions in Q1 and 7 in Q2, the same number of students admitted to dislike them (Q1: 8, Q2: 12). Interesting to note is the general

tendency towards a more unfavourable attitude in the second questionnaire. Two illustrative examples of the students' comments are:

"I didn't like the grammar exercises because there was not enough explanation about *how* the respective structure had to be used."

"It was not really interesting to do the fill-the-gap exercises on grammar because there were always several possible answers."

With respect to the constructivist self-reflection exercises based on the model solution, a completely different picture emerges: They were mentioned positively only once in Q1, and negatively two times in Q1 and five times in Q2. Illustrative example comments:

"I really found it interesting to compare my own summary to a model. Like this, you discover different ways to write about the same topic."

"Because I wrote the text myself and eliminated the errors I could find before uploading it, I was barely able to improve it by myself. If I cannot compare [my assessment] to a score given by a teacher, how should I know whether I assess my text correctly?"

*Post-hoc focus groups*²⁰

Immediately after the last online session, two focus group discussions took place with randomly selected students (F1: 6 students, F2: 7 students). In order to ensure that the participants felt free to express their opinions, they were not led by their teacher, but by an experienced researcher unknown to the students. Two of the key questions (K1, K2) were directly related to the students' attitudes towards exercise and feedback types. The salient tendencies in the discussion about these two key questions are summarized below.

(K1): "Generally speaking, there were two types of exercises: In one type, you had to fill in gaps or select an answer between several choices, and you got direct feedback in form of an automated answer. In the other type you had to fill in a sentence or a text, and then you had to compare your answer to a model. Which type did you prefer personally?"

All of the seven participants of F1 declared to prefer the instructivist exercises. The main reason given was that the answers were clear-cut, while students felt insecure about the limits of acceptability of their own formulations when comparing them with

²⁰ see full transcription of relevant parts of the discussions in the supplementary materials provided online

the model. The attitudes in F2 seemed to be more diverging. The general trend of the discussion in F2 was that both exercise types have their strengths, and that a good mix of both is probably the best for an online module.

(K2): “You had to evaluate your own summaries, comparing them with a model solution. Did you find this easy? Did it help you?”

In F1, all but one student did not find the model solutions helpful. One student was inclined to find them helpful but disliked the fact that she had to assess her own text with a score. In F2, the participants mentioned two concrete items they learnt through the models, provide information about the source text, and address the "wh"-questions in the introduction. However, also in this group, most participants declared to feel insecure about the self-assessment scores because they lacked a framework for reference, and they could not assess the accuracy of their own texts. One student, who was favourable towards the usefulness of the models stated she always tried to implement what she had learnt from the model in her next summary. This statement reinforces that feedback needs to be sustained in order to gain its full potential (Storch, 2010:42).

Discussion

The comparison of pre- and post-tests indicates a positive development in students' skills through the intervention. Nevertheless, students' self-assessment behaviour and self-reported learning gain reveals that they are seemingly unaware of this learning gain.

The first two research questions can thus be answered in the following way: Whereas students actually learnt to elaborate their summaries both linguistically and in terms of content, they did not fully appreciate the value of these skills. Instead, they remained insecure about their progress, because they focused on accuracy which they felt unable to evaluate themselves. Indeed, as the model only represents one out of a vast combination of possible formulations, the chance to discover language-related problems is low. In the following, we discuss this mismatch based on students' expressed attitudes in order to answer research question (3): What is the role of students' attitudes as possible mediator between actual and perceived learning gain?

The discrepancy between the students' expectations and the constructivist approach underlying the assessment part of the online module can be attributed to three different, yet related, key factors: (1) evaluation, (2) linguistic focus and (3) learner motivation (see table 3 for an overview).

Table 3 Three key factors in mismatch between students' expectations and adopted pedagogic approach

Key factor	Students' expectations		Adopted pedagogic approach
Evaluation	External judgement	≠	Internal self-disclosure
Linguistic focus	Lower-order concerns (accuracy)	≠	Higher-order concerns (content and coherence)
Learner motivation	Performance goal	≠	Mastery goal

1) The mismatch in evaluation is linked to the fact that constructivist feedback does not provide an external judgement in the form of KR. This clearly confused the students who were not used to self-evaluation, and therefore, did not seem comfortable relying on it, as is evidenced by the following quotes:

"We didn't get feedback on our summaries, so how can we know whether they were good? You can't make progress like that." (post-hoc questionnaire)

"I was disappointed because we were expected to evaluate our own summaries. I think that a combination of online exercises and a reliable text correction by the teacher would be ideal." (post-hoc questionnaire)

Drawing on Hattie & Timperley (2007), effective feedback should "answer three major questions (...): Where am I going? (What are the goals?), How am I going? (What progress is being made toward the goal?), and Where to next? (What activities need to be undertaken to make better progress?)" (Hattie & Timperley, 2007: 86). Clearly, the second question was not answered to the students' satisfaction.

2) The linguistic focus mismatch is rooted in the students' (short) history as learners of German as a foreign language. Second-year bachelor students are not used to writing longer texts in German which requires a focus on HOCs, like coherence and cohesion. Instead, their understanding of good writing is intrinsically linked to accuracy, which leads them to concentrating on LOCs like grammar and word choice. This might also reflect secondary school teachers' feedback behaviour. The following two statements illustrate this focus on LOCs:

"I don't know if I made a lot of grammatical errors, so I can't assess the quality of my summary." (post-hoc questionnaire)

"Maybe you make the same errors every week again without noticing it. (...) Like this, there's a risk that your errors fossilize." (post-hoc questionnaire)

This confirms Hyland's (2003) and Diab's (2005) findings about students' preferences towards corrective feedback in writing.

- 3) A third mismatch can be detected between the students' perceptions of their learner role on the one hand, and the role they should adapt in a constructivist learning environment, on the other. In this context, Duijnhouwer, Prins and Stokking (2012) refer to personal goals, differentiating between "performance goal" and "mastery goal":

"As students have a stronger mastery goal they have a stronger focus on developing their competence (...). As students have a stronger performance goal they have a stronger focus on getting their competence positively judged" (*op. cit.*:173)

The following two selected quotes clearly show that the students struggled with the role they needed to adopt for self-evaluation:

"I found it strange that we had to evaluate ourselves. After all, we're not teachers, so how can we mark our own work?" (post-hoc questionnaire)

"Because I wrote the text myself and edited it carefully before uploading it, I was unable to pick up on any mistakes myself. If I can't compare my own judgement to that of a teacher, it's difficult to know whether I was right or wrong." (post-hoc questionnaire)

However, it is important to note that not all students expressed negative attitudes towards the constructivist feedback they received in the online module. Some statements indicate an open attitude towards this new feedback experience, positively mentioning the goal-orientedness that is reinforced by the model solution. That is, the Hattie & Timperley's (2007) question "Where to next?" has been answered to those students' satisfaction.

"Thanks to the model solutions, the teacher's expectations were clear. You knew how you actually should do it." (post-hoc questionnaire)

"I also learnt to always mention the source and the 5 WH-questions; I think I always learnt something from the comparison and tried to do better next time" (post-hoc questionnaire)

"When you have to fill in whole sentences, you've got to think about it (...) and when you compare it to the model solution, it's like 'ah, I have to keep this in mind'." (post-hoc focus group interview 1)

Moreover, when comparing the different feedback types they received in the online environment, students also mentioned problematic aspects of **instructivist** feedback, self-critically reflecting on their own behaviour:

1) Retention problem:

"If you need a piece of information that you read about two exercises ago, you already can't recall it anymore." (post-hoc questionnaire)

The respondent refers to the "exercise suites" in the instructivist part of the module, building upon knowledge provided in the feedback of immediately preceding exercises. Clearly, this stimulus to read the feedback failed to yield the desired effect. Van Beuningen (2010) also discusses the limited capacity of corrective feedback to foster transfer: "[F]ocused CF is rather a form of explicit grammar instruction than a focus-on-form intervention (...). This might make it more difficult for learners to transfer what is learned from the feedback to new writing situations" (*op. cit.*:11).

2) "Click-away" behaviour:

"I think you click away after filling in an exercise, I personally do not read the whole feedback but just click on 'OK'." (post-hoc focus group interview 2)

In the same vein, Heift (2010) concludes from her literature review that 18% of learners in CALL environments neglected to look up answers altogether. She specifically mentions the problem of recasts being neglected in a CALL environment: "[T]here is no longer a need for the learner to attend to the feedback generated by the computer - the correct answer has already been supplied." (*op. cit.*:199). This "click-away" behaviour that challenges the pedagogic intentions of the instructor-designer can be subsumed under the umbrella term "Instructional disobedience" coined by Elen (2013).

Concluding from the above variety of statements, the attitude clearly depends on the individual's learning motivation, and will ultimately determine the benefit the students draw from the model solution. It is therefore necessary that students are conducted towards developing a more mastery-oriented mind-set to optimally benefit from constructivist online tasks and "maximize their potential for self-repair" (Heift, 2010). This can be achieved by learning how to use cognitive and meta-cognitive strategies for writing and self-evaluation (Segev-Miller, 2004). Feedback - as important as it might be - is "only part of the equation", as Hattie & Timperley (2007) pointed out:

"A major task for teachers and parents is to make academic goals salient for all students, because students who are prepared to question or reflect on what they know and understand are more likely to seek confirmatory and/or disconfirmatory feedback that allows for the best opportunities for learning." (*op. cit.*:103f)

Conclusion

The main goal of this exploratory study using naturalistic classroom data was to broaden our understanding of students' attitudes towards different feedback types in an online environment while dealing with a complex writing task. For this purpose, the actual learning gain as measured in pre- and post-tests was examined in combination with self-reported learning gain as measured in self-assessment, and triangulated with the students' attitudes as expressed in questionnaires and interviews.

The results of the study revealed an interesting dichotomy between the achieved results and the perceived learning gain. A satisfactory overall learning gain was detected in summary writing through constructivist online support that consisted of repeated and guided self-evaluation based on model solutions. In stark contrast to these findings is the students' low self-confidence showing in the self-assessment scores and the reported insecurity concerning their overall learning gain.

The students' attitudes were found to go some way towards explaining the observed contrast. Indeed, the dichotomy is caused by a mismatch between the students' expectations of a learning environment on the one hand, and the adopted constructivist approach on the other. This mismatch can be attributed to three different key factors: (1) evaluation, (2) linguistic focus and (3) learner motivation.

The findings are in line with previous research that has pinpointed the importance of motivation (Duinhouwer *et al.*, 2012; van der Kleij *et al.*, 2012) for the uptake of feedback. The activation of the learner's ZPD in an individual online learning environment requires a mature learner role, striving towards mastery (instead of performance). It is important for teachers to bear in mind learner histories, and consequently raise their students' awareness of this aspect of (language) learning before the latter engage in constructivist online learning activities: "[T]eachers should help students to develop practices of feedback use which will scaffold and engage them as they develop their own self monitoring capabilities" (Storch, 2010:180).

An important limitation of the adopted study design, a single-group pre-post-test intervention, is that a possible "learning by doing"- effect could not be controlled for. However, there are two aspects that make a strong case for the effect of the online intervention: (a) in the pre-post-tests comparison, the focus was on specific aspects of writing (HOCs) that formed an important part of the intervention and that are highly unlikely to be achieved through "learning by doing"; (b) at the moment of the pre-test, the students had already received face-to-face instruction on summary writing during six weeks, including writing practice, and clearly had not learnt these specific aspects.

Another shortcoming of the present study is that the two important cognitive mediators of learning apart from attitudes, viz. strategy development and self-efficacy beliefs, were not operationalized in a consistent way, and therefore had to be excluded from the analysis. While in recent research, the importance of self-efficacy beliefs for

L1 writing development has been underlined (Pajares, 2003; Woodrow, 2011), this area to date is being under-investigated in the field of L2 writing research (Kormos, 2012). Future research should investigate the role of self-efficacy beliefs in the context of computer-assisted advanced L2 writing more thoroughly.

This study revealed that the use of models as feedback in an online individual module for summary writing can actually enhance students' writing performance regarding HOCs. Additionally, it might be an interesting focus for further research to investigate whether the concept of noticing (Schmidt, 1990, 2001) can be expanded to HOCs, as previous research focusing on the use of models to foster attention on LOCs has provided evidence of noticing (Hanaoka, 2007; Martínez & Roca de Larios, 2010). Such an in-depth analysis of students' response to the models might reveal how processes of noticing can be linked to development in advanced L2 writing.

Acknowledgements

I would like to thank Mat Schulze and the anonymous reviewers for their insightful comments and helpful suggestions on earlier drafts of this paper.

References

- Aljaafreh A. and Lantolf J. P. (1994) Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, **78** (4), 465-483.
- Cotos E. (2011) Potential of automated writing evaluation feedback. *CALICO Journal*, **28** (2).
- Diab R. L. (2005) EFL university students' preferences for error correction and teacher feedback on writing. *TESL Reporter*, **38** (1), 27-51.
- Doyle W. (1977) Paradigms for research on teacher effectiveness, *Review of research in education*, 163-198.
- Duijnhouwer H., Prins F.J. and Stokking K.M. (2012) Feedback providing improvement strategies and reflection on feedback use: Effects on students' writing motivation, process, and performance. *Learning and Instruction*, **22** (3), 171-184.
- Elen J. (2013) "Instructional disobedience": Challenging instructional design research. *Earli 15th biannual conference*, München, Germany: unpublished keynote speech.
- Enginarlar H. (1993) "Student response to teacher feedback in EFL writing", *System*, **21** (2), 193-204.
- Felix U. (2002) The web as a vehicle for constructivist approaches in language teaching. *ReCALL*, **14** (1), 2-15.
- Felix U. (2005) E-learning pedagogy in the third millennium: the need for combining social and cognitive constructivist approaches. *ReCALL*, **17** (1), 85-100.
- Hanaoka O. (2007) Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research*, **11** (4), 459-479.
- Hattie J. and Timperley H. (2007) The Power of Feedback. *Review of Educational Research*, **77** (1), 81-112.
- Heift T. (2005) Corrective feedback and learner uptake in CALL. *ReCALL*, **16** (2), 416-431.

- Heift T. (2010) Prompting in CALL: A Longitudinal Study of Learner Uptake. *The Modern Language Journal*, **94** (2), 198-216.
- Hyland F. (2003) Focusing on form: Student engagement with teacher feedback. *System*, **31** (2), 217-230.
- Hyland F. (2010) Future Directions in Feedback on Second Language Writing: Overview and Research Agenda. *International Journal of English Studies*, **10** (2), 171-182.
- Karagiorgi Y. and Symeou L. (2005) Translating constructivism into instructional design: Potential and limitations. *Educational Technology & Society*, **8** (1), 17-27.
- Kintsch E., Steinhart D., Stahl G., Matthews C. & Lamb R. (2000) Developing Summarization Skills through the Use of LSA-Based Feedback, *Interactive Learning Environments*, **8** (2), 87-109.
- Kirschner P.A. (2002) Can we support CSCL? Educational, social and technological affordances for learning. In Kirschner P.A. (ed.), *Three worlds of CSCL: Can we support CSCL?*, Heerlen: Open University of the Netherlands. 7-47.
- Kormos J. (2012) The role of individual differences in L2 writing. *Journal of Second Language Writing*, **21** (4), 390-403.
- Kulhavy R.W. (1977) Feedback in written instruction, *Review of Educational Research*, 211-232.
- Mandernach B. J. (2005) Relative effectiveness of computer-based and human feedback for enhancing student learning. *The Journal of Educators Online*, **2** (1).
- Martínez N.E. and Roca de Larios J. (2010) The Use of Models as a Form of Written Feedback to Secondary School Pupils of English. *International Journal of English Studies*, **10** (2), 143-170.
- Nagata N. (1993) Intelligent Computer Feedback for Second Language Instruction, *The Modern Language Journal*, **77** (3), 330-339.
- Norris J.M. and Manchón R. (2012) Investigating L2 writing development from multiple perspectives: Issues in theory and research. In: Manchón R. (ed.), *L2 writing development: Multiple perspectives*. Boston/Berlin: Walter de Gruyter, 221-244.
- Pajares F. (2003) Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, **19** (2), 139.
- Pujolà J. -T. (2001). Did CALL feedback feed back? Researching learners' use of feedback. *ReCALL*, **13** (1), 79-98.
- Radecki P.M. & Swales J.M. (1988) ESL student reaction to written comments on their written work, *System*, **16** (3), 355-365.
- Robinson G.L. (1991) Effective feedback strategies in CALL: Learning theory and empirical research. In: Dunkel P. (ed.), *Computer-assisted language learning and testing: research issues and practice*. New York, N.Y.: Newbury House Publishers, 155-167.
- Rosselle M., Serco L. and Vandepitte S. (2009) Learning outcomes and learner perceptions in relation to computer-based feedback. *Indian Journal of Applied Linguistics*, **35** (1), 45-61.
- Ruddell R.B. & Boyle O.F. (1989) A study of cognitive mapping as a means to improve summarization and comprehension of expository text, *Reading Research and Instruction*, **29** (1), 12-22.
- Schmidt R. W. (1990) The Role of Consciousness in Second Language Learning. *Applied linguistics*, **11** (2), 129-158.
- Schmidt R. W. (2001) Attention. In: Robinson P. (ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 3-32.
- Segev-Miller R. (2004) Writing from sources: The effect of explicit instruction on college students' processes and products. *L1-Educational Studies in Language and Literature*, **4** (1), 5-33.
- Shute V. (2008) Focus on Formative Feedback. *Review of Educational Research*, **78** (1), 153-189.

- Storch N. (2010) Critical Feedback on Written Corrective Feedback Research. *International Journal of English Studies*, **10** (2), 29-46.
- van Beuningen C. (2010) Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, **10** (2), 1-27.
- van der Kleij F.M., Eggen T.J.H.M., Timmers C.F. and Veldkamp B.P. (2012) Effects of feedback in a computer-based assessment for learning. *Computers & Education*, **58** (1), 263-272.
- Warschauer M. (1996) Computer-assisted language learning: An introduction. In Fotos S. (ed.), *Multimedia language teaching*, Tokyo: Logos International. 3-20.
- Wolsey T.D. (2008) Efficacy of instructor feedback on written work in an online program. *International Journal on ELearning*, **7** (2), 311-329.
- Woodrow L. (2011) College English writing affect: Self-efficacy and anxiety. *System*, **39** (4), 510-522.

3.3 Further results and conclusions from the study

3.3.1 Evidence of “noticing” in the constructivist part of the module

Although “noticing” can certainly not be equated with learning, it has been described as *sine qua non* for it (Ellis, 1995; Schmidt, 1990, 2001). Furthermore it shows “a learner’s reactive response to the feedback” (Aljaafreh & Lantolf, 1994, p. 267) and is therefore an important indicator for L2 development. In IS1, noticing was stimulated through the presentation of models for the written summaries, accompanied by guiding questions. In AT terms, this kind of self-reflective action seeks to actively involve the *subject collective*, i.e., the students, in the process of noticing secondary contradictions between their own performance (outcome) and the teacher’s goal (object) in the activity.

Students compared their summaries to a model solution three times in three subsequent weeks. The repeated task allowed them to get used to and gain confidence in the procedure. When asked to compare their syntheses with the model solution (cf. Appendix B), students made very good observations evidencing noticing. The three standard questions asked each week were: (1) Are there certain items of the model missing in your summary? If so, which ones are they? (2) Did you mention items in your summary that are not in the model? If so, why did you find them important? (3) Are there any other differences between the model and your own summary? In order to analyse the answers given to these three questions, they were labelled according to (a) main topic (content, structure, or language) and (b) students’ stance towards their own summary (positive, negative or neutral).

The total number of observations over the three weeks amounts to 396, which means that the students reported on average 3.5 differences between the model and their own summary. A quantitative comparison of the three main topics revealed that, in line with the focus of the questions, content differences are noticed most often, totalling between 60-70% of the comments each week. Noticed differences in structure covered about a quarter of the comments. Finally, language-related differences were noticed to a far lesser degree (between 6 and 10%) (see Figure 7). Whereas the ratio between the three main topics remained more or less unchanged throughout the intervention, the total number of observations decreased each week.

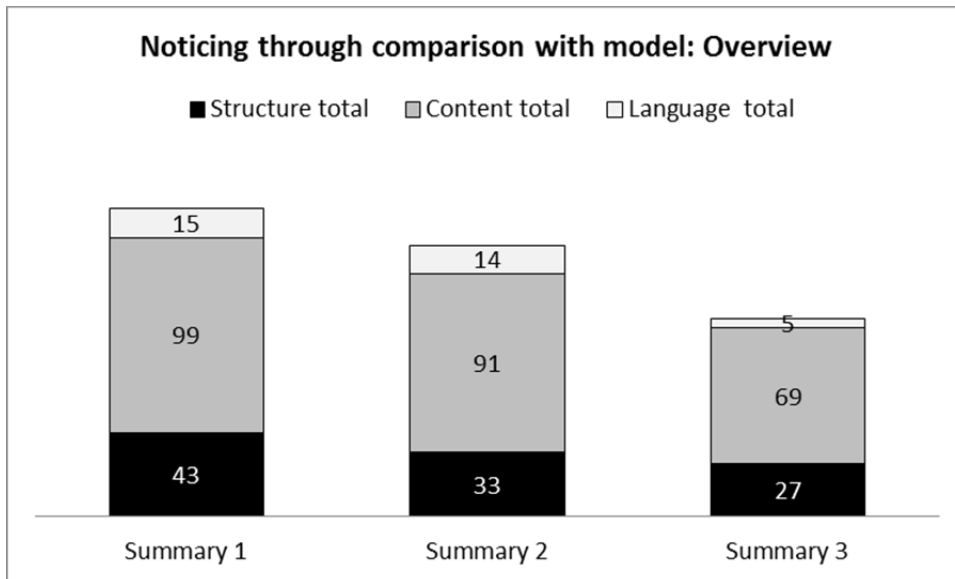


Figure 7 Number of differences registered between student's own summary and model

In the following, the main categories are presented with the respective sub-domains, indicating the frequency of mentioning between brackets and adding illustrating examples, where appropriate.

Content:

- Missing integration of specific details (n=80)

I did not integrate many figures in my summary. I thought they were not that important.²¹

- Added integration of specific details (n=75)

In the model, the fact that most freight traffic will go through the tunnel by 2017 is missing. I think this is quite important for transport in Europe.

I wrote that the project idea dates back to the 1940s, but this is of course not so important. In the model it says that the planning took a long time, which covers about the same information.

I added some more information about Facebook, because some readers, especially the elder generation, might not have the necessary background information.

- General congruence with model with respect to main items (n=48)

²¹ In the following, all students' quotes from their answers in the online module, in the questionnaires and the focus group discussions were translated by the author from the original Dutch or German to English.

- Information about the source text of the summary (n=38)

I forgot to mention [the source / the interview].

- General level of detail (n=10)
- Errors in specific details (e.g., wrong figures and misspelt names) (n=5)

Interestingly, when students reported about details in their summaries that were not in the model, and vice-versa, frequently they added a rationale, sometimes also mentioning reader-orientation. This indicates deep content elaboration which is the very purpose of providing this kind of constructivist feedback.

Structure:

- Overall similarity with model resp. source text (n=49)

I followed the structure of the source text whereas the model has a different structure.

My summary has a totally different structure. It is rather chronological, and does not combine advantages and disadvantages, like in the model.

- Introduction and/or conclusion (n=22)

I should have stated the main point in the first paragraph more clearly.

- Length (n=18)

- Overall structure and outline (n=14)

I think that my outline is not perfect. I described one of the advantages in the paragraph of the disadvantages.

Language:

- Literal copy of source text vs. rephrasing (n=15)

I need to rephrase more instead of copying what was said in the source text.

I copied the first sentence of the introduction in the source because it stands out. Perhaps that is not such a good idea for a summary.

- Grammar and orthographic errors (n=11)

I don't know whether there are still a lot of grammar errors in my summary that I didn't notice.

- Clarity and complexity of formulation (n=8)

I needed more words than the model to express the same ideas.

My summary lacks cohesive links.

In general, the language-related observations were formulated more vaguely. The noticed differences with the model mostly referred to the lack of rephrasing in students' summaries. A relatively large number of observations related to grammar, expressing students' uncertainty whether their texts contained unnoticed grammatical errors.

With regard to students' stance towards their own summaries expressed in the comments, interestingly, content-related observations were mostly reported in a neutral or even a positive way, adding a rationale to the difference in content selection (see Figure 8). Observations regarding differences in structure were mostly reported with a negative stance in the first week, while in the subsequent weeks, a more neutral tone was adopted (see Figure 9). Finally, language-related comments predominantly met with a negative attitude (see Figure 10).

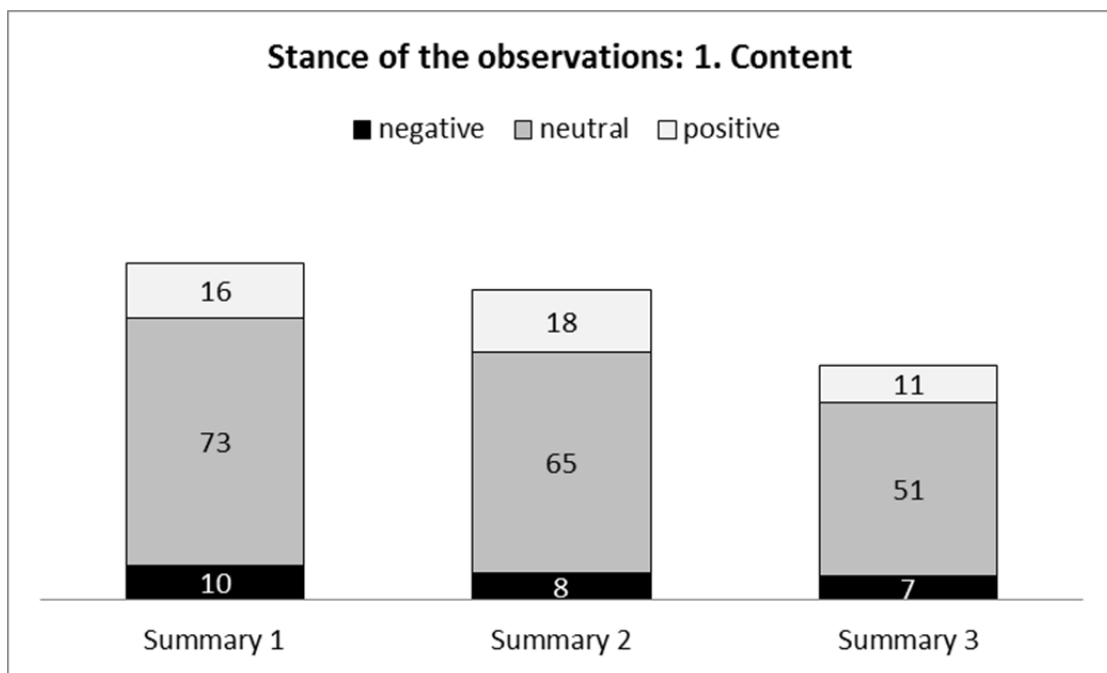


Figure 8 Students' stance towards their own summaries regarding content

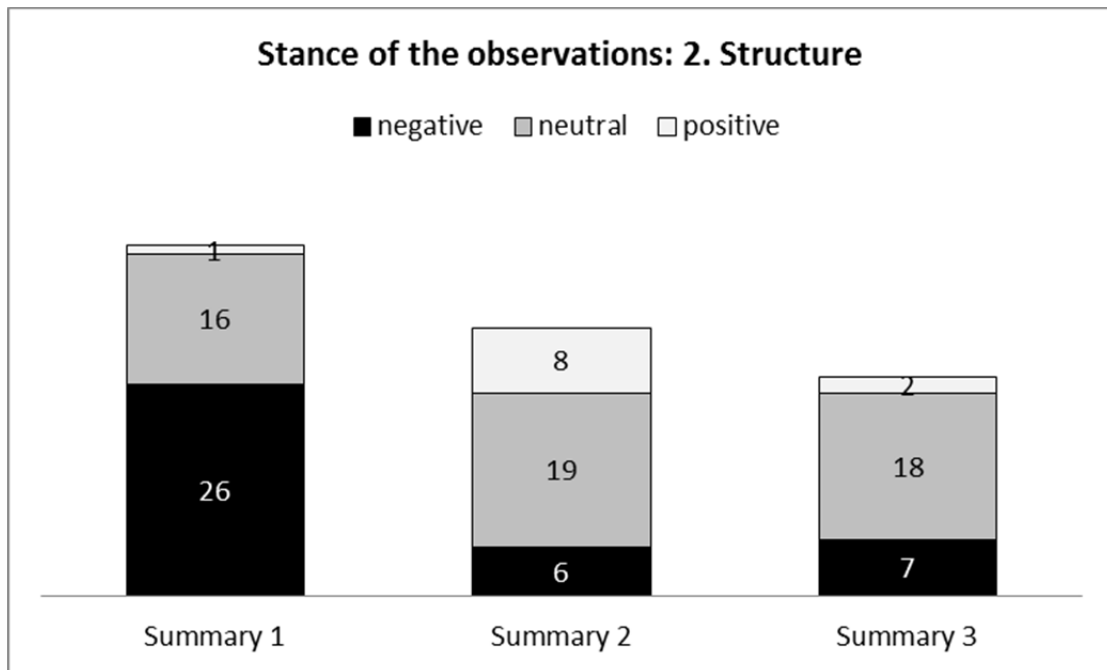


Figure 9 Students' stance towards their own summaries regarding structure

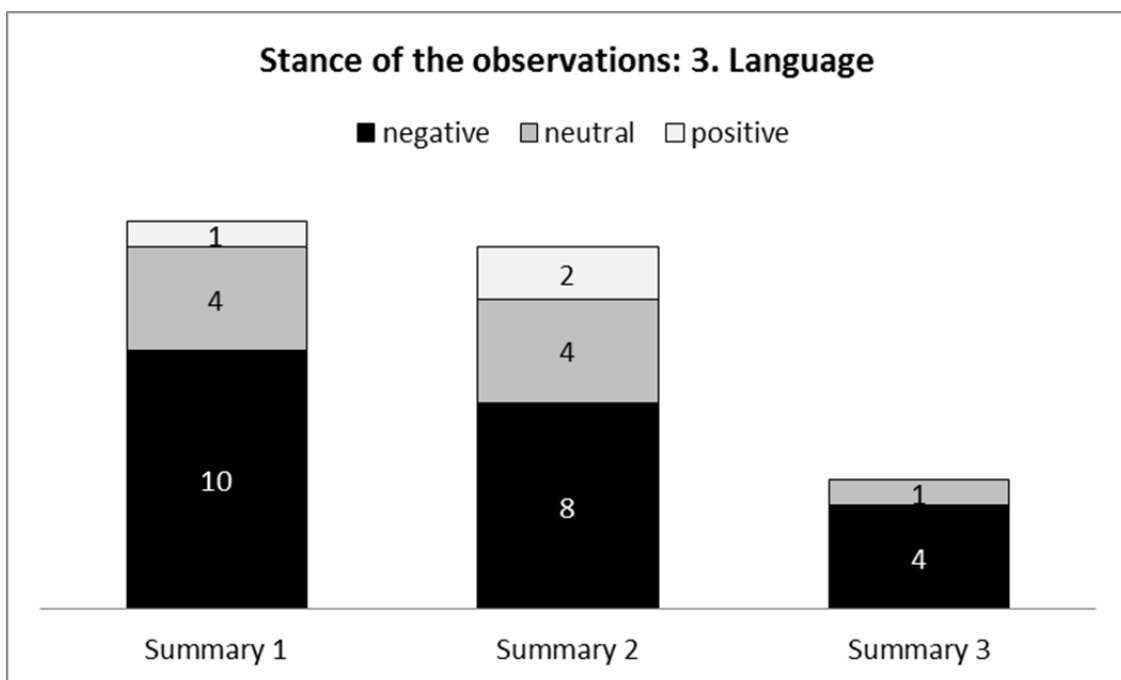


Figure 10 Students' stance towards their own summaries regarding language

To conclude, the answers of the students' given in the online module when asked to compare their solution with a model, in the questionnaires, and in focus groups, revealed that the constructivist approach to cognitively involving the students in the

evaluation process led to the intended effect. Especially with regard to the HOCs of text quality, i.e., content and structure, the model indeed stimulated noticing.

3.3.2 Insight into writing processes: How "planners" and "revisers" elaborate notes for summary writing

Besides analysing the impact of the intervention, this study also had an exploratory aim. In order to better support students in building up writing strategies, I wanted to gain insight into their writing processes and to determine a possible correlation of the latter with the outcome. This was deemed necessary, as the analysis of P2 (see section 2.2.2.2) revealed strong individual differences regarding writing processes, in accordance with literature that claims that these processes differ between skilled and unskilled writers (Sasaki, 2000; Scardamalia et al., 1984; Thorson, 2000). Therefore, the main research question for this exploratory part of the study was: "Are there writing processes that characterise the writing style of students that produce good summaries?". To answer this question, the writing processes of the post-tests were recorded and analysed with the screen-capturing software *Morae*. The results of this exploratory analysis were presented at the Junior Research Meeting of the German Association of Applied Linguistics in Essen in March 2011.

A comparison of the time to produce a first draft and the time spent reviewing the draft revealed big individual differences, with the time taken for reviewing varying between 0% and 200% of the production time of the first draft, and the mean value being 50%. This means that students spent on average double the time composing than reviewing. In order to distinguish "planners" and "revisers" (Torrance et al., 1994), two analysis units were combined, viz. the product/process-ratio (operationalized as the relation between the number of keystrokes of the final version and the total number of keystrokes during the writing process) and the pausing behaviour (operationalized as the relation between the total duration of pauses exceeding two seconds²² and the total time on task). Based on this analysis, four distinct "planners" who presented high values for both measures and four distinct "revisers" who had low values for both measures were identified. However, there appeared to be no direct relationship between writing process and product quality. Among the students with good post-test results were both "planners" and "revisers", as well as mixed types. However, a combination between high product/process ratio and low pausing, i.e., an outspoken linear writing approach that is characterised by little, if any, editing activity, had a negative correlation with

²² The threshold of two seconds is derived from the mean pausing time of adult writers reported in writing scholarship (Wengelin, 2006)

text quality. None of the writers with a result above average showed a purely linear writing path, whereas two of the writers who had a very weak performance were linear writers. From this observation, the conclusion was drawn that recursive writers had indeed, as literature suggests (Scardamalia et al., 1984), a better chance to produce good summaries, and that external motivation to stimulate recursive writing should be included in an online writing module.

Moreover, observing the processes of integrating notes taken during listening in the summary, two main techniques were identified, viz. the "patchwork technique" and the "skeleton technique". As a general trend, the former was used more by the on-line planners. It implies that they started to write from their notes immediately, inserting the propositions in the given order, deciding on-line where they would fit into the emerging text, and thus integrating them directly content- and language-wise, adapting the context immediately to the inserted chunk. To some extent, this writing technique resembles the composition style of unskilled writers that Uzawa (1996) labelled the "what-next"-approach. As a general trend, the latter technique was used more by revisers. *Skeleton writers* first jotted down notes in the given order of the source, after which they re-ordered the bullets to create a coherent skeleton of the summary to be written, and only then started writing, based on this skeleton, occasionally reshuffling items. Thorson (2000) also observed this technique in a qualitative study on L1 and L2 writing strategies for two different genres (letter and article). In IS1, this technique was applied by some highly skilled writers and led to good quality texts. Therefore, it was decided to integrate a "skeleton" stage in further interventions in which students would be required to produce an extended outline from a bulleted list, indicating the type of logic connection between the bullets.

To conclude, the research question posited for this observational part of the study can be answered positively. Recursive writers who established a skeleton for the text to be written in a pre-writing planning phase mostly produced better summaries. Therefore, in the following interventions, these aspects were specifically addressed in the online instructions.

3.3.3 Attitudes towards PC use in classroom

A major concern in this study was to investigate the attitudes of the students towards the innovative teaching method employed. Besides attitudes towards specific exercise types (see section 3.2), also general attitudes towards PC use for educational purposes were included in the four questionnaires that were administered in week 1, 3, 5, and 6 of the intervention (see Appendices 4-7).

In the post-hoc-questionnaire (see Appendix 7), students were asked to compare their learning gain in the F2F-classroom-sessions (i.e., six lessons prior to the

intervention), and in the sessions based on the online module. All of the students (n=12, i.e., 32.5% of the participants) who indicated they had learnt more in the F2F-lessons specified their answer pointing to "individual" or "more" feedback in these lessons. At first sight, this statement seems contradictory, as the individual students effectively got a lot more (automated) feedback during the online module. This suggests that students did not *perceive* the feedback as such, because it was not provided by a human. As the second most important flaw of the online sessions, the "lack of interaction" was mentioned (n=4). Clearly, these students preferred human-human interaction to human-computer interaction in an educational setting. The following statements from the post-hoc-questionnaire (see Appendix 7) illustrate this finding²³:

In the beginning, it was great, but now I think it is a pity we don't talk in the classroom

We spend so many hours on the computer for facebook and so on, so I really don't want to sit in front of the computer at school, too, if it's not really necessary. Think of the radiation.

The computer is for relax time, so it doesn't fit into a classroom.

Data from the pre-hoc-questionnaire that aimed at getting a general idea of participants' PC literacy and daily use of the computer revealed that they spend nearly 2.5 hours on average in front of the PC, 45% of which in free-time activities (i.e., social networking sites), and 55% in study-related activities. Regarding language learning activities, the majority (91.5%) use the PC on a regular basis to consult ancillary programs for text production (grammars and dictionaries) and 66% use the automatic correction function of word processors. In sharp contrast, only 5.7% use the PC regularly for interactive exercises and 26% for communication in the L2. Considering that all of them communicate daily in their L1 via social networks, this seems to be a small percentage. The figures also revealed that these students of Applied Languages clearly perceive the PC as an ancillary instrument for information retrieval, but not as an instrument for learning through interaction.

Another important factor that caused students' distress was a mismatch between the innovative instructional form and the end-of-term exam. The following quote illustrates this concern:

I think it is useful during the classes. But as the exam is going to be different, I also think it is useless.

²³ All statements presented in this section were translated from original German or Dutch (students used both L1 and L2 to answer the questionnaires)

This perceived mismatch in part is also rooted in students' rather traditional, paper-based way of preparing for an exam. The following statements illustrate a certain "pen-and-paper" - paradox of this so-called "Net Generation" (Tapscott, 2008):

It would be more helpful to have these important grammar items on paper. Then we could get back to them while preparing for the exam. Of course the exercises remain available online, but when I am preparing for the exam, my computer is switched off.

I liked the exercises, but I prefer to have the theory on paper instead of on the screen. Personally, I am convinced that one tends to quickly forget the information one reads on screen.

I feel sorry that I did not take a screenshot of all exercises and pasted them into a word document. Now, I have nothing on paper.

3.3.4 Discussion of the study in the light of Activity Theory

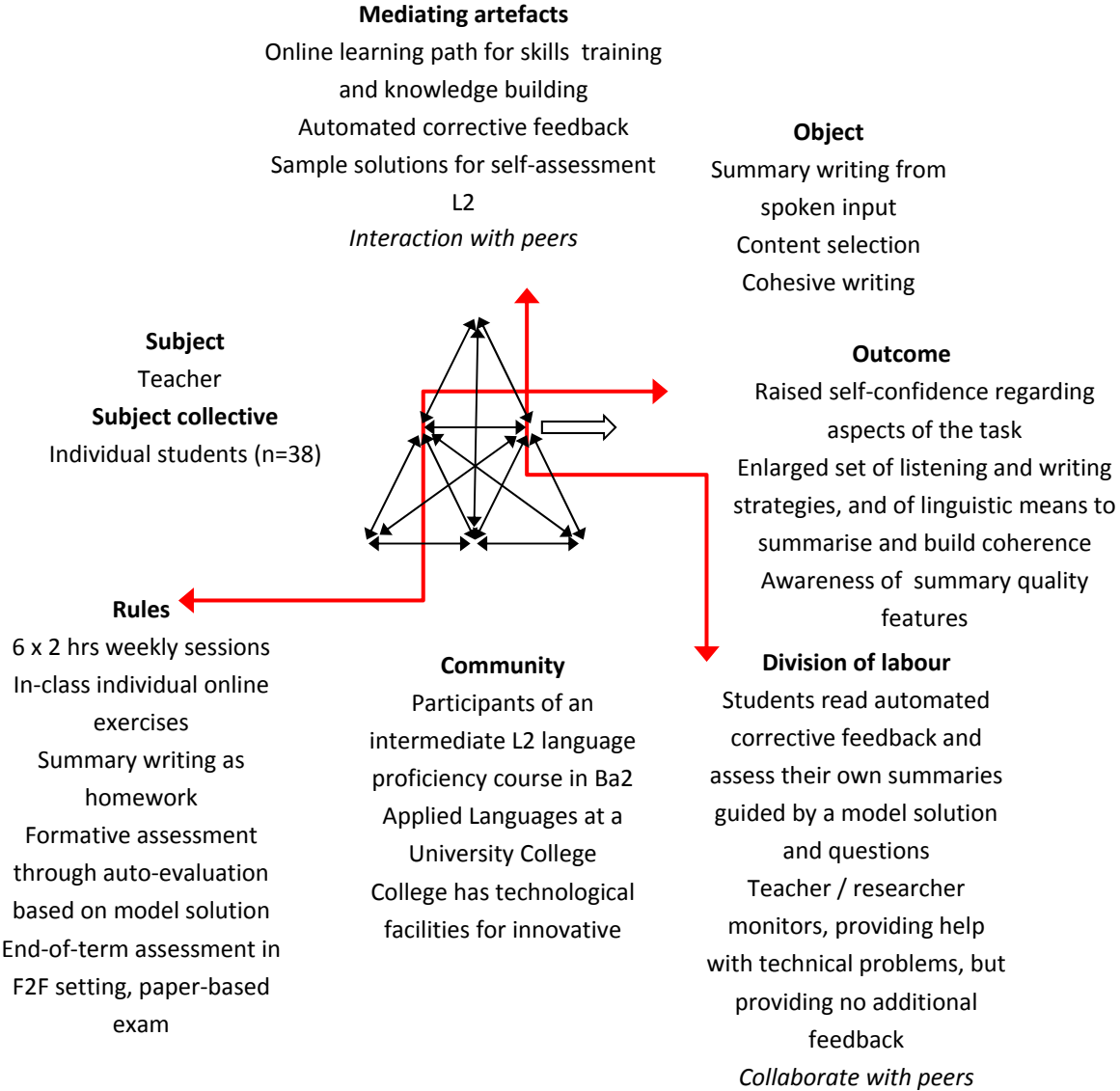


Figure 11 System of the educational activity in IS1

As the graphical representation of the AS reveals (see Figure 11), two major contradictions arose in the activity of IS1.

1. A new conflict between rules and outcome was stated by the students in the post-hoc questionnaire (see section 3.3.3): The innovation was only part of the classroom-based instruction, whereas the exam, due to institutional rules, was held in a traditional pen-and-paper based way. This caused participants distress, as their main objective of the activity was to pass the end-of-term exam, and they did not perceive the activity as useful to reach this objective. To resolve this conflict, a change in institutional rules would be required to allow students to be scaffolded online as part of the exam. Thus, actually, the tension occurs between

the rules of the classroom activity's AS and the rules of the institution's AS, therefore exemplifying a *tertiary* contradiction.

2. Students' reactions on the innovative teaching method in the post-hoc questionnaire also revealed an interesting contradiction between their perception of themselves as language learners in a classroom AS and the role computers play herein: They declared to miss personal feedback and social interaction, which caused a lack of learning satisfaction. In activity-theoretical terms, there is a contradiction between the division of labour and the artefacts used in the activity. The conclusion was drawn that human-computer interaction needs to be counterbalanced with human-human interaction in order to lead to learning satisfaction.

The consequences that were drawn from the analysis of IS1 for the design of the next intervention study (IS2) imply that the intervention should not directly be linked to exam forms nor have repercussions on exam results in order to avoid contradiction 1. Therefore, IS2 was carried out in a general language course, taking the form of a general "learning-to-write unit". To avoid contradiction 2, it was decided to integrate peer interaction using the computer as a tool in the educational design, thus, adding a social constructivist component. In order to explore the added value of this new approach, a comparative study design was set up that combines the new, socio-constructivist, approach with the cognitive constructivist approach that was taken in IS1, using the computer as a tutor in an individual setting.

Chapter 4

The effect of pre-programmed feedback and direct peer feedback on process and product
(Intervention Study 2)

4.1 Background information on the study

Intervention Study 2 targeted the comparison of an individual online learning module following cognitive constructivist principles with a collaborative online learning module following social constructivist principles. It was carried out in a classroom-based setting within a German language proficiency course taught in the second bachelor year of the Linguistics and Literature study programme at UGent (n=48). The intervention took place in April of 2011 and comprised two 110-minutes sessions in two successive weeks in two class groups. In addition, it was decided to combine two different, yet related, tasks in this intervention that are frequently used in L2 academic writing instruction, viz. summary writing from spoken input and synthesis writing from written input. These two tasks were chosen in order to compare the (perceived) usefulness of the support provided by online technologies and pedagogies as well as the tracking facilities that these technologies provide to investigate the writing process.

To tackle the effect of conditions (individual vs. collaborative writing and writing from spoken vs. written input) and at the same time adhere to ethical requirements, a cross-sectional study design was adopted (see Table 5). Consequently, each student wrote one text individually, and one in a randomly assembled collaborative group of three, one summary from spoken input, and one synthesis from written input.

Table 5 Cross-sectional design in Intervention Study 2

	Group 1	Group 2
	Topic "Stuttgart 21"	
Week 1	<i>Collaborative synthesis from written input</i>	<i>Individual synthesis from written input</i>
	Topic "Gotthardtunnel"	
Week 2	<i>Individual summary from spoken input</i>	<i>Collaborative summary from spoken input</i>

The two individual online module were created in *curios*, an authoring system for the creation of online tests and exercises that was developed at UGent and is accessible through the institution's VLE. The online module used in the intervention in week 2, group 1 was a slightly adapted version of the module used in IS1 week 1 (see Appendix 8a). For week 1, group 2, a new individual online module was developed. It concerns the same topic and integrates the same source materials as the collaborative task (see Appendix 13), but is structured differently in order to fit cognitive constructivist principles for learning in that the students are guided in the task process (source text elaboration and planning of student's own synthesis) by the sequence of exercises, and

in the reflection about their own answers by model solutions. The two collaborative modules (week 1, group 1, and week 2, group 2) were presented as learning paths, simulating a collaboration script, in the institution's VLE (see Appendix 13 for the learning path including input material for week 1, group 1). They adhere to social constructivist principles for learning in that the students are stimulated to give each other feedback and scaffold each other during the whole task process. The tools used were a group forum (provided in the institution's VLE) for synchronous content elaboration of the source texts and planning of the joint summary or synthesis in class, MS Word for (asynchronous) individual planning and text fragment production, and *Google Docs* for (a-)synchronous composition and revision of the collaborative texts.

In the classroom sessions preceding and following the intervention, a pre- and a post-test as well as anonymous questionnaires were administered. The pre-post-tests consisted of two parts, viz., a writing-from-input test (see Appendix 11a and 11b for the writing prompts) and a discrete items test on cohesive strategies (see Appendix 12). A pre-hoc-questionnaire concerning learning style (see Appendix 10) was filled in on paper, whereas the two pre- and post-hoc questionnaires were administered online through the University's VLE, using the authoring tool *curios*. They contain questions about individual writing strategy use, PC use for (language) learning, and expectations or attitudes towards the intervention and the online technologies used (see Appendices 9a and 9b).

Part of the study results, more specifically, a comparison of the individual vs. the collaborative condition for the synthesis task, was published in a special issue of the international double blind peer-reviewed *CALICO Journal* on "Web 2.0 and Language Learning: Rhetoric and Reality", edited by Michael Thomas and Mark Peterson (see following section).

4.2 Affordances of Web 2.0 technologies for collaborative advanced writing in a foreign language (Article in *CALICO Journal*)²⁴

Can online collaboration yield a positive effect on academic writing in a foreign language? If so, what exactly is the added value, compared to individual writing, and (how) does it translate to better output? These are the central questions addressed in this paper. Second Language (L2) writing research has long highlighted the benefits of collaboration in terms of both L2 learning and text quality. Most recently, the positive effect of co-ownership and peer feedback on process and product has been emphasized in studies on Computer-Supported Collaborative Writing. What has remained underexplored is the impact of Web 2.0 technologies on *advanced* L2 writing. The present paper bridges this gap through an empirical study combining Web 2.0 technologies with an academic writing task. Collaborative and individual writing processes and products are compared by applying a mixed-methods approach. The results shed new light on claims made in previous studies that collaboration leads to higher text accuracy. No statistically relevant difference was found between the individual and collaborative syntheses in terms of complexity, accuracy, and fluency. However, collaborative texts score significantly higher on appropriate content selection and organization. Analysis of the process data shows that this is due to in-depth discussions during the planning phase.

KEYWORDS

Computer-Supported Collaborative Writing (CSCW), Advanced foreign language writing, Revision process, Text quality, Peer feedback

INTRODUCTION

The task under investigation in this study, namely writing a synthesis from several written sources in a foreign language, is highly complex because it requires a multiple focus on input processing, content selection and organization, along with production. Can online peer collaboration help students to master it? Web 2.0 technologies offer great opportunities for the collaborative co-construction of knowledge. Wikis and online editors facilitate the co-construction of a joint artifact, in this case, a

²⁴ Strobl, C. (2014). Affordances of web 2.0 technologies for collaborative advanced writing in a foreign language. *CALICO Journal*, 31(1), 1-18. doi: 10.11139/cj.31.1.1-18. Double-blind peer-reviewed.

collaboratively written text. Due to increasing levels of sophistication and usability, these tools have received growing interest from writing instructors and researchers alike. Is it possible, though, to find empirical evidence for the supposed benefit of the online collaboration process in the product, in this case, a synthesis from written sources (for a detailed task description see the study design section of this paper)? Evidence-based research on this question has remained scarce to date (Wang & Vásquez, 2012). This paper answers the need for an empirical study of the benefits of Web 2.0 tools for L2 text production.

The research idea for this study is driven by two complementary lines of scholarship: L2 writing research, and Computer Supported Collaborative Writing (CSCW). L2 writing research has long highlighted the benefits of collaboration in terms of both L2 learning and text quality (Arnold, Ducate, & Kost, 2009; De la Colina, & García Mayo, 2007; Elola & Oskoz, 2010; Fernández Dobao, 2012; Kessler, Bikowski, & Boggs, 2012; Kost, 2011; Storch, 2005; Wigglesworth & Storch, 2009). Where text quality is concerned, texts resulting from pair work have been found to achieve a higher level of accuracy than those written by individuals (Storch, 2005; Wigglesworth & Storch, 2009). Fernández Dobao (2012) even found a significant difference in accuracy dependent on the number of participants involved in the collaborative process (pairs vs. groups of three). She attributed this difference to the higher amount of correctly resolved language-related episodes in the groups of three.

Besides product enhancement, a possible connection between the pooling of linguistic resources in collaborative dialogues and L2 learning has been explored (for a comprehensive review cf. Swain, Brooks, & Tocalli-Beller, 2002). Studies in CSCW have emphasized the positive effect of co-ownership and electronic peer feedback regarding reader-orientation (Blin & Appel, 2011; Kessler, 2009; Lee, 2010). Departing from an Activity Theoretical framework, the emerging text (a "digital material-semiotic artifact") is interpreted simultaneously as an object and as a tool, and therefore, research interest has shifted from the product toward the writing process: "[D]igital material-semiotic artifacts (...) provide data not only on language use and students' interactions, but also on the evolving social and communicative structure of the activity, or the hidden curriculum, as it is negotiated and possibly re-constructed by the participants as the activity unfolds." (Blin & Appel, 2011, p. 477) Specific attention has been given to the role of feedback during the writing and reviewing process, particularly formative feedback, which can either be given by the teacher or by peers. In the former case, the teacher "shares the responsibility to assess with the students, but still supervises the revision process" (Alvarez, Espasa, & Guasch, 2012, p. 388), whereas in the latter, this responsibility is shared among peers. In addition, the process can be supported by the teacher through a script that provides instructions on how to interact, promoting a fruitfully structured interaction (Rummel, Spada, & Hauser, 2009).

Peer collaboration in the reviewing process has been found to focus mainly on surface-level, meaning preserving changes (Paulus, 1999), also called LOCs (lower-order concerns). What has remained underexplored, however, is the impact of Web 2.0 technologies on *advanced* L2 writing (Thorne & Reinhardt, 2008) where HOCs (higher-order concerns) like content selection and organization play an important role. The present paper aims to bridge this gap through an empirical study. In this study, the use of Web 2.0 technologies throughout the writing process is combined with a task, synthesis writing, which belongs to the field of advanced foreign language proficiency and imposes a high cognitive load on the learners.

The research questions for the present study are:

(1) What impact does collaboration have on the final text? Do the collaborative syntheses differ from individually written ones with regard to complexity, accuracy and fluency (CAF) and/or content and coherence? Do collaboratively produced texts (mainly) achieve higher accuracy levels?

(2) How does the fact that writing takes place online and is a joint, collaborative activity, influence the writing process, and especially the reviewing phase? Is it true that peer-induced revision mainly focuses on surface-level, meaning-preserving changes, that is, LOCs?

STUDY DESIGN AND RESEARCH METHODOLOGY

In order to explore the difference between collaborative and individual online writing, an empirical study was set up using a university virtual learning environment and Google Docs as tools. The research was carried out in an intact class (n=48) of university students of German L2. All of them were Dutch native speakers and had an advanced proficiency level of German writing (CEF-level B1-B2²⁵). Prior to the tasks under investigation, all students received an introduction to synthesis writing and to peer reviewing. The instructor modeled the revision of several authentic students' syntheses from a different class in a guided classroom activity.

For the purposes of the actual study, the class was divided into two groups. All participants were required to write two syntheses in crossed conditions, one individually, and one in a randomly selected collaborative group of three, one from aural, and one from written input (see Table 1 for an overview). As the two synthesis

²⁵ CEF: Common European Framework; The CEF levels B1 and B2 correspond with the ACTFL Proficiency Guidelines levels "Advanced low" and "Advanced mid".

tasks differed in terms of input type and outcome, only one of them, synthesis from written input, will be reported on in this paper.

Table 1
Overall Study Design: Two Tasks under Crossed Conditions

	Group 1	Group 2
Week 1	Introduction to synthesis writing and peer reviewing Pre-hoc questionnaires on learning style and attitudes	
Week 2	Individual writing	Collaborative writing
Synthesis based on written sources*		X
Week 3	Collaborative writing	
Summary based on aural source		
Week 4*	Post-hoc questionnaire on attitudes	

* Results reported in this paper

In this task, students were asked to integrate and synthesize information from multiple sources on a complex topic into a coherent and informative text of a prescribed length. This means that they needed to thoroughly select the content of the input material for their own text. The source material for this task, however, was not of an academic nature. It consisted of three different pieces of textual information related to the planned construction of a new station in Stuttgart (*Stuttgart21*) which had been the subject of a heated public debate in Germany: (1) A list of bullet points providing background on the timelines of the project, its ecological and economic impact, and the acts of civil protest it provoked; (2) the declaration of the Association for the German Language (GfdS) about the word of the year 2010, *Wutbürger*²⁶, and its relatedness to civil protest gaining momentum in Germany; and finally (3) a selection of quotes, representing a variety of viewpoints on the project.

The planning phase was organized based on constructivist learning principles, according to which knowledge is constructed in social or cognitive interaction. While in the collaborative groups, social interaction was secured among the peers through a script and an online forum, the individual writers needed a different, cognitive stimulus

²⁶ This neologism translates as "enraged citizens" and refers to massive civil protests against several expensive public infrastructural projects

to interact with, and challenge, their ideas. For this purpose, the individual writers followed an online exercise module with sample solutions: After submitting an answer, they were immediately prompted to compare it with the respective sample solutions. In this way, they were given the opportunity to profit from the sample solution in the following step of the planning phase.

The collaborative groups were supported throughout the planning and writing process by a script helping them organize their collaboration. The effectiveness of scripts for collaboration in computer-mediated settings has been demonstrated by CSCL research. As Rummel et al. (2009) state: "By enforcing specific kinds of activities among the collaborators, scripts are expected to prompt cognitive and social processes by participants that might otherwise not occur" (p. 73). According to Dillenbourg and Hong's (2008) classification, the script used in this study can be considered a micro-script in that it prescribes the collaborators' actions down to the level of the dialogue.

The exploration of the source material was conceptualized as a jigsaw exercise in the collaborative groups: Each member was asked to explore one source and explain it to the others in an online forum. The individual writers were guided in this phase by reflective questions in the online exercise module. Next, both individual and collaborative writers received online guidance to plan their own texts, leading them to build a "text skeleton", that is, an outline interspersed with cohesive ties. After submitting their own skeleton proposal, the individual writers received two sample skeletons. The collaborative groups did not receive sample solutions.

The planning phase of approximately 45 minutes took place in class but was completely online. The writing and reviewing phases were carried out by all participants outside the classroom in Google Docs, an online document-sharing and versioning tool, with a one-week deadline for the delivery of the final text. Again, the collaborative groups were supported by a micro-script guiding the interaction along a timeline. The individual writers received general guidelines for the writing and review stages.

To measure the intervention outcome in terms of product and process, a mixed-methods approach was adopted. For product analysis, the classical CAF triad was measured in a semi-automated way. Complexity was computed using an algorithm by Schulze, Wood, & Pokorny (unpublished manuscript) that combines four textual surface measures into one numerical value for so-called "balanced complexity", each of them covering : lexical variation (Giraud's type-token ratio), lexical sophistication (mean word length), syntactic variation (unique bigram ratio), and syntactic sophistication (mean period unit length). Accuracy was calculated as errors per word ratio (E/W), and fluency as total number of words. In addition, the final texts were evaluated regarding readability by assigning a joint holistic score on a scale of 1 to 5 for coherence and cohesion. To measure content selection, all so-called "missing propositions," that is, main propositions from the source texts that had not been included in the final text,

were counted. Independent samples t-tests on comparison of means were conducted with a view to detecting statistically significant differences between the texts produced individually and collaboratively. To control for the small data set size, non-parametric tests (Mann-Whitney U test) were also run.

Insight into the writing process was gained through both qualitative and quantitative instruments. For the qualitative analysis, the Google Docs revision history of all seven collaborative texts and seven individual texts was traced. Furthermore, the online communication in the collaborative groups was observed analyzing the forum posts (used for planning) and the comments made in Google Docs (used for writing and revision). In addition to this qualitative analysis, the collaboration activity level in the groups during writing and revision was measured by counting all comments in Google Docs collectively (both language and non-language related contributions), as well as the discussions consisting of more than two turns. For a taxonomy of the collaboration activities, the comments were assigned to twelve topic-based categories. This taxonomy draws upon previously established rubrics in order to describe revision behavior (Arnold et al., 2009; Elola & Oskoz, 2010), adapting them to the data and focus of the study. According to the topics that were found to play a predominant role in the peer contributions, some LOC-items were bundled into one category (e.g., punctuation and spelling), and the categories "cohesive ties choice" and "semantics" were added. The revision type (addition, change or deletion) was neglected because the present study focuses on revision content (HOC vs. LOC).

As groups were randomly assembled, a possible influence of the group constellation in terms of proficiency level on the collaboration activity level and/or the final outcome of the intervention was investigated using a non-parametric correlation test (Spearman's Rho). The mean of the individual group members' proficiency scores represents the group proficiency score. The individual scores were calculated as a combination of several previous study results related to language proficiency. To measure the final outcome, a holistic score was assigned to all final drafts, both the collaborative and the individual ones.

Furthermore, the answers of the students in the evaluative post-hoc survey were analyzed both quantitatively and qualitatively and provided additional insight into their experiences with, and attitudes toward, the two writing conditions.

RESULTS

Product Analysis

Quantitative Results Regarding Text Quality

In this section, the results of the quantitative analyses regarding CAF, on the one hand, and content selection and presentation, on the other hand, will be presented. An overview of all numeric results can be found in Table 2.

Table 2
Quantitative Results of Text Quality Measurements for Individual and Collaborative Texts

	Individual texts (n=21)				Collaborative texts (n=7)			
	Mean	Min	Max	SD	Mean	Min	Max	SD
cb (complexity)	39.23	34.93	43.3	2.55	39.37	32.87	43.52	3.23
E/W (accuracy)	.08	.03	.13	.03	.06	.02	.1	.03
# W (fluency)	327	260	463	43	354*	346	371	10
Missing propositions	2.64	1	6	2.47	1.43**	0	4	1.27
Cohesion + coherence	3	1	4	.95	3.14	1	5	1.35
Holistic score (/20)	12.3	8	18	2.47	12.43	9	17	3.05

*Statistically significant difference at the .05-level between individual and collaborative texts ($t=-2.652$, $df=24.76$, $p=.014$)

** Statistically significant difference at the .05-level between individual and collaborative texts ($t=2.080$; $df=26$, $p=.045$)

The descriptive statistics suggest that the collaborative texts (n=7) achieve higher levels of accuracy, fluency and content selection than the individual texts (n=21). On average, collaboratively written texts contain fewer errors, more words, and there are fewer main propositions missing. The E/W (errors per word) ratio of the individual texts averaged .08, with a large spread ranging from .03 to .13. In the collaborative texts, the spread is smaller (ranging from .02 to .1), and the mean is lower (.06). Regarding text length, the difference in spread is even higher (203 versus 26 words).

For all text quality categories, a t-test on comparison of means was performed in order to detect a possible statistical significance which was set at p-value <.05. In order to account for the small overall sample size and the difference in size between the two independent samples, a non-parametric Mann-Whitney U test was also run which confirmed the results of the t-test. The difference in fluency and content selection is statistically significant, but this is not the case with accuracy ($t=1.201$, $df=26$, $p=.241$). Moreover, individual and collaborative texts do not differ in cohesion and coherence ($t=.311$, $df=26$, $p=.759$), nor in balanced complexity ($t=.113$, $df=26$, $p=.911$).

Process Analysis

Qualitative Analysis of the Planning Phase: Content Selection

The results presented in this section summarize the qualitative analysis of two data sources: a comparison between the text skeletons and the final text drafts, and the planning conversations in the forums.

In order to detect the influence of social and cognitive stimuli on the planning process, the text skeletons were compared to the final text drafts in terms of proposition selection and order. The following example illustrates the difference between the two writing conditions (individual and collaborative) in terms of content selection. One of the source materials to be integrated into the synthesis was the declaration of the GfdS concerning the selection of *Wutbürger* as word of the year. This macroproposition is not easy to integrate into the rest of the source material, and was suggested as an introductory or conclusive item in the sample solutions to the individual writers. Nevertheless, 70% of them chose not to integrate the proposition into their final texts and simply dropped it. This is in stark contrast to the collaborative groups, all of whom integrated the proposition into their texts. The following excerpt of a forum discussion reflects the kind of difficulties they encountered while trying to fit the proposition into their text skeletons.

Excerpt 1: Forum discussion (English translation followed by the German original):

S1: I would not mention Stuttgart 21 as example for „Wutbürger“ but just as a stand-alone proposition. After all, the text does not depart from the concept of „Wutbürger“, but from S21. The „Wutbürger“-concept and its implications could then be used as an introduction.

→ S2: You are right. In my skeleton proposal it looks like „word of the year“ was the main topic instead of „Stuttgart 21“.

S1: **Ich würde** Stuttgart 21 nicht als Beispiel für 'Wutbürger' nennen, sondern einfach als Feststellung darstellen. Der Text geht nämlich nicht vom Begriff 'Wutbürger' aus, sondern von S21. Der Begriff 'Wutbürger' und dessen Erklärung kann man dann als Einführung benutzen.

→ S2: **Damit haben sie Recht.** Wie ich es in diesem Textskelett vorschlage, sieht es ja so aus, als ob der Text um das Wort des Jahres gehen wird, statt um Stuttgart 21.

An interesting aside is that most groups held their discussions in the foreign language, German, although this was never explicitly suggested in the micro-script. On the one hand, this might be beneficial for their L2 communication skills training, as they employ hedges and other typical discourse markers (see words in bold in the above excerpt). On

the other hand, it might also prevent them from deepening the discussion due to language-related limitations.

Qualitative Analysis of the Writing Process

As the writing and reviewing phases are very much intertwined in a recursive writing process, they will be treated simultaneously in this section. This approach seems particularly suited for the description of the asynchronous collaborative writing process where the two phases are hardly distinguishable: while one group member still engages in writing a first draft of a text passage, another member may be reviewing already written text of the same document simultaneously. This means that often, emerging text and already written and/or revised text are synchronously present in the same document.

At this point, a caveat is in order. Although the students were instructed to write all texts in Google Docs, some of them clearly preferred to use their usual text editor, and subsequently cut and pasted their individual contributions, or even their complete final drafts, into the Google Docs editor. This is true for about 80% of the individual texts, where only a final edit, or no action at all, can be tracked via the Google Docs history function. This being the case, no quantitative conclusions regarding revision cycles and activities can be drawn. Instead, a qualitative analysis of the text geneses that could be fully tracked (all collaborative, and seven individual writing and reviewing processes) was performed. This analysis revealed two trends: (1) The asynchronous collaborative writing process in all groups is characterized by a constant intertwining of writing and revising (deleting, rewriting, reshuffling) activities. (2) The recursive writing style that was observed in most collaborative groups stands in stark contrast with most of the observed individual writers. The latter clearly depict a rather linear writing process, starting from the text skeleton, copying and reformulating the propositions in the given order, occasionally changing a cohesive tie or the syntax, but rarely or never changing the proposition order and/or choice.

The screenshots in the appendix illustrate two representative text geneses. Figures 1 (“Writing start in a collaborative text”) and 2 (“Final edit in a collaborative text”) represent screenshots from the beginning and final phase of the writing process in one group. The different colors represent the recent changes made by different contributors working simultaneously in the document (a Google Docs built-in function). Figure 1 shows the simultaneous, but diverging, writing start of two group members. In the upper part of the screenshot, one student used the bullet list he had copied into the document in order to reformulate the facts in his own words, bit by bit deleting the bullets after having integrated the respective content in his text fragment. Meanwhile, another student started writing her fragment from scratch (in the lower part of the screenshot). Figure 2 shows the final edit carried out by one member who had been assigned by the group for this task (as recommended in the micro-script). Prior to this

final edit, several revision cycles had taken place in which all group members were involved. The screenshot in Figure 3 illustrates the writing style that was identified as typical of the individual setting. The copied text skeleton serves as a model to formulate new text and is gradually deleted. The screenshot was taken in the final phase of writing the first (and final) draft, after which only minor editing activities took place.

Quantitative and Qualitative Analysis of the Review activities

In order to gain insight into the collaborative review activities, the peer comments provided in the Google Docs were analyzed both quantitatively and qualitatively. The intensity of the activities varied greatly among the groups. The episodes count ranges from 3 to 34 comments. Most comments led to a direct editing action without further discussion. Few comments were neglected. The number of comments that elicited discussions of more than two turns was also different from group to group, from 1 in the least interactive group to 15 in a group that engaged in lively interaction. The group that showed by far the highest level of review activity consisted of three highly proficient students. This group also obtained the best final score. Although there was a noticeable tendency of a higher review activity level in groups consisting of at least two highly proficient students, no significant correlation between those two variables could be detected. Table 3 shows the results of a Spearman’s Rho correlation test between final score, group proficiency level, and interaction intensity.

Table 3 Spearman’s Rho Correlations Between Final Score, Group Proficiency Level, and Interaction Intensity

	Group proficiency	Final text quality
Review activity	r=.382 p=.454	r=.482 p=.274
Final text quality	r=.574 p=.234	

The contribution topics covered in the peer interaction were assigned to twelve categories. Following Kessler’s et al. (2012) subdivision of contributions in collaborative learning situations, nine of the twelve established topic categories refer to language-related contributions (LRCs), while the other three refer to non-language-related contributions (NLRCs): layout questions, workflow and appraisal. Table 4 shows an overview of the contribution topics in order of frequency.

Table 4 Topics Covered in the Review Contributions in order of Frequency

% of total (n=119)	Mean / group	Topic	Type	Remark
22.4*	3.38	<i>Morphology, incl. morphosyntax</i>	LRC	Discussed in all but 1 group
17.6	2.63	<i>Style</i>	LRC	Best performance group: 29% of total
14.3	2.13	Content	LRC	Best performance group: 32% of total
12.6	1.88	<i>Lexical choice</i>	LRC	Discussed in all groups
9.2	1.38	<i>Orthography and punctuation</i>	LRC	Discussed in 4 (/8) groups
8.4	1.25	<i>Cohesive ties choice</i>	LRC	Discussed in 4 (/8) groups
5.9	0.88	<i>Syntax</i>	LRC	Discussed in 4 (/8) groups
3.4	0.50	Coherence	LRC	Discussed in 3 (/8) groups
2.5	0.38	Workflow	NLRC	Meant to be discussed in forum
2.5	0.38	Semantics	LRC	Discussed in 2 (/8) groups
2.4	0.38	Layout etc.	NLRC	Discussed in 3 (/8) groups
1.7	0.25	Appraisal	NLRC	Only in best performance group

* Italicized items relate to meaning-preserving changes (LOCs)

There is a noticeable predominance of morphology-related discussions, covering 22.4% of the total number of contributions. Lexical choice is the only topic that was covered in all groups. Following Faigley and Witte's (1981) taxonomy of revisions, 67% of all contributions (italicized in Table 4) can be accounted for by meaning-preserving surface changes, that is, LOCs. However, 61% of the peer feedback in the group with the best outcome concerns the HOCs style and content. This group is also the only one where praise statements were included in the peer feedback.

The following excerpt from a group discussion through Google Docs comments illustrates a high level of reflection on language. In this discussion about a lexical choice, different arguments are presented (semantics, lexical diversification) and external authorities (two renowned bilingual dictionaries) are cited. The example also aptly illustrates the problem-solving process of the group, which is trying to establish a compromise by referring to different arguments and authorities.

Excerpt 2: Group discussion in Google Docs comments section (English translation followed by the German original, italicized)

S1: Could we find another word for this, I think "Lösung" would be more appropriate, although I recognize [S3]'s attempt to avoid the recurrence of "Lösung" that was already used in the previous sentence...

S2: I used "Klärung" instead of "Lösung" in the conclusion in order to avoid recurrence.

S3: Is "Klärung" a proper synonym for "(Auf)lösung"???

S1: Indeed, I also wonder...

S1: In Duden [main unilingual German dictionary], I find "Lösungsmöglichkeit", sounds somewhat abstract, but is certainly a synonym...

S2: aah, ok

S2: in 'Van Dale' [main bilingual German-Dutch dictionary]-> 'een probleem tot een oplossing brengen' : 'ein Problem einer Lösung/Klärung zuführen'

S3: hmm I think that 'Möglichkeit' is better, but if it's 2:1 => change

S2: I also like "Möglichkeit", it might be the better option.

S1: Können wir kein anderes Wort nutzen, Lösung wäre hier besser, aber ich verstehe was [S3] tun will, im vorigen Satz steht auch schon Lösung...

S2: Ich habe in meinem Schluss Klärung benutzt, anstatt Lösung, um die Wiederholung zu vermeiden

S3: Aber ist Klärung ein reiner Synonym für (Auf)lösung???

S1: ja, frage ich mich auch?

S1: In Duden finde ich "Lösungsmöglichkeit", etwas abstrakt, aber sicherlich synonym...

S2: aah, ok

S2: in 'Van Dale' -> 'een probleem tot een oplossing brengen' : 'ein Problem einer Lösung/Klärung zuführen'

S3: hmm ich meine, dass 'Möglichkeit' besser ist, aber wenn 2-1 => ändern

S2: Ich finde Möglichkeit auch gut und vielleicht besser

Attitudes Analysis

The students' attitudes and preferences expressed in the post-hoc survey helped shed light on their perception of strengths and weaknesses of the collaborative writing and review process. Therefore, this data source will be included in the discussion of the process analysis results with a view to triangulating the qualitative observational data concerning the following aspects: (1) their experience of specific tools and aspects of the collaboration, including hindering factors, and (2) their preference toward individual or collaborative writing in this specific setting and task.

(1) 29% of the students declared their attitude toward online collaborative writing had changed for the better during the project thanks to the high usability of Google Docs. The following quote from the survey illustrates this statement.

Excerpt 3: Quote from post-hoc survey (English translation followed by the Dutch original, italicized):

Normally, I am against group work because this implies a lot of practical organization. However, this time I did not feel hindered because the group members were able to post their contributions in their own time and at their own pace.

Normaliter ben ik tegen groepswerken, omdat het veel praktische planning vraagt. Op deze manier vond ik het echter helemaal niet storend, de groepsleden kunnen op hun eigen tempo wanneer het past hun bijdrage leveren.

Collaboration in itself was perceived as fruitful, especially during the review process, as the mean rating of the following statements on a five-points Likert scale suggests (English translations from the German original): "The collaboration went well in my group" (3.27/5), "I learned from *receiving* peer feedback" (3.25/5), "I learned from *providing* peer feedback" (3.09/5), "Our feedback helped improve the final text quality" (3.71/5), and finally, "I am happy with our final text" (3.56 /5).

Nevertheless, some problems were reported. Besides free-riding (which even led to the breakdown of collaboration in one group), style and pace matching and the higher workload were mentioned as drawbacks of collaborative writing.

(2) When juxtaposing the two writing conditions in this specific task, students displayed two apparently diverging reactions: On the one hand, they positively pointed out the personal engagement of peer discussions in comparison to the anonymity of pre-programmed constructivist feedback, as the following quote shows.

Excerpt 4: Quote from post-hoc survey (English translation followed by the German original, italicized):

Peer feedback feels more personal than a sample solution.

Man hat die Idee, Kommilitonen geben mehr persönliche Feedback als eine Musterlösung.

Moreover, regarding feedback handling in the individual planning condition, 48% of the students stated that they only "superficially browsed" the extra information provided in the feedback.

On the other hand, 90 % of the respondents stated that they preferred the individual (guided) planning activities to the collaborative planning in the online forum. The main problem reports related to diverse individual work pace and technical issues.

Consequently, students suggested carrying out the collaborative planning face-to-face rather than online. We will discuss this issue in the next section of the paper.

On the whole, 70% of the respondents declared preferring the individual writing condition to the collaborative one. The following statement sums up the perceived pros and cons of both writing conditions.

Excerpt 5: Quote from post-hoc survey (English translation followed by the Dutch original, italicized):

Individual tasks are easier because you can just do it your way, even though you don't make a lot of headway. You are limited by your own capacities and style. Group tasks are more demanding in terms of organization and are more time-consuming. Nevertheless you learn more, especially from the direct feedback you exchange with your peers. Actually, I think a combination of both is best. In the end you have to be able to do it on your own.

Individuele oefeningen zijn gemakkelijker, in die zin dat je gewoon je zin kunt doen. Op die manier leer je echter niet veel bij. Je blijft binnen je eigen mogelijkheden en stijl. Groepsoefeningen vragen wat meer organisatie en je werkt er ook langer aan. Daartegenover staat dat je er veel meer door bijleert, niet in het minst door de directe feedback die je elkaar kan geven. Eigenlijk vind ik de combinatie van beide best. Je moet het uiteindelijk ook kunnen zonder de anderen.

DISCUSSION

Research question 1: Differences between the collaborative and individual syntheses with regard to text quality

As the results of the quantitative analysis regarding text quality reveal, the individual texts differ from the collaborative ones in (1) fluency, and (2) content selection, but not in (3) accuracy and complexity nor in (4) cohesion and coherence. In the following, possible factors explaining these results will be discussed in the above order, also taking into consideration results of the process analysis, as and where appropriate.

Fluency

The significant difference in fluency is not considered important as the text production was not timed, and the number of words was limited. The greater length in the collaborative texts can be attributed to a conglomeration effect of the individual contributions.

Content Selection

Far more interesting is the finding that the collaborative texts score significantly higher on content selection than the individual texts. Groups dropped far fewer of the propositions of the source texts that were considered important for the synthesis by the teacher than individuals did. Analysis of the process data revealed that this is due to in-depth discussions during the planning phase. This means that the selected communication tool (online forum) and scaffolding method (script) for this task actually served their purpose. However, the students themselves perceived the forum as inconvenient for synchronous planning activities, even suggesting that collaborative planning should take place face-to-face rather than online. The main problem reports related to diverse individual work pace and technical issues. Moreover, some group members disregarded the step-by-step instructions in the micro-script, which made for a rather chaotic planning phase in some groups. The danger of "over-scripting" as described by Dillenbourg (2002) might come into play here. To sum up, although collaborative online planning was perceived as inefficient by the students, it turned out to be ultimately effective.

On the contrary, the sample solutions did not cause the individual writers to change their own previous selection of propositions. This kind of canned constructivist feedback apparently did not lead to the intended effect of reflection. Perhaps students did not even *examine* the sample solutions in any great detail, as their responses in the post-hoc survey suggest. Why did they, then, declare to prefer individual planning? Time efficiency seems to be the key to understanding the students' attitudes here. Moreover, writing habits also might play a role, as this was the first experience with online collaborative writing for most of the students.

Accuracy and Complexity

The results do not confirm claims made in previous studies that collaboration leads to higher accuracy compared with texts produced by individuals (Storch, 2005; Wigglesworth & Storch, 2009; Fernández Dobao, 2012). On a cautionary note, it should be stated that the small sample size in the present study makes it difficult to detect significant differences. Furthermore, in the above-mentioned studies, the ratios EFC/C (error-free clauses / clauses) and/or EFT/T (error-free T-units / T-unit) are used instead of, or in addition to, the E/W measure to determine the accuracy rate. However, as accuracy is not the main concern of this study, one measure was deemed sufficient to establish a trend. Another plausible explanation for this discrepant finding is the relatively high L2 proficiency among the students in the present study compared to the aforementioned studies. The proficiency level might also explain the fact that hardly any difference was found regarding balanced complexity.

Cohesion and Coherence

It is somewhat surprising that the individual texts do not achieve higher levels of cohesion and coherence than the collaborative texts. After all, the groups had to agglutinate three individually written pieces, which could have resulted in a less coherent joint text. Apparently, most groups were able to overcome this potential problem by setting out clear agreements in the planning phase and by a thorough edit in the revision phase (both suggested in the micro-script).

Research question 2: Influence of the writing condition (individual or collaborative) on the writing process

The qualitative introspection data reveal that the writing process is highly influenced by the writing condition: While individual writers depict a rather linear writing pattern, in collaborative text production, a constant intertwining of the writing and reviewing phases was observed. This confirms the finding of previous scholarship (Fung, 2010) that collaborative writing displays features of a recursive writing process. In this way, collaborative writing experience might help overcome the observed linear writing habit which is unfavorable in academic writing.

This recursive writing pattern fostered an intense reviewing activity with fruitful discussions on language and content-related issues in some groups. Two thirds of all peer comments during the writing and reviewing process are related to meaning-preserving surface changes, that is, LOCs. The observed predominance of morphology-related discussions can be attributed to the L2 in question, German, which is a strongly inflecting language. The fact that surface changes were the major preoccupation of most groups is in line with previous research (Paulus, 1999). However, there is a tendency among the better performing groups (the groups with the highest scores on their final text) to have more contributions about style and content than the other groups. Some group discussions exhibited a keen awareness of HOCs, especially when two or more high achieving students were involved. Excerpt 2 exemplifies the high level of reflection on language in some collaborative groups which has been described as "pooling of knowledge about language" (Storch, 2005) and is believed to be beneficial for language learning.

The final survey results concerning attitudes confirm that the students were aware of the benefits of collaboration, especially when it came to reviewing their text. They were convinced that their final texts had benefited from the peer review activities. Although an overall superior quality of the collaborative texts in comparison with the individual texts could not be verified (see discussion of research question 1), it might still be true that some collaborative texts improved through changes in response to peer comments. In order to answer this question, a detailed analysis of the revision activities induced by peer comments would be needed, which exceeds the scope of this article.

Interestingly enough, notwithstanding the positive attitudes expressed toward the collaborative writing experience as a whole, more than two thirds of the students declared preferring the individual writing condition to the collaborative one. While this might just be a matter of habit, as already mentioned, they also reported problems compromising a good collaboration, like differences in style and work pace. In one group (of seven), collaboration even seemed to have failed due to the free-riding attitude of two of the three members. Their collaboration broke down after the planning phase, and their Google Doc only contained one text skeleton. The only active group member finally decided to upload an individual text instead of a collaborative one.

CONCLUSION

The study presented in this paper was a short-term intervention under quasi-experimental conditions. It served to establish claims about possible benefits and pitfalls of collaborative L2 synthesis writing using Web 2.0 tools, compared to an individual setting. The gathered data are rich in terms of triangulation and specific in their focus on advanced L2 writing. The answers to the research questions were obtained through a mixed-method analysis.

With regard to the impact of the writing condition on the *final text*, two important findings should be highlighted. First, collaboratively written texts show a significant tendency toward better content selection and organization. This is caused by a higher processing depth in the synchronous collaborative planning than in the individual planning guided by sample solutions. Although perceived as tedious and time-consuming by the participants, the forum discussions led to a significantly better content selection in the collaborative texts. This confirms that collaborative online planning, although perceived as *inefficient* by the students, turned out to be ultimately *effective*.

Second, no statistically significant difference between the individually and collaboratively produced texts could be detected in terms of accuracy. These results are not in line with previous research findings on collaborative writing that attribute mainly higher accuracy to collaboratively produced texts. This might partly be attributed to the fact that the participants already had a medium to high proficiency in the L2. A replication of this small-scale pilot study with more participants of a comparably high L2 command would be desirable in order to enlarge the statistical power of the tests.

With regard to the impact of the writing condition on the *writing process*, two main characteristics of online collaborative advanced writing have been detected. There is a tendency that online collaboration naturally stimulates recursive writing. This seems promising. Collaborative writing experience might help overcome the rather linear writing pattern observed in individual student writers which is unfavorable in academic

writing. Further, instances of "pooling of knowledge about language" (Storch, 2005) that are believed to be beneficial for language learning could be observed. In addition, the pooling of cognitive resources apparent in several group discussions also proved to have a positive effect on the outcome. This is an important benefit of collaboration in a task like synthesis writing that imposes a high cognitive load on the learner. Although LOCs were indeed the major preoccupation of most groups, a keen awareness of HOCs could be observed in the discussions of the better performing groups.

Drawing upon socio-constructivist theories, the processing depth of newly acquired knowledge is known to be greater when discussion (with peers) is involved (Swain et al., 2002). Therefore, it is to be expected that the recursive writing patterns and the pooling of language knowledge present in collaborative writing have an enduring effect on individual language and/or (academic) writing skill acquisition. Future research should explore this possible effect in long-term studies in order to provide empirical evidence for the benefits of collaborative advanced writing in a foreign language, using Web 2.0 tools as a platform.

REFERENCES

- Alvarez, I., Espasa, A., & Guasch, T. (2012). The value of feedback in improving collaborative writing assignments in an online learning environment. *Studies in Higher Education* 37(4), 387-400.
- Arnold, N., Ducate, L., & Kost, C. (2009). Collaborative writing in wikis: Insights from culture projects in German classes. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 115-144). San Marcos, TX: CALICO.
- Blin, F., & Appel, C. (2011). Computer Supported Collaborative Writing in Practice: An Activity Theoretical Study. *CALICO Journal* 28(2), 473-497.
- De la Colina, A. A., & García Mayo, M. d. P. (2007). Attention to Form Across Collaborative Tasks by Low-Proficiency Learners in an EFL Setting. In M. d. P. García Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 91-116). Clevedon NY Ontario: Multilingual Matters Ltd.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P.A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL* (pp. 61-92). Heerlen: Open Universiteit Nederland.
- Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, 3(1), 5-23.
- Elola, I., & Oskoz, A. (2010). Collaborative Writing: Fostering Foreign Language and Writing Conventions Development. *Language Learning and Technology* 14(3), 51-71.
- Faigley, L., & Witte, S. P. (1981). Analyzing revision. *College Composition and Communication* 32, 400-414.
- Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing* 21, 40-58.

- Fung, Y. M. (2010). Collaborative Writing Features. *RELC Journal* 41(1), 18-30.
- Kessler, G. (2009). Student-initiated attention to form in wiki-based collaborative writing. *Language Learning & Technology*, 13(1), 79-95.
- Kessler, G., Bikowski, D., & Boggs, J. (2012). Collaborative Writing among Second Language Learners in Academic Web-Based Projects. *Language Learning & Technology* 16(1), 91-109.
- Kost, C. (2011). Investigating Writing Strategies and Revision Behavior in Collaborative Wiki Projects. *CALICO Journal* 28(3), 606-620.
- Lee, L. (2010). Exploring Wiki-Mediated Collaborative Writing: A Case Study in an Elementary Spanish Course. *CALICO Journal* 27(2), 260-276.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing* 8(3), 265-289.
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning* 4(1), 69-92.
- Schulze, M., Wood, P., & Pokorny, B. (unpublished manuscript). *Measuring balanced complexity*.
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing* 14, 153-173.
- Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 171.
- Thorne, S., & Reinhardt, J. (2008). Bridging Activities, New Literacies, and Advanced Language Proficiency. *CALICO Journal* 25(3), 558-572.
- Wang, S., & Vásquez, C. (2012). Web 2.0 and Second Language Learning: What Does the Research Tell Us? *CALICO Journal* 29(3), 412-430.
- Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing* 26(3), 445-466.

APPENDIX

Figure 1
Writing Start in a Collaborative Text (Screenshot)

"Stuttgart 21" ist ein Projekt, in dem man das heutige Kopfbahnhof abrisen wird und ein neuen Durchgangsbahnhof bauen wird: Stuttgart 21 genannt. Das Projekt wurde seit Jahrzehnten geplant, endgültiger Beschluss zur Durchführung ist Anfang 2010.

Aber Während des ganzen Jahres gab es immer mehr Demonstrationen. Zum Schluss waren schon zehntausenden Bürgern gegen das Projekt.

Die Hauptgründe für die Proteste waren **nicht nur die Milliardenhöhe Baukosten, aber für das Projekt war auch die Abholzung des beliebten angrenzenden Schlossgarten mit alten Bäumen nötig. Der Bürger hatte aber kein Mitspracherecht.**

- ~~Planung des Projektes seit Jahrzehnten, endgültiger Beschluss zur Durchführung Anfang 2010~~
- ~~Während des ganzen Jahres 2010 Demonstrationen von immer mehr (zum Schluss zehntausenden) Bürgern gegen das Projekt~~
- ~~Hauptgründe für Proteste: Baukosten in Milliardenhöhe; Abholzung des beliebten angrenzenden Schlossgarten mit alten Bäumen nötig; kein Mitspracherecht der Bürger~~
- Parkschützer: Umweltschutzorganisationen (u. a. BUND, Robin Wood) und Bürger, die den Schlossgarten erhalten wollen. Aktionen seit 12/2009 (u. a. Besetzung des Parks, um so das geplante Fällen der Bäume zu verhindern)
- Eskalation des Konflikts am 1.10.2010: Polizei setzt Wasserwerfer und Tränengas gegen Demonstranten ein, um Abholzung des Parks mit Gewalt durchzusetzen. > Viele Verletzte
- Schlichtungsgespräche (NL *bemiddeling*) zwischen Projektgegnern und -befürwortern (live TV-Übertragung): Okt.-Nov.2010; Leitung: Heiner Geißler (CDU; 80); Ergebnis: „Stuttgart 21 Plus“: Pläne für Stuttgart 21 müssen nachgebessert werden.

(verschiedene Quellen, u. a. www.dw-world.de, www.wikipedia.de)

Haupttext:

Die Meinungen in Bezug auf dieses Projekt sind offensichtlich sehr verschieden **und ist mit Demonstrationen verbunden.**

Zitat: "Das war so ein gutes Gefühl, der Zusammenhalt und die Freude der Menschen, da etwas tun zu können."

Die Leute kommen zusammen aus verschiedenen Motiven. Zunächst gibt es viele Leute, die sich gegen das Projekt richten. Aber auch Leute, die sich gegen die Stadtentwicklung richten, die Art, wie die Regierung mit der Stadt und Bevölkerung umgeht. Viele protestieren eben gegen das Vorgehen der Polizei.

Figure 2
Final Edit in a Collaborative Text (Screenshot)

Stuttgart 21

Stuttgart 21 ist ein Verkehrsprojekt, wobei der heutige Kopfbahnhof Stuttgart Hauptbahnhof abgerissen wird und ein neuer Durchgangsbahnhof zwischen Stuttgart und Ulm hergestellt wird.

Das Projekt wurde seit Jahrzehnten geplant, aber endgültiger Beschluss zur Durchführung gab es erst Anfang 2010. Ziel des Projekts ist die Kreation eines neuen Herzens Europas. ~~Die Bevölkerung von Stuttgart sehen diese Pläne jedoch als Gledach wirft die Bevölkerung von Stuttgart den Initiatoren einen Großenwahnsinn anver~~ und organisierten während des ganzen Jahres 2010 Demonstrationen. Schließlich sind schon zehntausende Bürger gegen das Projekt.

Stuttgart 21 ist ~~alsofeiglich~~ mit wöchentlichen Demonstrationen verbunden, die aus verschiedenen Motiven geführt werden. Zunächst richtet sich eine wesentliche Gruppe gegen das Projekt. Ihre Beweggründe sind nicht nur die hohen Baukosten, ~~die das Projekt mitbringt,~~ sondern auch die Abholung des beliebten Schlossgarten ~~mit alten Bäumen~~. Die Parkschützer vereinigten sich deshalb in Umweltschutzorganisationen wie BUND und Robin Wood. Eine Parkschützerin und Mitglied der BUND beschreibt das allgemeine Gefühl während ~~der~~ Demonstrationen: "Das war so ein gutes Gefühl, der Zusammenhalt und die Freude der Menschen, da etwas tun zu können." ~~Sie hatte ja Glück, dass die Polizisten immer friedlich waren, wenn sie dabei war. Dennoch bleibt ein sehr wichtiger PEin anderer immer wiederkehrender Protestgrund ist~~ das gewalttätige Vorgehen der Polizei ~~und führten die Demonstrationen unvermeidlich zu einer Eskalation des Konflikts am-~~

~~Eine Eskalation des Konflikts wäre sozusagen unvermeidbar. Am 1.10.2010. Dieses gewaltsame Z führten die Manifestationen tatsächlich zu einem gewaltsamen Zusammenstoßen der Demonstranten mit der Polizei. Die Räumung des Schlossparks hatte mehr als hunderte Verletzte zur Folge, indem die Polizei bei der Räumung des Schlossparks Wasserwerfer und Tränengas gegen die Aufständigen einsetzte. Schlichtungsgespräche, unter der Leitung von Heiner Geißler fanden folglich statt. Geißler fordert die Fortsetzung des Projektes, aber erachtet entscheidende Verbesserungen an dem Projekt für notwendig. Stuttgart 21 Plus, die ~~selch~~ eine solche Nachbesserung erzielen ~~willsell~~, wird jedoch nicht ohne weiteres akzeptiert, denn im Facebook, sowie auf der Straße, wird der Protest unerschütterlich fortgesetzt: "Wir wollten kein S21, wir haben es verhindert. Wir wollen kein S21+, wir werden es verhindern." Das Projekt hat klar einen~~

Figure 3
Individual Writing, Copying from Text Skeleton (Screenshot)

(dank) Schlichtungsgespräche > zwischen Projektgegnern und befürwortern (live TV Übertragung): Okt. Nov. 2010; Leitung: Heiner Geißler (CDU; 80)
> Ergebnis: „Stuttgart 21 Plus“: Pläne für Stuttgart 21 müssen nachgebessert werden. (jedoch) "Wir wollen kein S21+, wir werden es verhindern." (Zitat FB-Nutzerin)

Die Bürger und Bürgerinnen die die andauernde Proteste gegen das Stuttgart 21 Projekt im Jahre 2010 anführten, haben endlich einen Durchbruch erreicht. Der Grund der Unzufriedenheit waren vor allem die geplanten Kosten in Milliarden Höhe, als auch die Abholzung des geliebten Schlossparks. Mittlerweile wurde das Projekt umbenannt und heißt es "Stuttgart 21 +", das eine Nachbesserung des alten Projekts ist.

Das Projekt "Stuttgart 21" besteht darin, dass man den heutigen Bahnhof abreißt und einen neuen Durchgangsbahnhof baut. Die Planung zum Bau hat Jahrzehnten gedauert und der endgültige Beschluss zur Durchführung wurde Anfang 2010 vollzogen. Aber nicht alle waren mit dem Projekt zufrieden. Umweltorganisationen und Bürger organisierten sich als so genannte "Parkschützer" die den Schlossgarten behalten wollen. Aktionen gab es sogar schon bevor den Abruchanfang, seit Ende 2009. Die Schützer besetzen den Park um das Fällen der Bäume zu verhindern. Professor Häußermann sagt das es wirklich erstaunlich ist, dass dieses Aktivitätsniveau sich über eine so lange Zeit sich hält und dass es offensichtlich sogar immer mehr Leute anzieht. Es ist also offensichtlich, dass die Bürger und Bürgerinnen nicht bereit sind ihre Ideale aufzugeben trotz u.a. das harte Auftreten der Polizei am 1. Oktober 2010. Dabei worden Wasserwerfer und Tränengas gegen Demonstranten eingesetzt, so dass es trotzdem zu einer Abholzung kommen könnte. Es worden leider viele verletzt.

Dank den Schlichtungsgesprächen die es - geleitet von Heiner Geißler (CDU; 80) - zwischen Projektgegner- und befürwortern im Oktober und November des vorherigen Jahres gab und die auch live im Fernseher übertragen worden, sind die Pläne also nachgebessert worden. Seitdem gibt es "Stuttgart 21 Plus", das eine Vereinbarung zwischen den beiden Parteien ist. Jedoch gibt es immer noch Unzufriedenheit. So schreibt eine Facebook-Nutzerin: "Wir wollen kein S21+, wir werden es verhindern." Es steht aber fest, dass die Gegner jetzt nur noch wenige sind.

4.3 Further results and conclusions from the study

4.3.1 (Perceived) usefulness of the technologies to support tasks

To answer the question about the usefulness of the selected technologies to support students in acquiring the necessary skills and strategies to carry out the tasks under investigation, viz. summary and synthesis writing, two parameters were investigated: On the one hand, the learning gain from the intervention was measured by comparing the written products in pre- and post-tests. On the other hand, the attitudes of the participants regarding the perceived usefulness of the tools and pedagogies used were investigated by means of their answers in the post-hoc questionnaire.

4.3.1.1 Learning gain from the intervention

To investigate a possible learning gain from this two-week intervention concerning academic writing tasks, a pre-post comparison of the writing test was carried out. On a cautionary note it should be stated that the pre-post-test task in this intervention is not identical with the tasks under investigation. This is due to the fact that the intervention combined two different tasks which are highly time-consuming. As the time allotted for the intervention in this language course did not allow to administer extended pre-post-tests, a task was created for this aim that requires a relatively short time investment and at the same time represents a sub-step common to both intervention tasks, i.e., writing from a bulleted input source.

First, the global score (holistic measure) of pre- and post-test-writing test was compared for all participants who wrote both tests in class ($n=35$). The comparison revealed a small and marginally significant increase of the global score between pre- and post-test. ($t=-2.0; df=35; p=.053; D=-.29$) (see Figure 12 Pre-post-test differences in individual writing test, IS2 (all students, $n=35$))

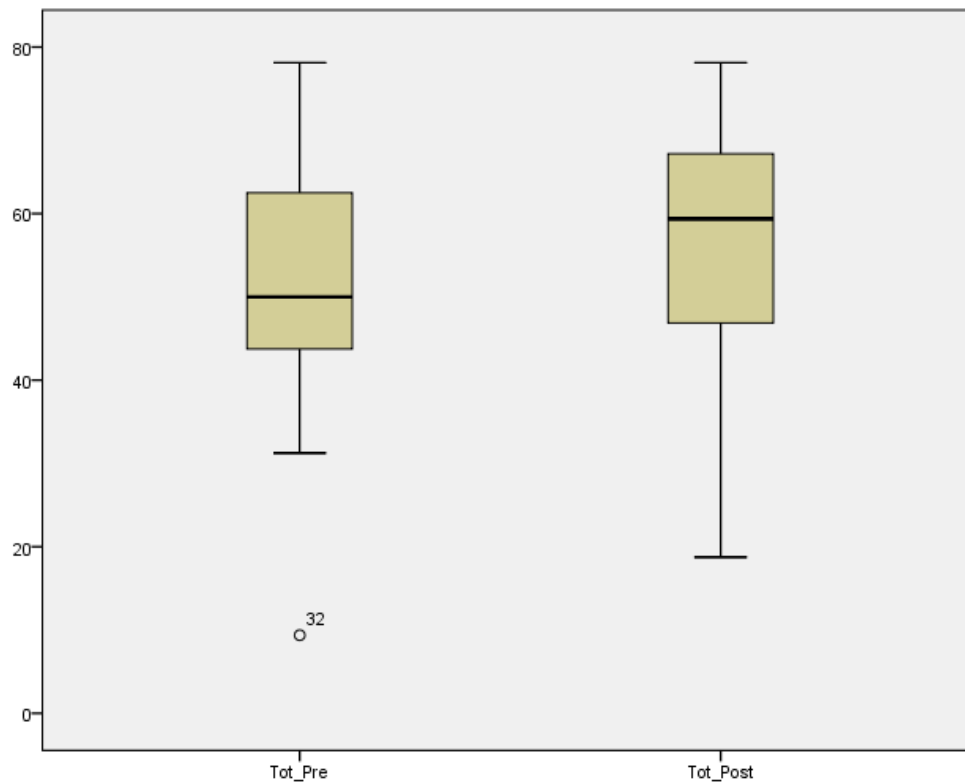


Figure 12 Pre-post-test differences in individual writing test, IS2 (all students, n=35)

For a better insight into the differences of text quality, the pre-and post-tests of those 12 participants who completed both pre- and post-tests within the 20 minutes foreseen in the test design were selected. Due to the low number of cases, only descriptive statistics were used, comparing the mean results.

Table 6 presents the results of the CAF analysis (means for 12 students). Traditional CAF measures were used, one each for accuracy and fluency, and six for complexity, including the four variables that compose the value for "balanced complexity" as proposed by Schulze et al. (Schulze, Verspoor, Wood, & Pokorny), viz. UBR (unique bigram ratio), MPL (mean period length), MWL (mean word length), and GTTR (Giraud's type-token ratio). The most noticeable result is an increase in fluency, as expressed in the number of words written within the 20-minutes time limit. This result has to be qualified by stating that the input prompt for the post-test also contained more words (+22%) than the one for the pre-test. The considerable decrease in accuracy in the post-test that was computed in the ratio of error-free clauses to the total number of clauses, is surprising. A more comprehensive research of the type of errors committed might help to understand the cause of this result. It might be related to source prompt differences as well as to a trade-off with the other dimensions of textual quality, e.g. complexity and fluency, which both showed a positive development (Skehan, 2009). The main changes in complexity concern syntactical and lexical diversity.

Table 6 Changes in CAF between pre- and post-tests, IS2 (means for 12 students)

Dimension of text quality	Measure	Explanation	Pre	Post	Change pre>post
Fluency	W/T	Total number of words	154	190	+24%
Accuracy	EFC/C	Error-free clauses per clause	.53	.39	-26%
Complexity	}	DC/C	.32	.31	-3%
			(dependent clauses per clause)		
		W/C	9.25	9.72	+3%
			(words per clause)		
		UBR	12.06	13.51	+12%
	MPL	14.83	15.02	+1.3%	
	GTTR	9.12	9.76	+7%	
	MWL	6.18	6.15	-.5%	

To investigate the use of cohesive ties, a difference was made between logical connectors and several co-reference devices. Their frequency in pre- and post-tests was compared, taking into account the total number of clauses. The results are presented in Table 7 (means for 12 students). They point towards a trade-off between the use of synonyms (pre-test) and recurrence (post-test), and a stable use of the other cohesive ties types.

Table 7 Changes in cohesive ties use between pre- and post-tests, IS2 (means for 12 students)

Cohesive type	ties	Explanation	Pre (per clause)	Post (per clause)	Change pre>post
Recurrence		repetition of antecedent	.33	.50	+50%
Synonym		including words derived from the same stem	.16	.10	-31%
Pro-forms		personal, relative, and demonstrative pronouns	.38	.37	-3%
Logical linking devices		conjunctions, adverbs	.56	.55	+1%
Total		All types	1.43	1.52	+6%

To conclude, it can be stated that there is no clear difference between pre-and post-tests with regard to the global holistic score (all students), the general text quality features

based on the CAF dimensions, and the use of cohesive ties (selection of 12 students, i.e., one third of the participants). The observed patterns for within-subjects difference for the latter two parameters are irregular. Therefore, no direct measurable impact of the intervention on individual writing proficiency based on text quality can be stated. Two possible explanations for these inconclusive findings are that a) a two-week intervention might be too short to observe a clear progress in text quality features in pre- and post-tests in text quality features, or b) the chosen analysis units (holistic scores, CAF measures, and cohesion) might not be transparent for changes from pre- to post-tests. In addition, the difference in writing prompts for pre- and post-test (input length, complexity) might have caused a confounding effect.

Consequences that were drawn from these insights for further research (i.e., the design of IS3) include the following decisions:

1. New measures should be defined to capture the changes targeted in the intervention instead of using general CAF measures seem to be better suited to investigate long-term changes in writing
2. The comparability of pre- and post-test prompts should be guaranteed in order to avoid bias.²⁷ Furthermore, the pre- and post-test should concern the same task as the intervention in order to draw conclusions about a possible learning effect originating from the intervention on post-hoc individual writing.

4.3.1.2 Attitudes of participants towards the intervention

To investigate the attitudes of the participants towards the online intervention, more specifically, the tools and pedagogical approaches used, an anonymous post-hoc questionnaire was administered (see Appendix 9b). In the following, the main aspects concerning participants' attitudes towards a) the technologies and b) the innovative aspects of writing pedagogy implemented are summarised, illustrating the results from closed questions with quotes from open answers.

Technologies used

Concerning the use of technologies for a learning-to-write task, 73% of the participants declared in the post-hoc questionnaire to prefer a writing course with online technologies to one without. This positive general statement is remarkable given the

²⁷ Actually, the best solution to avoid bias due to writing prompts is to apply a cross-over design, i.e., administer each prompt in the pre-test to one half of the group, and in the post-test to the other. Unfortunately, I only got acquainted with this method after carrying out IS 3.

fact that online learning was new to 85% of the students. Some of the open answers allow to link this positive attitude with the tools used:

Before, I would have said that online learning can never have the same effect as a F2F lesson. Now I think it's a good idea to combine both.

I am a dummy in computers, but I learnt a lot, for example how to use Google docs, and I think I'll be using this again.

Before, I didn't even know it was possible to write texts online, and on top of that in a group. Furthermore, it is easier to use online dictionaries and other online resources.

Google Docs is very useful for collaboration, very efficient. It really improved my experience with collaborative writing.

As the above statements corroborate, students found Google Docs (in the following abbreviated as GD) - which was unfamiliar to 67% before the intervention - very useful for collaborative writing and reviewing. An interesting aside is that participants did not use GD for individual composing (outside class), although they had to hand in their finished texts as a Google document. Instead, they preferred writing in their habitual word processor and cut-pasted their finished texts into a Google document. This was the case for about 80% of the individual texts, where only a last editing, or sometimes even no action, could be tracked in the document's history facilitated by GD. This hindered a comparative investigation of the individual and the collaborative writing processes.

In the same vein, the usefulness of the modules in *curios* to scaffold individual content elaboration and planning was rated very high (see Table 8). The online forum on the institution's VLE in combination with the script provided by means of a learning path for collaborative content elaboration and planning was judged less appropriate, to the extent that 21% of the students explicitly stated they would rather do the collaborative planning F2F than online.

Table 8 Perceived usefulness of tools for individual and collaborative writing used in IS2

Tool	Implemented in	(very) useful	(rather) disruptive
<i>curios</i>	Individual content elaboration and planning	93.5%	6.5%
Google Docs	Writing and reviewing	89%	11%
Forum	Collaborative content elaboration and planning	48%	52%

To sum up, the technologies implemented in this intervention study, apart from the forum which was found too cumbersome for quick interactions, had a high degree of acceptance by the participants.

Innovative aspects of writing instruction

Regarding the attitudes towards new forms of writing instruction that were implemented in this intervention, in the following, two aspects are elucidated, viz. the collaborative writing condition and the focus on writing strategies.

Individual vs. collaborative setting for learning-to-write

In the post-hoc questionnaire, 66% of the students declared to prefer an individual online learning setting to a collaborative setting for writing instruction. Especially the synchronous collaborative planning in the forum was perceived as chaotic by learners, in part because script instructions were disregarded by individual group members, as the following statements exemplifies:

I was rather critical in the beginning, because personally, I am not a fan of online learning. But it includes so many facilities: the teacher and the students can easily communicate. However, I think the facilities don't suffice for good collaboration. It was too complex and difficult. As always: when somebody doesn't collaborate, e.g., not answering on forum posts, this is a pity.

Therefore, 90% of the learners preferred individual (guided) content elaboration and planning. Concerning the actual writing process, an individual setting also was preferred to a collaborative setting by 77% of the students. The positive aspects of individual writing that were mentioned are mainly related to pace (mentioned 4 times), style (2 mentions), and independence (2 mentions). These statements are in part mirrored by the mentioned negative aspects of collaborative writing which include dependency on peers (5 mentions), "unequal workload" (4 mentions), i.e., the freeriding behaviour of group members, the fact that collaboration is time-consuming (4 mentions), sometimes "chaotic" (6 mentions), and "difficult" (5 mentions).

However, there were also many positive aspects of collaboration mentioned in the open answers. Most of them regard the collaborative reviewing process, e.g., learning from your peers' feedback (5 mentions) as well as from their errors (2 mentions). This is corroborated in the fact that peer feedback was rated as helpful (3.7 on a Likert scale of 5), as the following quotes also illustrate:

I prefer to write individually AND receive peer feedback.

Peer feedback is more personal than a sample solution.

Furthermore, twelve participants commented that collaboration in writing can lead to a better product, five of them pointing at the benefit of collaborative discussions, as the following statement illustrates:

Collaboration leads to a deeper understanding of the content.

The collaborative task was also welcomed as an innovative element in a highly individualised curriculum:

It's good to collaborate for a change in a curriculum that mainly promotes individual work.

To conclude, one interesting remark should be highlighted and discussed. It concerns the nature of peer feedback in an educational setting, stating that it was "not always spontaneous but a bit fake". This remark is corroborated by a curious and coincidental finding of a "secret conversation" between three male students while collaborating in GD. As the collaborative writing took place outside the classroom, the process could not be recorded. Instead, the version tracking function of GD was used to reconstruct the text genesis and the contributions of the individual students to the joint product. This function allows to view both the text creation and the comments on the text posted in the margin. However, the chat that is activated when two or more people are logged in synchronously, definitely disappears after closing the document, and is not visible in the document history. This was not the case for this specific group document. Apparently due to a technical bug, the chat conversation that the group members had on the night before the deadline to hand in their final text reappeared in the main document (see Appendix 15). From the conversation, it can be concluded that these students used the chat as a "parallel channel" to discuss not only workflow issues (who reads and comments on which text part), but also the formulation of the comments to be integrated in the main document. It is important to note in this context that the quality of peer comments was included in the individual score for the task that accounted for 50% of the whole collaborative task score. Having this in mind, and clearly feeling unobserved, these students planned and formulated comments in Dutch, or rather in a typical chat-interlanguage interspersed with emoticons and Flemish dialect, which they would then integrate in the main document, translated to German. They also kidded each other in the chat, reacting on comments posted by the peers in order to show their commitment and to "please the researcher". The following nice fragments of this "secret conversation" that were clearly not meant for the researcher's eyes illustrate the possible lack of authenticity of peer conversations in (online) collaborative tasks (translated to English from Dutch original):

Student 2: I have integrated no less than 5 idioms. Fishing for good points

Student 1: brilliant

Student 3: Wait, we'll discuss this in German

Student 1: I don't see any more issues in the main part

Student 3 [who wrote this fragment]: of course!!! It's totally perfect

[Student 2 writes a positive comment on this fragment in the document]

Student 1: haha, [Student 2], funny cunt

Student 3: haha [Student 2] :) Wait, I'll add a few good points to your account at Carola's [referring to the researcher]

Student 3: she's not going to understand a fucking thing of this [referring to the posted comments], fuck if she's got to do this with everybody [referring to the researcher reading all comments]

Student 2: haha true :p

In Activity theoretical terms, this "secret conversation" reveals a primary contradiction between their "double role" both as participants and students in the activity, i.e., a contradiction between "use value" and "exchange value" (Engeström, 1987, p. 130). In order to achieve their main goal (object) of getting good marks, they faked a spontaneous collaborative revision to comply with the set rules and with the alleged object of the researcher. However, while "faking" their commitment, they actually *did* (quasi incidentally) engage in the intended reflection about their writing.

Focus on writing strategies

The writing intervention using online technologies was perceived as (very) useful by 71% of the participants, 10% were undecided, and 19% did not find it useful. Two thirds of the students stated that their writing had changed through the intervention, pointing more specifically to their increased attention to cohesion and coherence (23 mentions in open answers), planning (10 mentions), and structure (3 mentions). Being forced to compose a skeleton, i.e., an enriched outline, prior to writing the actual text was considered as (rather) helpful by 80% of participants, vs. 20% who found this strategy too time-consuming.

I think more about my sentences, especially when I use "aber". Then I ask myself whether there's really a contradiction involved or whether I just like to use the word. Mostly, it's the latter.

Three weeks ago, I always had a dictionary at hand when writing. Now, I feel more secure about choosing the correct cohesive ties. Therefore, I write faster now.

First think about the structure of the text, then write. Not just start to write.

These quotes from answers in the post-hoc questionnaire exemplify that students indeed employed more strategies and, at the same time, that their meta-cognitive awareness regarding their own strategy use increased through the intervention. On a cautionary note, again, the possibility should be kept in mind that the students wrote these reflections merely to please the researcher. However, as the questionnaires were filled in online and anonymously, the motive of "fishing for good points" that was revealed in the previous section can be excluded here, and therefore, this explanation seems rather unlikely.

4.3.2 Discussion of the study in the light of Activity Theory

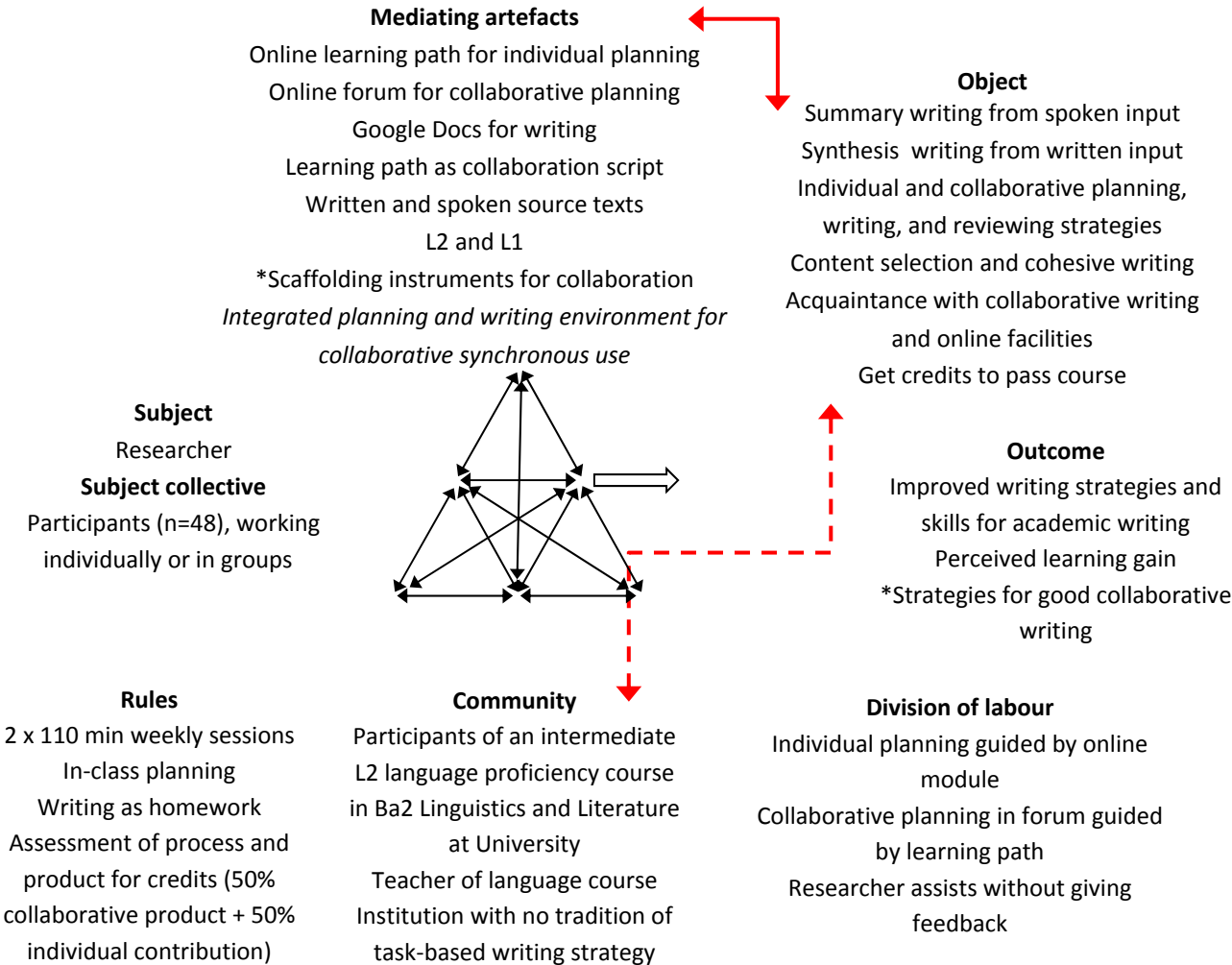


Figure 13 System of the educational activity of IS2

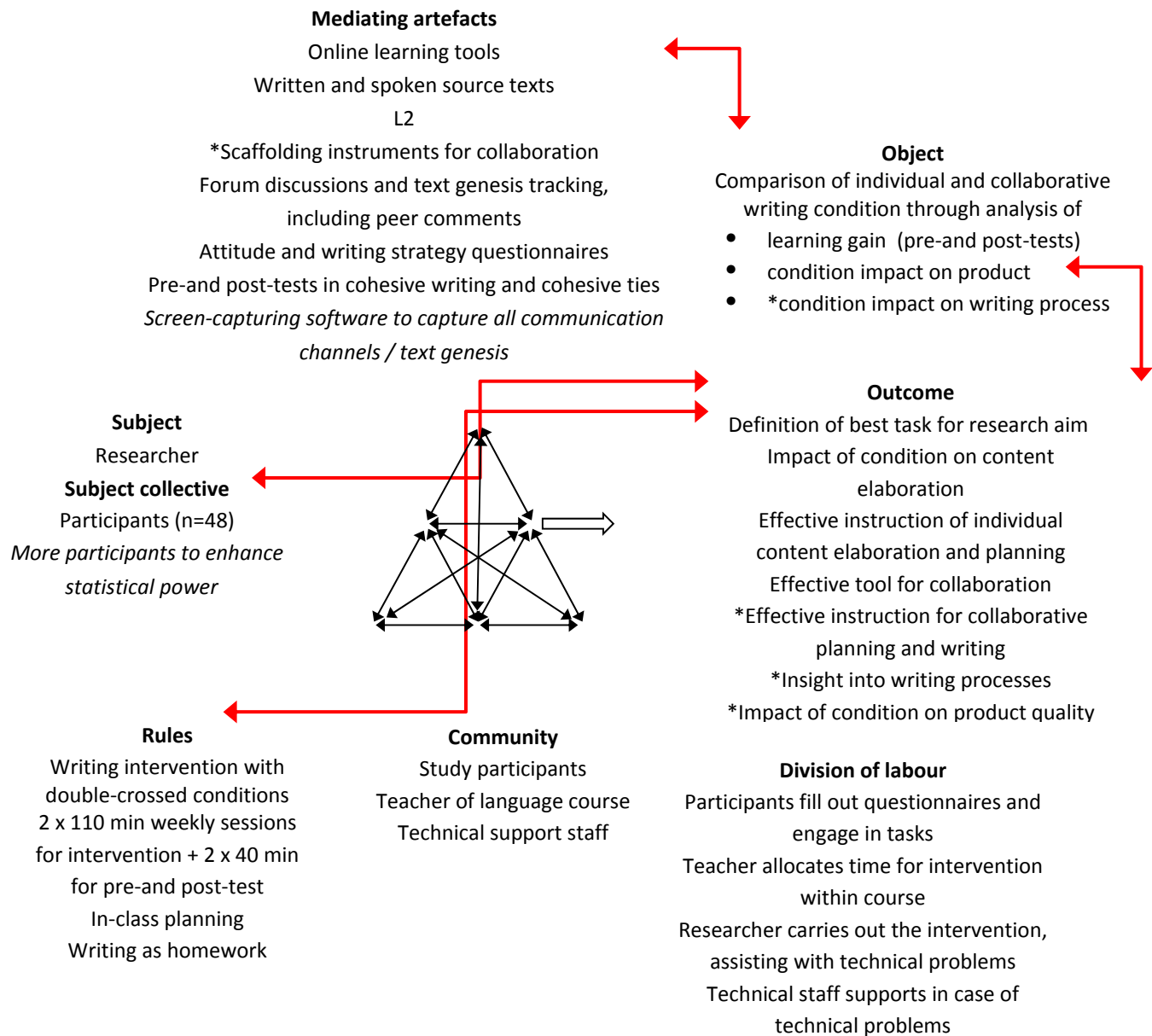


Figure 14 System of the research activity of IS2

Contradictions in the AS of the learning and the research activities

The graphical representation of the AS of the educational activity (see Figure 13) reveals one contradiction that arose in the activity of the IS2 between the mediating artefacts used and the object of the activity. More specifically, the combination of forum and online path used to scaffold and facilitate collaborative content elaboration and text planning turned out to be counterproductive. The required window-hopping caused students to disregard the work flow script presented in the learning path. In many groups, the collaborative planning was therefore perceived as chaotic and unsatisfying.

The dashed double-sided arrow indicates a *possible* contradiction (my own convention). As the study programme of Linguistics and Literature is traditionally rather content-focused than skill-focused and language courses are taught in a rather theoretically-based way following a grammar-translation approach, the students are not used to task-based skills training and strategy instruction. This could have caused a tension between the object of the activity and the subject collective. However, as the attitudes expressed in the questionnaires revealed, this was not the case. Instead, students welcomed this (short-term) intervention as an innovative element affording the curriculum fresh scope.

The AS of the research activity (see Figure 14) reveals four important contradictions, three of them involving one objective of the research, i.e., the comparison of the individual and the collaborative writing condition by means of an analysis of the writing processes and the written products.

1. The first contradiction regards the number of participants (*subject collective*). It was stated in the quantitative analysis of the products (*outcome*) that the statistical power of the tests was very low due to the low number of collaborative groups (n=7 in each condition).
2. The second contradiction concerns the *mediating artefact* used for the process analysis, viz. Google Docs. As explained above, this tool facilitates the a posteriori tracking of text genesis and comments, but not of the communication in the chat which also plays an important role in the writing process.
3. The third contradiction is closely related to the second inasmuch as the writing process investigation was also hindered by the *rule* that the writing process itself took place outside class. This led to the situation that students used a different tool than the one foreseen for individual writing which made tracking impossible. Furthermore, it is possible that groups met F2F in order to carry out the collaborative task, and consequently, valuable information about the processes might have been lost.
4. Finally, an intended *outcome*, viz. the detection of an impact of the "individual vs. collaborative writing" condition on the product, was only hinted at by

means of the analysis (*object*). It was found that the units of analysis used only to a limited extent allowed to detect an impact of the condition.

Consequences for the design of Intervention Study 3

1. From the contradiction found in the educational activity concerning the affordance of the combined mediating tools forum and learning path, the consequence was drawn that the design should be adapted in order to facilitate a collaborative workflow in *one* environment/tool. Furthermore, the collaboration should be scaffolded by other means (than an online micro-script) in order to avoid ineffective work flow.
2. From the contradiction found in the research activity concerning the number of participants which led to a lack of statistical power, the consequence was drawn that more participants were needed than the usual class size of the study programmes involved in the intervention studies so far. Therefore, it was decided to carry out IS3 twice, in one class of each study programme each, i.e., Linguistics and Literature as well as Applied Language Studies. In addition, it was decided to carry out IS3 with students in the *third* year of their undergraduate programme in order to be able to focus exclusively on writing strategies concerning HOCs (content elaboration and selection) and abandon the additional focus on LOCs (e.g., grammar, cohesive ties), assuming that third-year students would need less input regarding LOCs than second-year students due to their higher general proficiency level.
3. The data loss regarding the writing process due to the combination of the tool used, GD, and the rules applied, i.e., writing as a homework, led to the insight that the whole process from planning through composing to reviewing should occur inside the classroom, and that a screen-capturing tool was needed to facilitate an analysis of *all* interactions between collaborating peers, including the chat conversation.
4. Finally, the most important insight came from the contradiction between research object and outcome concerning impact of condition on the final product, i.e., the texts. It was found that a thorough comparison of both settings was very complex and would require the definition of units of analysis that combine process and product. Furthermore, a clear impact of the collaborative setting was detected regarding content elaboration and selection through fruitful discussions in the planning phase, which is the utmost important strategy to be acquired when engaging in writing from sources. For those two reasons it was decided to rather focus on the collaborative setting, abandoning the comparative research approach.

Chapter 5

Collaborative online writing, supported by a script and a model (Intervention study 3)

5.1 Background information on the study

"Bear in mind that students need and want constraints in order to explore individual freedom in collaborating"
Jim Slotta, EARLI 2013, Symposium CSCL

5.1.1 Overall study design

The third intervention study was carried out in the spring semester of 2012-13, more specifically during five consecutive weekly lessons of 110 minutes each in February and March 2013. In order to increase the number of participants and thus enhance the representativeness of the study, it was carried out in parallel at two institutions, as a part of two different courses. The first group of participants (n=42) followed a course in advanced writing which is part of the third-year study programme of Applied Languages at HoGent, the second group (n=34) participated in a linguistics course which is part of the third year programme of Linguistics and Literature at UGent.

The study targeted collaborative writing of syntheses from several written sources as a preparation for the students to write literature reviews for their bachelor paper. The main aim of the study was to explore the impact of instructional support in the form of a script and a model video on collaborative writing via the online editor Google Docs. Figure 15 illustrates the original overall research design. The independent variable in this intervention was the condition "scaffolding sequence" (see next section). The students produced three collaborative syntheses during three successive weekly sessions of the course. Proficiency was used as control variable in the sense that the collaborative groups were composed according to proficiency levels in low-proficiency, high-proficiency, and mixed-proficiency triads. As a second control variable, the individual learning style of the participants was taken into account which was determined through a questionnaire administered before the intervention (see Appendix 20). The dependent variables are a) the collaborative writing process that was carried out in Google Docs and recorded with the screen-capturing software *Morae*, and b) the collaboratively produced syntheses. The second research question concerns the individual learning gain of the students through this online collaborative writing experience. To investigate their progress, students wrote individual syntheses before and after the collaborative intervention as pre- and post-tests. Furthermore, students' attitudes towards the activity were monitored by means of pre- and post-hoc questionnaires and post-hoc focus groups.

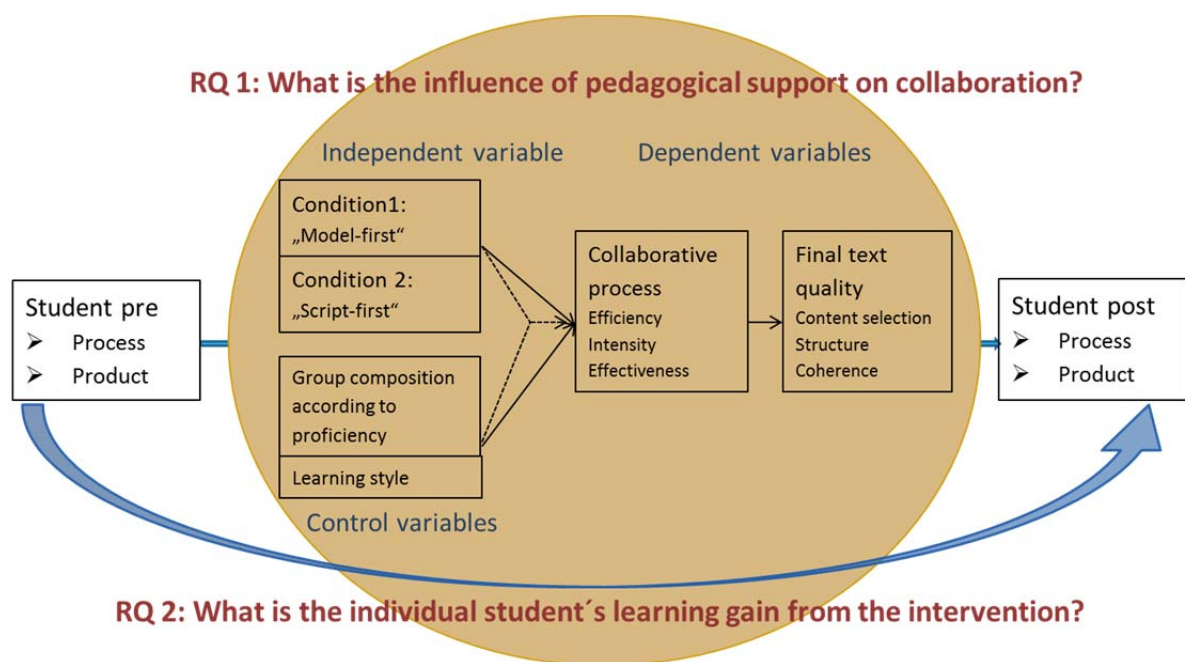


Figure 15 Overview over whole study design and main research questions of IS3

The results of the intervention study in the light of the two main research questions were disseminated in two international peer-reviewed publications. The first question concerning the impact of scaffolding instruments on the collaborative writing process is being dealt with in the book chapter "Learning to think and write together: Collaborative synthesis writing supported by a script and a video-based model" (see chapter 5.2). The second question concerning the individual learning gain gave rise to the article "Learning to write syntheses: Changes in individual writing after collaboration" (see chapter 5.4). The co-authors of the latter publication, Annelies Raes and Tammy Schellens, helped with shaping the focus of the article, determining and implementing the statistical tests to be used, and reviewing the text. Furthermore, a third publication emerged out of the qualitative analysis of the collaboration process: "Assessing the quality of computer-supported collaborative writing in a foreign language" (see chapter 5.3). The first author of this publication, Lara Hoste, was a student of the "Master in multilingual communication" programme at UGent in 2013-14 and wrote her Master thesis entitled "Qualitative Analyse von online Schreibkooperationen anhand eines selbst entwickelten Bewertungsrasters" under the supervision of this PhD's author. For her thesis, she carried out a fine-grained analysis of the collaboration in three triads. To this aim, we developed an assessment scheme for CSCW based on a validated instrument for the investigation of CSCL processes (Meier, Spada, & Rummel, 2007). After finishing her thesis, she analysed the collaboration processes of three more triads based on the new instrument, working as a job student. We presented the results of this part of the research at the Conference on Writing Research 2014 in Amsterdam in a joint paper presentation (comparison of the two

conditions based on the qualitative process analyses) and in a poster presentation (development of the new rating scheme) by Lara Hoste. Based on these presentations and Lara Hoste's Master thesis, the article was written that presents and discusses the new rating scheme and its validation through the qualitative process analyses.

5.1.2 Design of the collaborative activity

Although the design of the collaborative activity is described in a nutshell in the article presented in chapter 5.2, in the following, a more elaborated version is provided, together with a rationale of the design principles.

The two groups formed the basis for the main conditions which were defined by the sequence of support instruments applied to scaffold students in the collaborative writing process (see Figure 16).

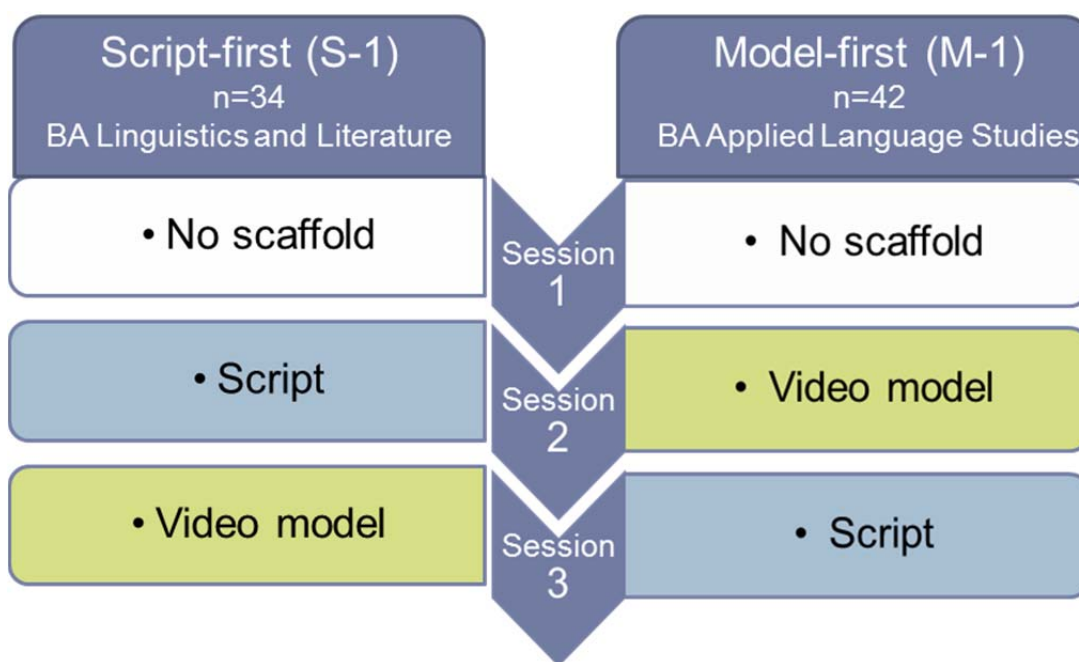


Figure 16 Study design of the collaborative activity: crossed conditions

Intentionally, in the first collaborative session, there was no scaffolding offered. This has a twofold reason. First, in line with a socio-constructivist approach to learning, it was interesting to see how the collaborative groups would be dealing with this complex task, and what solutions they would find to problems they encountered during its accomplishment. Second, from a research perspective, this first learning-by-doing session provided the ground to investigate the impact of the (different) scaffolding mechanisms on the collaboration processes that would be implemented in the following session. In the second and third collaboration session, a script and a model video were

administered as scaffold instruments in a cross-over design in order to explore whether the implementation order mattered, i.e., whether "script after model" had a different impact than "script-first", and idem for the model. The quantitative results of the comparison between the two conditions were presented at the CALL 2014 conference in Antwerp and a short paper was integrated in the conference proceedings. In chapter 5.5.1, the main outcomes are highlighted.

Designing a collaborative activity in an online environment requires careful consideration of several factors. Strijbos et al. (2004, p. 418) established a list of 29 questions grouped into six interconnected considerations (type of learning objective, expected interaction, task type, structure, group size, and computer support) that need to be addressed when designing CSCL activities. In the following, a brief overview of the decisions that shaped the design of IS3 is provided, based on Strijbos et al.'s (2004, p. 418) catalogue of questions.

The learning objective of the collaborative activity can be subdivided into four sub-goals that are identical for all learners: (a) elaborate source content in view of synthesising it, (b) plan and write the text finding one's own structure / introduce the writers' voice(s) for the synthesis, (c) revise the text focusing both on LOCs and HOCs, and (d) establish an effective and efficient work flow for the collaborative online writing activity. According to Strijbos et al.'s classification, the first three sub-goals target "open skills" that can best be trained in *ill-structured* tasks. These are characterised by the existence of multiple or open solutions and by their authenticity, i.e., they reflect the complexity of real-life problems without outlining sub-goals (Lodewyk, Winne, & Jamieson-Noel, 2009). However, the fourth, subservient goal can be considered a closed skill that calls for a high level of pre-structuring. In the design of IS3, the activity was ill-structured only in the first, unscaffolded collaboration session. In the second and third collaboration sessions, a script and a video model were used as scaffolding mechanisms to structure both the activity and the collaboration work flow (see Figure 16). This was deemed necessary to prevent cognitive overload during task performance. After all, the learners had no prior experience with collaborative writing, let alone via online channels. In this context, it has to be noted that the level of task control plays an important role in motivation. On the one hand, when the level of (external) task control is very high, e.g., by providing a fine-grained script that leaves few options to be explored, this can lead to motivational loss. On the other hand, a very low task control can lead to "mathemathantic" situations for learning, especially in a complex task (Kirschner et al., 2004).

The targeted activity in IS3 fully corresponds with the definition of "collaboration" by Ede and Lunsford (1990), as it involves (a) the substantive interaction in all stages of the writing process, (b) the shared decision-making power over, and responsibility for, the text produced, and (c), the production of a single written document (Ede & Lunsford, 1990). However, the actual text production stage in this particular case can

rather be described as a *cooperation*, as it involved individual writing. Therefore, both synchronous and asynchronous communication modes needed to be supported, although the activity was preponderantly based on synchronous CMC. The collaborative interaction was situated on three different levels: The students needed to (a) coordinate the work flow (synchronous mode), (b) exchange ideas about the relevance of content items and about the structure of the collaborative text (synchronous and asynchronous), and (c) give each other feedback on individually written parts of the collaborative final product (both in the synchronous and the asynchronous mode).

To facilitate the interaction on these different levels, a groupware that provides different communication channels was needed. Several options were compared, among which was TC3. TC3 is a groupware that was specifically designed for collaborative writing based on pedagogically sound principles, providing different spaces and facilities for collaborative planning, translating, and reviewing (Erkens, Jaspers, Prangma, & Kanselaar, 2005). However, as the use of this tool within the given institutional context was not straightforward, and technical support in case of possible problems with tool stability could not be assured, I opted to use Google Docs (GD). This groupware provides parallel channels for chat (synchronous), comments (asynchronous), and the emerging text itself (synchronous and asynchronous). Moreover, automatic backups are made every few seconds, and the retrieval of earlier versions including the visualisation of changes between versions is facilitated. In this way, in case of incidental data loss through invalid recording files, GD tracking history could be used as a safety net. However, GD is less suited for simultaneous interaction in the emerging text, as the document is automatically saved every few seconds, and therefore, version conflicts can arise when simultaneous changes are implemented. Therefore, a separate space for individual writing was needed. For this function, MS Word was chosen, as students are familiar with this word processor and like to rely on its editing facilities (e.g., the spellchecker).

Fruitful collaboration requires a feeling of joint ownership of the product. As Kirschner et al. (2004) stated, the need for this feeling is based upon the two pedagogical principles of individual accountability and positive interdependence. To assure individual accountability, i.e., increase the extrinsic motivation for equal participation of all group members in the collaboration, the jigsaw principle that is often applied in CSCL tasks (Dillenbourg & Jermann, 2007) was used, i.e., each student was only assigned one of the source texts (see Figure 17). In this way, every student was actually forced to elaborate the content of "his/her" source in order to provide the peers with the necessary background information to proceed with the synthesis task. The grading system that was applied reflected this positive interdependence, as all students received two scores, viz. an individual one for their contribution to the joint discussion and product, and a group score for their collaborative product. Furthermore, positive interdependence helps to prevent undesirable behaviour like free-riding behaviour that

can entail motivational loss of productive group members, also called "the sucker effect" (Kerr & Bruun, 1983).

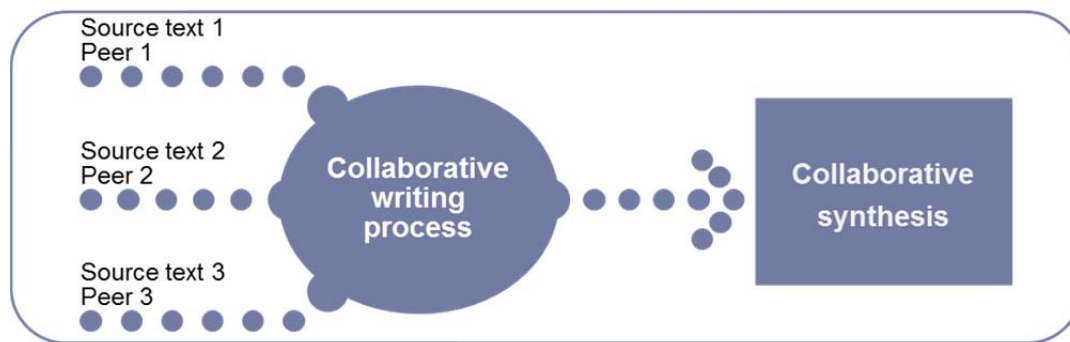


Figure 17 Jigsaw principle for the collaboration in triads

The group size was determined based on previous literature on L2 collaborative writing. In focus-on-form tasks targeting peer revision, pair interaction has been studied most frequently and has been found to be beneficial (Hanjani & Li, 2014; Storch, 2005; Wigglesworth & Storch, 2009, 2012). However, the required content elaboration and restructuring in a synthesis task implies a need to negotiate. For this kind of interaction involving possible disagreement, CSCL research has pointed out small groups as preferable size (Strijbos et al., 2004). Therefore, a group size of three was implemented in this intervention study.

5.2 Learning to think and write together: Collaborative synthesis writing supported by a script and a video-based model. (Chapter in: Guasch, T. & Deane, M. (Eds.), *Learning and teaching writing online: Strategies for success*)²⁸

Abstract

Synthesis writing is a cognitively and linguistically demanding task, and even more so in a foreign language. Collaboration can help students to learn how to elaborate the content of the source texts and integrate it in a new coherent text. But to be effective, online collaboration has to be pedagogically supported. In this chapter, a promising approach to support students in this complex task is presented and evidence for its effectiveness is provided. The effect of two instructional support instruments, a video-based model and a script, on the development of collaborative writing processes is described. Using a mixed-methods approach, collaboration intensity and efficiency are measured quantitatively, and collaboration effectiveness is observed qualitatively. The data originate from an in-class intervention study: 42 university college students grouped in triads wrote three syntheses in subsequent sessions, collaborating in Google Docs, and being supported in two of the sessions. The analysis results confirm the benefit of the support instruments used and broaden the understanding of collaborative writing processes: qualitative insight shows that joint information processing is intensified through scaffolding, and leads to a better content selection and to good peer-induced revisions of the jointly produced text on different levels. Regarding of collaboration efficiency, work flow was improved through scaffolding, albeit to a less noticeable degree for the high-proficiency groups.

Introduction

Learning to write a synthesis from multiple source texts is important preparation for academic writing. At the same time, it is a challenging task that imposes a high cognitive load on student writers. It is hybrid, as it involves both reading and writing skills, and it is complex, both on content and language level. Regarding content, writers

²⁸ Strobl, C. (2015). Learning to Think and Write Together: Collaborative Synthesis Writing, Supported by a Script and a Video-based Model. In G. Rijlaarsdam (Series Ed.) & M. Deane & T. Guasch (Vol. Eds.), *Studies in Writing: Vol. 29, Learning and Teaching Writing Online* (pp. xx-xx). Leiden: Brill. [print publication expected in April 2015]

are required to elaborate and integrate information from multiple sources. The degree of elaboration is the key to understanding the sources (Solé, Miras, Castells, Espino, and Minguela, 2013), which, in turn, is a *sine qua non* for producing an effective synthesis. Untrained student writers, though, tend to display reproductive, rather than elaborative, patterns when writing from sources: They copy-paste source fragments in the given order, occasionally reformulating chunks, but rarely summarizing and combining ideas in their own words (Solé *et al.*, 2013; Raedts, Rijlaarsdam, van Waes, and Daems, 2007).

Regarding language use, the challenge for untrained student writers lies in the required focus on text level (coherence and cohesion), whereas they naturally tend to confine their attention to word or sentence level when writing and revising (Paulus, 1999; Baurmann, 2002). This tendency is even more prominent when students produce syntheses in a foreign language (L2) (Van Steendam, Rijlaarsdam, Sercu, and Van den Bergh, 2010). This chapter presents a promising approach to support the students effectively in the learning process of this complex writing task: A repeated collaboration in groups of three, scaffolded by a video model and a script.

A beneficial effect of peer collaboration on foreign language (L2) writing has been documented by various scholars for different tasks, proficiency levels, and settings (Elola and Oskoz, 2010; Kessler, Bikowski, and Boggs, 2012; Kost, 2011; Storch, 2005; Würffel, 2008). On a cognitive level, peer collaboration has been shown to foster reader-orientation (Blin and Appel, 2011; Würffel, 2008). On a metacognitive level, the benefit of collaboration lies in the raised awareness of effective strategy use (Würffel, 2008). Finally, there is also an affective bonus consisting in reduced anxiety of the individual writer through the shared responsibility for the outcome (Würffel, 2008). In L2 synthesis writing, two more important benefits can be added to the ones already mentioned: (a) Peer collaboration can help students to master the complexity of this task by pooling their linguistic (Storch, 2005) *and* cognitive resources; (b) Peer collaboration can help overcome the reproductive patterns observed in inexperienced synthesis writers because the setting creates an extrinsic need to exchange information about, and therefore cognitively engage with, the source information.

For this collaboration to be effective, students need to be supported in the collaborative process (Slavin, 1992) and trained in peer feedback (Min, 2005). But what type of instructional support works best for this specific task and target public, that is, advanced foreign language students in higher education? Does this depend on the group's proficiency level? These are the main questions of the empirical study described in this chapter.

Two state-of-the-art instructional methods to support online collaboration are *scripting* and *observational learning*. Scripting is a well-researched instrument in Computer Supported Collaborative Learning (CSCL) (Fischer, Kollar, Mandl, and Haake, 2007). It has been stated that the main effect of a script consists of a better organization

of the collaborative process (Rummel, Spada, and Hauser, 2009), and that it depends on group characteristics (Hämäläinen and Arvaja, 2009; Schellens and Valcke, 2000).

Observational learning has been shown to foster effective results both for individual synthesis writing (Raedts *et al.*, 2007), and for peer revision in L2 writing (Van Steendam *et al.*, 2010). The present study provides evidence for the following hypotheses, building on previous scholarship on computer-supported collaborative learning (CSCL) and collaborative L2 writing:

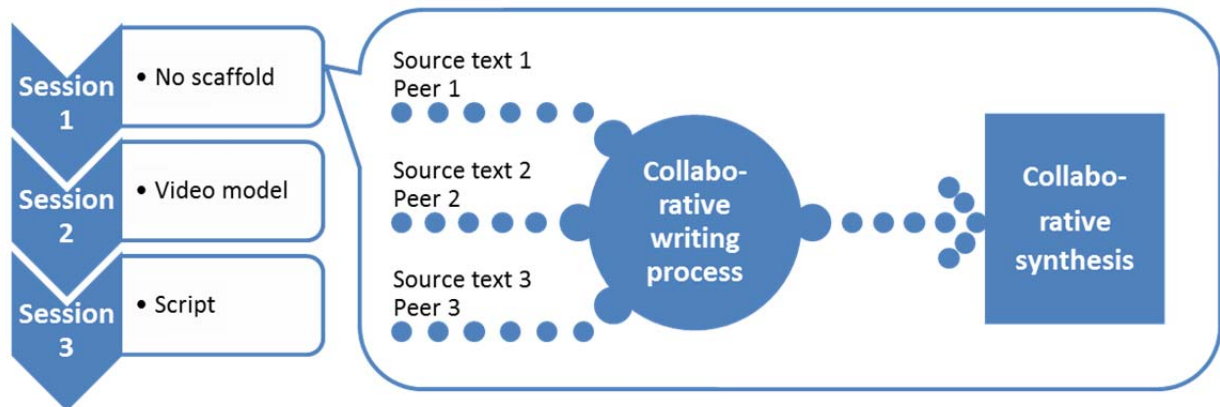
1. Instructional support leads to a higher interaction density which translates to a better outcome, both in terms of content elaboration (Rummel *et al.*, 2009) and in terms of language accuracy (De la Colina and García Mayo, 2007).
2. Instructional support leads to a better coordination of the collaborative process (Rummel and Spada, 2007).
3. Observational learning leads to increased amounts and, to a lesser degree, to diversification of peer feedback (Van Steendam *et al.*, 2010). More discussions and editing activities regarding both text level concerns and word or sentence level concerns take place after students have observed the model.
4. Both model and script exemplify successful collaboration and writing strategies which some high-proficiency learners already use. Nonetheless, the high-proficiency student groups also benefited from the instructional support, albeit more to reinforce the strategies they already used.

Study design

To measure the effect of both instructional support methods on computer supported collaborative synthesis writing, a three-week intervention study was carried out in an L2 writing class at a university college (see flowchart in figure 1 for an overview of the intervention and the task set-up). The participants ($n=42$) are third-year bachelor students of an applied languages programme including German L2. All are Dutch native speakers with an advanced language proficiency for German (B2-C1 of the Common European Framework)²⁹. To tackle the effect of proficiency level, the students were grouped in triads based on pre-test results into low-proficiency groups ($n_{\text{groups}}=3$), high-proficiency groups ($n_{\text{groups}}=3$), and mixed-proficiency groups ($n_{\text{groups}}=8$).

²⁹ Correspond with ACTFL Proficiency Guidelines levels ‘Advanced mid’ and ‘Advanced high’.

Figure 1
Flowchart of the intervention (left), and of the task structure in each session (right)

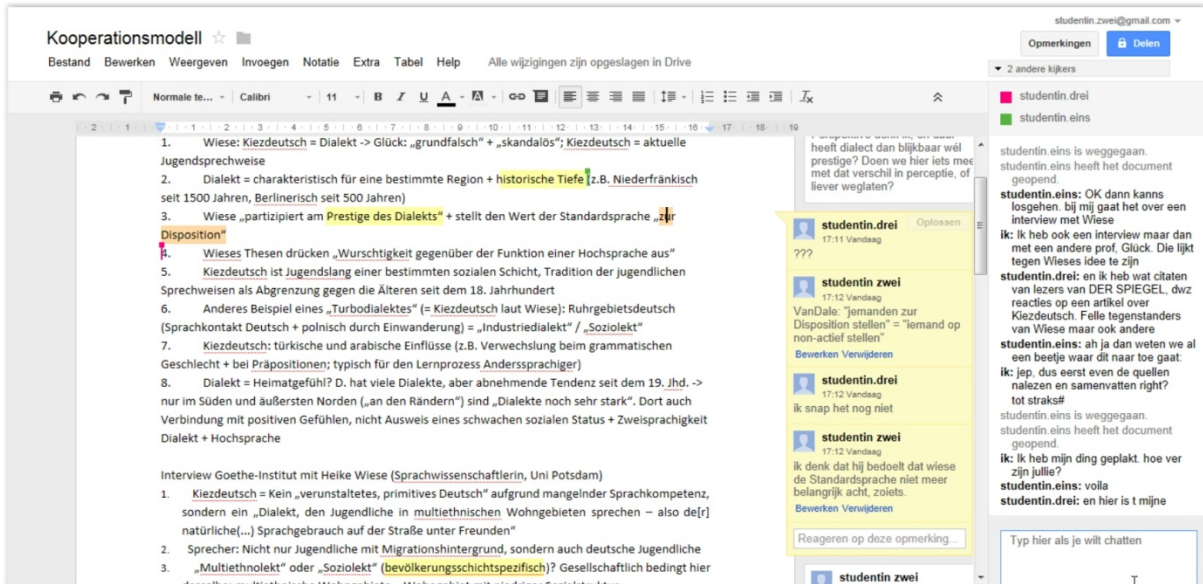


Each week, the students wrote a synthesis in German in these collaborative groups of three in class. They had 90 minutes on average to perform this task, which consisted in synthesizing the information of three popular scientific texts on a common topic in one collaborative text. Each week, a different topic was chosen, all relating to language variation. Before actually writing the synthesis, the main ideas of the three source texts had to be summarized. This part of the task was carried out individually, following the *Jigsaw concept* (Dillenbourg and Jermann, 2007): each student only read one source text and was held 'responsible' for summarizing its main ideas. This way, the individual commitment of all group members, which is a hallmark for effective collaboration, was ensured. At the same time, this information gap serves as extrinsic motivation to elaborate the content of the source texts through negotiating meaning. After all, every group member needs to get the gist of all source texts in order to proceed with the task.

The students used a text editor for the individual part of the task and an online document sharing and versioning tool, Google Docs, for collaborative synthesis writing. Google Docs allows for synchronous interaction in a text and provides communication facilities that match the different types of interactions occurring during text elaboration (see screenshot in figure 2): The chat lends itself to interactions of quick consensus building, like work flow discussions, whereas the comments, that are linked to a highlighted section in the running text, allow for in-depth discussions related to specific parts of the text, regarding both content and language use.

Figure 2

Screenshot from the model video illustrating communication channels in Google Docs: the collaborative text on the left, the comments in the middle, and the chat on the right



In session 1 ('no scaffold session'), all students received a short introduction to peer collaboration using Google Docs facilities, but were given no further instructions on how to proceed with the collaborative task. In the two subsequent weeks, instructional support was provided through a model and a script: In session 2 ('model session'), the students watched a screencast video depicting a collaboration model prior to engaging in their own task. This 13-minutes-video shows a collaborative synthesis writing process conducted by a dummy peer group. Crucial moments of the planning, writing, and revising processes are modelled, including commonly occurring problems. In session 3 ('script session'), the students received a macro-script (Dillenbourg and Jermann, 2007) that lists six process steps, including recommendations on the expected outcome, the tool to be used, and the time to be spent on each step.

All writing sessions were recorded and analysed using the screencast software Morae that allows for customized coding of process phases and single actions. These recordings were analysed using a mixed-methods approach: A quantitative approach is used to detect patterns in the collaboration of all groups in terms of the development of intensity and efficiency. The qualitative approach, zooming in on the interaction of three randomly selected groups, grants an insight into the effectiveness of the collaboration process.

Results: collaboration efficiency and intensity

Development of Collaboration Efficiency in Terms of Work Flow

One of the major challenges for all groups was time efficiency, as they had to accomplish the task within approximately 90 minutes. In order to reveal how groups organized their collaboration to meet this time limit, the task phases were coded and their duration was measured. The coding was based on the steps that were recommended in the script and depicted in the model (see table 1).³⁰

Table 1
Task phases used for work flow coding

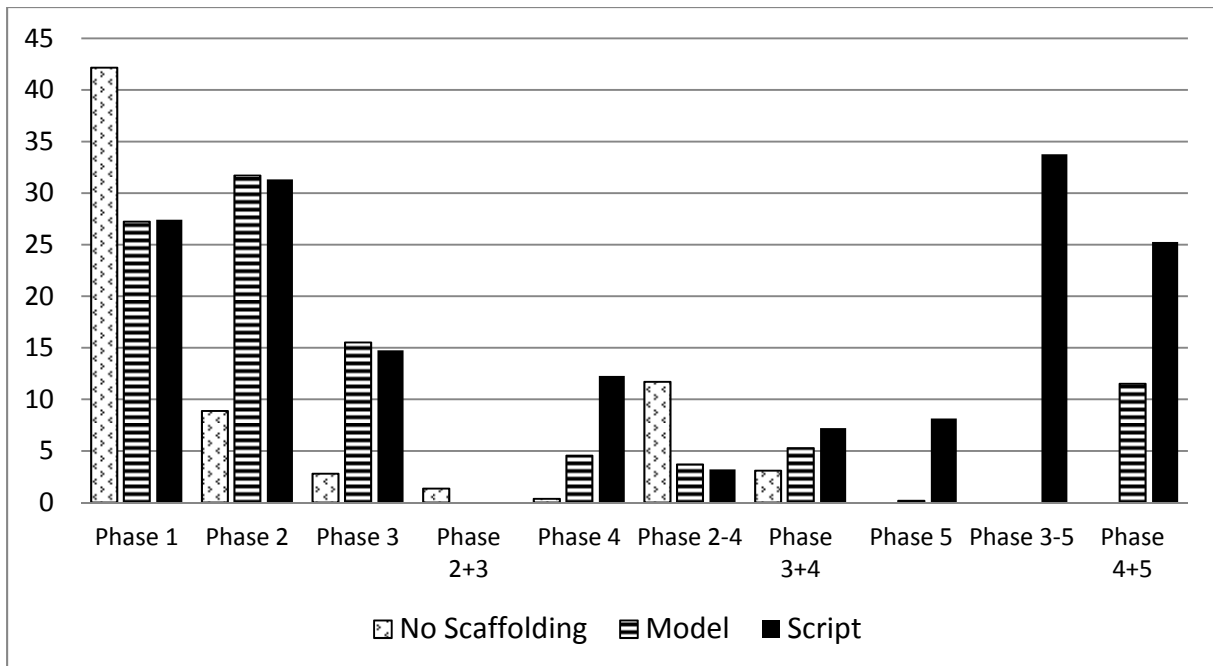
Nr.	Short description	Phase
1	Individual summary of one source text	} Planning
2	Collaborative content discussion and synthesis planning	
3	Individual writing of a synthesis fragment	} Writing
4	Collaborative integration of the fragments	
5	Collaborative revision of the final text	Revising

As the steps were only revealed in sessions 2 and 3, the triads developed their own work flow in the first session, which explains the occurrence of hybrid phases covering more than one of the task steps. Figure 3 gives an overview of the mean time spent on the task steps in each writing session.

³⁰ Steps 2 and 3 of the script were merged into Phase 2 for the quantitative workflow measurement, as they were carried out simultaneously, and therefore could not be distinguished from each other.

Figure 3

Time on task per phase - Mean of all groups per session of approximately 90 minutes (in minutes)



The time spent on the first individual preparatory phase, the summary of one source text, was reduced by one third on average in the scaffolded sessions (mean 1st session: 42.16 min > mean 2nd + 3rd sessions: 27.33 min). The model played an important role here: In the first session, 43% of the students wrote out a fully elaborated text summary of their source. After watching the model, all but four students followed the depicted strategy, jotting down the main ideas of the source text in numbered bullets. Conversely, the groups dedicated a lot more time to the second planning phase, the collaborative content discussion and selection after viewing the model. No group engaged in a final collaborative revision in the first session. In fact, most groups did not even get as far as integrating their individual fragments into a collaborative text. Besides inefficient time management and work flow - most groups tried to combine phases 2-4 in their Google Doc which led to rather chaotic synchronous planning and writing - an additional hindering factor was the fact that, due to unforeseen circumstances, the students had to finish the task in 70, instead of 90, minutes in this session. The quality of the few syntheses that effectively were completed in the no scaffold session was rather poor. Groups basically cut-and-pasted their individual summaries into one piece without integrating them at content or language level.

Furthermore, figure 3 shows a trend towards a greater diversification of the task phases in the scaffolded sessions, especially in the script session. Not only did the groups start earlier to actually write the synthesis (phases 3+4), the different task phases are clearly distinguishable in the work flow. This indicates that the steps modelled in the video, and explicitly written out in the script, were effectively followed. The script

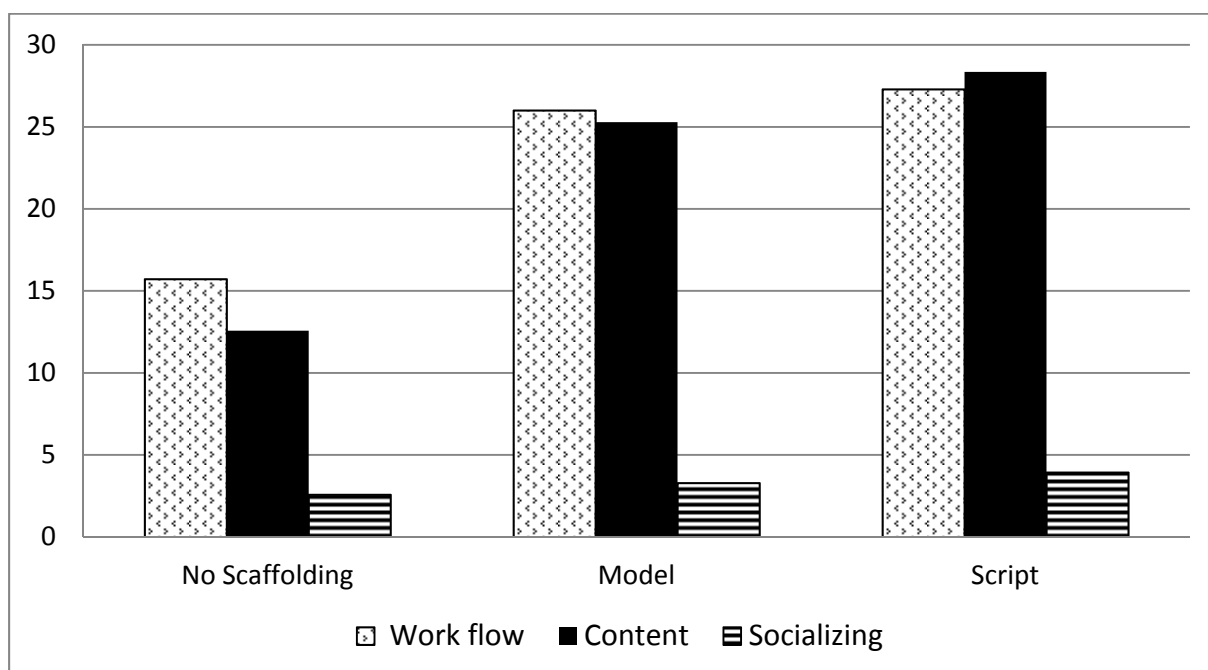
seems to have had a decisive impact on the task organization. It was only then that all triads finished their syntheses, and most triads revised their texts in a collaborative effort. This result should be qualified by saying that they were allowed an additional 10 minutes in the script session, a deliberate measure to facilitate collaborative revision, which, after all, forms part of the research focus. Other factors which might have contributed to the improvement of work flow, are a learning-by-doing effect and a cumulative effect. Indeed, based on quantitative data exclusively, it is difficult to discern whether the improvement was mainly an effect of the scaffolding mechanisms. Qualitative introspection in the next section, though, provides evidence for the high impact of pedagogical support.

Another interesting observation concerns the workflow of the groups according to proficiency. The high-proficiency groups already split the tasks in up to four phases in the no scaffold session, whereas in the other groups, only two (hybrid) phases can be distinguished, except for one mixed-proficiency group with three phases. This difference is less prominent in the model and script sessions, which indicates that the low- and mixed-proficiency groups benefited more from the scaffolding in terms of work flow.

Development of Collaboration Intensity in Terms of Interaction Density

In order to quantitatively measure collaboration intensity, the communication of the groups in two different channels was examined: (1) All chat turns were counted and classified according to their main focus on (a) work flow, (b) content, or (c) socializing, and (2) all comments were counted and classified according to their focus on (a) lower-order concerns (LOC), that is, discussions about issues concerning word or sentence level, such as lexical and grammar problems, or (b) higher-order concerns (HOC), that is, discussions on text level. The HOC discussions, in turn, were subdivided into content-related issues (HOC1) and issues related to coherence and cohesion (HOC2). The bar chart in figure 4 provides an overview of chat usage in all groups.

Figure 4
Number of chat turns per session, means of all groups

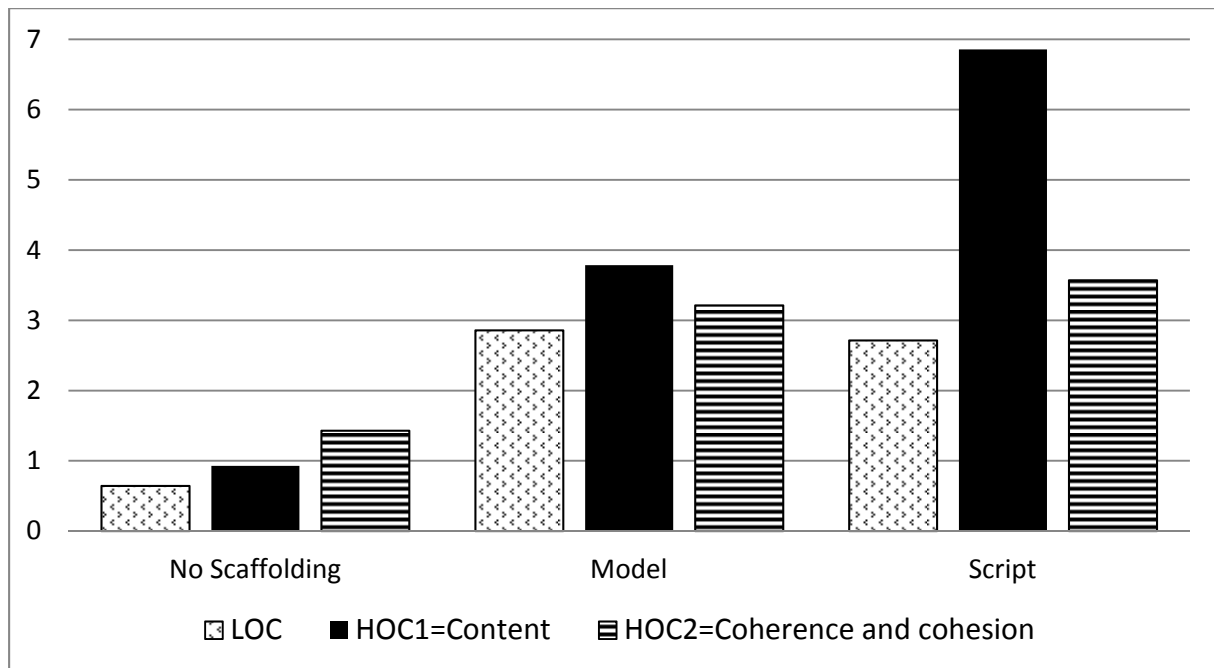


Chat related to work flow issues (for example, information about individual task step progress or discussions about the next step) is represented by the dotted columns, content-related chat (for example, discussions about the outline of the synthesis) by the black columns, and socializing chat (for example, building group spirit) by the striped columns. Overall, a considerable increase in chat usage can be observed in the model and script sessions compared to the no scaffold session. Work flow communication increased by more than 60 % in the model session, and remained stable in the script session. Content communication in the chat even doubled after viewing the model, and continued to increase slightly in the script session. The development of collaboration intensity through chat varied considerably between triads (for more details, see appendix 1). While in some triads, chat usage remained relatively stable throughout the three sessions, in others, observing the model literally led to a communication explosion in the chat. A decrease of chat usage in session 3 in six triads was related to a diminishing need to discuss work flow.

The second quantitative instrument to measure collaboration intensity is the comments that students gave on the collaborative text emerging in the Google Doc. For this analysis (see figure 5), only the original threats were counted, irrespective of the number of responses they triggered. All comments related to content or language, and were used predominantly in two specific task phases. In task phase 2 (elaboration of the outline), comments were often used to verify the meaning of a concept or wording, and to discuss the relevance of specific information for the common synthesis (these content discussions were labelled HOC1). In task phases 4 and 5, they were predominantly used for revision purposes, either to discuss the sequence of text

fragments (coherence) and the lexical links between them (cohesion) (labelled HOC2), or to discuss language use at the surface structure regarding vocabulary, grammar, or punctuation (labelled LOC).

Figure 5
Number of comments per session, means of all groups



In the no scaffold session, comments were scarcely used, although their function had been exemplified beforehand. In the model session, comment usage shows an exponential growth (LOC: x3.5, HOC1: x3.1, HOC2: x1.2). Interestingly, in the script session, the amount of comments related to *content* continued to grow by a factor of .8, whereas the communication intensity regarding HOC2 and LOC was not affected by the script. This corroborates the chat results. In both communication channels, content discussions increased throughout the collaborative writing experience. This result indicates that the students grew more aware of the importance of content elaboration and selection for synthesis writing in the course of the intervention study. Again, this might as well be mirroring a learning-by-doing effect. However, there is a clear indication of the model effect. Some groups hardly used any comments (if at all) in the no scaffold session (for more details, see appendix 2). All but one of these groups, however, did communicate lively via chat in this session. This indicates that they had not discovered the advantage of using the two channels for different communication purposes. After watching the model, all groups engaged in discussions via comment *and* chat.

Results of the Quantitative Analysis

To conclude, the quantitative analysis led to the following insights:

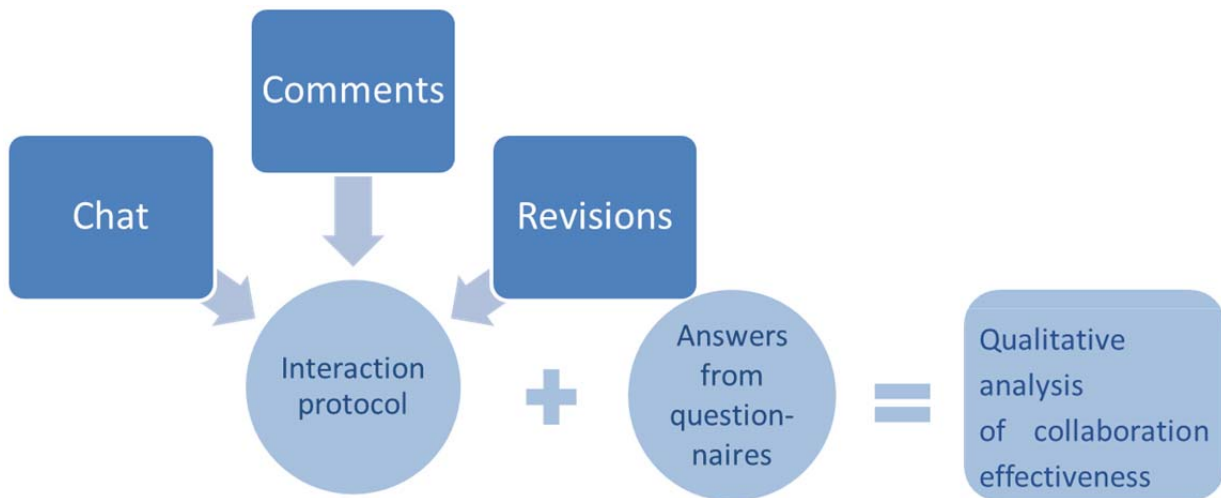
1. Scaffolding had a clear bearing on effective work flow organization. This was more prominent in the low-proficiency and mixed-proficiency triads than in the high-proficiency ones;
2. Collaboration density increased considerably after watching the model video, irrespective of triad composition according to proficiency;
3. Collaboration density for both LOC and HOC issues increased to a similar extent throughout the sessions;
4. The number of content discussions increased throughout the sessions, indicating a growing awareness of the importance of content elaboration for synthesis writing.

Using quantitative analysis, it was possible to detect significant tendencies in the development of online collaborative writing processes thanks to the impact of scaffolding mechanisms. However, there are some open questions. How exactly do script and video model benefit the work flow process? Does increasing collaboration intensity translate to effectiveness, that is, a deeper elaboration leading to a better content selection and/or more instances of peer revision, leading to a higher text quality? To answer these questions, we rely on qualitative data which will be provided in the following section of this chapter.

“You use the *Konjunktiv* here out of the blue”: Qualitative insight into collaboration effectiveness

For a detailed qualitative analysis of group interactions, three triads, each representing one of the composition types (one low-proficiency, one high-proficiency, and one mixed-proficiency triad), were randomly selected. All interactions contributing to the text genesis including chat, comments, and revisions, were transcribed and connected in an interaction protocol (see figure 6 for an overview of the data used in the qualitative analysis).

Figure 6
Data sources used in the qualitative analysis of collaboration effectiveness



As a theoretical background for the analysis of the group interaction, the framework of Meier *et al.* (Meier, Spada, & Rummel 2007) was found to be helpful. The authors established a rating scheme for a quality analysis of online collaboration consisting of nine dimensions, which they bundled into five basic aspects: good online collaboration requires good communication and coordination management, strategies for joint information processing, and the interpersonal skills needed to build a (working) relationship and maintain task motivation high. In the present analysis, Meier *et al.*'s dimensions (Meier *et al.*, p. 68) were used as a background for the selection of critical episodes that provide evidence for the effect of instructional support. These episodes relate to the following aspects of collaboration: (1) efficient task coordination and (2) interaction effectiveness, translating to (2a) deep content elaboration, and (2b) text improvement through successful peer-induced revisions. In addition, data from the questionnaires administered in the course of the intervention were used to triangulate the interaction protocol data where appropriate.

In the discussion of the selected episodes, the aspects and dimensions of the Meier *et al.* framework are italicised. All discussed episodes are listed in appendix 3. Quotes from the questionnaires are presented in the running text, directly translated into English.

Episodes of Efficient Task Coordination

Two sequences of critical episodes were selected to demonstrate how *coordination* induced by the model and the script leads to better *task division* and *time management*.

Episode 1 (see appendix 3.1) originates from the high-proficiency group interaction in the no-scaffold and in the model session. It shows that, while in the former, good individual attempts of coordination were already present, the group only gains confidence in their work flow management through the video model as common

ground, to which they explicitly refer several times. This development is supported by their statements in the questionnaire administered after the model session:

I actually knew beforehand that this was the best way to do it but it was difficult to carry out until everyone had seen the model (questionnaire excerpt)

For me, the model was really necessary to collaborate efficiently. Last week we tried to find our own way which caused uncertainty and stress. The model made us feel more secure, it worked like a fallback. We saved a lot of time by following the modelled steps. (questionnaire excerpt)

In many groups, one member tacitly took on the role of time guardian. The time guardians interfered when difference in work pace threatened to thwart the task progress. The script seems to be of great help here, because the time guardians used it as backing for their reminders, as two episode sequences from two mixed-proficiency groups exemplify (see appendix 3.2). In episode 2b, one group member explicitly appealed to the *individual task orientation* of her peers:

I really would like to finish the task this time, it's not impossible but everybody has to stay focused on the task

To conclude, the episodes discussed in this section show that both script and model helped coordinate the work flow (R7), and that this was the case for all groups under analysis, including the high-proficiency one (R8). The added value of the script was the time indications that worked as a fall-back.

Episodes of Interaction Effectiveness 1: Deep Content Elaboration

Episodes 3 and 4 (see appendices 4.3 and 4.4) illustrate how *joint information processing* following the model video leads to deep content elaboration in different groups and at different stages of the collaborative process.

Episode 3 relates to *grounding processes to sustain mutual understanding* in a high-proficiency group. The participants discuss the individual summaries they have just pasted in the common Google Doc to make sure everyone gets the gist of each other's source texts and to prepare the content selection. An interesting aside of this episode is the evidence of a constructive *interpersonal relationship*. Two group members try to help and encourage the third, who seems overwhelmed by the sheer volume of information.

Episode 4 illustrates the joint effort to create an outline for the synthesis. After having reached consensus on a rough structure, the students discuss a possible conclusion. In doing so, they *interact as equals* in the solution process thanks to their high *individual task orientation*. Inappropriate ideas are discarded in a non-face-threatening way, and a *consensus is reached* by combining several individual suggestions.

The two selected episodes in this section demonstrate that good collaboration copying the model leads to deep content elaboration by *joint information processing*. The group members *interact as equals*, contributing to the consensus by giving constructive critique and by suggesting solutions.

Episodes of Interaction Effectiveness 2: Successful Revision Sequences

To gain an overview of the revisions carried out in the three sessions by the three selected groups, some quantitative results are listed below. The revisions under investigation are those made in the collaborative text in phases 4 and 5 of the task process.

Table 2
Revision analysis for three selected groups

Aspect	Revisions measured	H2 (low-proficiency)	H4 (high-proficiency)	H11 (mixed-proficiency)
Intensity	Total number	11	18	18
Origination	Peer-induced	7 (64%)	14 (78%)	7 (39%)
Focus	HOC-related	5 (45%)	10 (56%)	8 (44%)
Success	Successful	8 (73%)	16 (89%)	12 (67%)

Table 2 reveals inter-group variance in some aspects that might be attributed to proficiency. With regard to origination, the highest number of peer-induced revisions was observed in the high-proficiency group. This might be attributable to their high self-efficacy beliefs in their writing skills. The participants of this group rated the corresponding statements (in the pre-hoc questionnaire³¹) on average 1 to 1.5 (out of 5 on a Likert-scale) higher than the other two groups. Also, the percentage of successful revisions is higher in the high-proficiency group than in the other two.

The difference in revision intensity is due to the fact that the low-proficiency group only reached the revision phase in the script session, whereas the other two groups revised their texts in the model and in the script session. In the no-scaffold session, no

³¹ English translation of the statements to be rated (from Dutch original): “*I believe I can write a well-structured / fluent / accurate text on a complex subject in German*”

collaborative revision took place. There are no salient differences between the three groups in terms of revision focus.

To conclude, the high-proficiency group seems to have profited most from the collaboration in terms of revision. On a cautionary note, it has to be stated that since the total number of analysed revisions is low, the tendencies observed have to be corroborated by a larger revision corpus in order to allow for generalizing statements.

In the following, two episodes are presented to illustrate the effect of *joint information processing* on text quality through successful peer-induced revision.

Episode 5 (see appendix 3.5) refers to a HOC revision concerning a quotation in a low-proficiency group. Apart from cohesion improvement, this interaction led to a clear learning effect, as one group member realized that using a conjunctive form without actually referencing the source does not make sense.

The second HOC revision episode (see appendix 3.6) originates from the end of the revision phase in a high-proficiency group. As they realized that their synthesis exceeded the 450 words limit, they engaged in a lively debate about which propositions to eliminate, concentrating on the examples. One group member tried to defend 'her' examples at first, but was finally persuaded by the convincing arguments of her peer.

The revision overview and the selected episodes provide evidence for the positive influence of the so-called 'pooling of knowledge about language' (Storch, 2005), as well as the pooling of cognitive resources on the final text quality. Indeed, factual and linguistic errors are detected, and coherence, cohesion, and lexical choices are improved by critical peer observations. Although the high-proficiency group seems to have benefited most from peer-induced revision, episodes from the other two groups show that this benefit is not necessarily dependent on proficiency.

Conclusions And Recommendations

The study reveals the usefulness of *both* scripting and observational learning as instructional support methods for online collaborative synthesis writing in higher education. The quantitative analysis supports most of the posited hypotheses regarding collaboration *intensity* and *efficiency*, and the episodes selected from the qualitative analysis provide evidence for the *effectiveness* of collaboration.

Results from the questionnaires corroborate the benefits of scaffolding for this collaborative writing task. Students rated the usefulness of both instructional support instruments high. The video model, that is, the first instrument, was rated 4.1 (on a scale of 5), and the script was rated 3.65. Students specifically mentioned the high usefulness of the video due to the immediacy of the model. Positive aspects of the script that were reported include the clear sequencing of the work steps and the time indications. The somewhat lower overall satisfaction rate for the script might be attributed to an effect of 'over-scaffolding', expanding Dillenbourg's concept of 'over-scripting' (Dillenbourg, 2002), as this was the second scaffold instrument administered

in a row. Given that both instruments have their strengths, it seems nonetheless recommendable to administer them both, if possible at the same time: Students could watch the model prior to engaging in the task, and have the script as a fall-back at hand while executing it. Drawing on Mauri and Onrubia's suggestion (this volume), another possibility would be to allow students to choose the scaffolding instrument for themselves, according to their individual preference for concrete visual aid including examples (the model) vs. abstract guidelines (the script).

What remains unanswered in the presented setting is the question whether the sequence of the two scaffold instruments played a role in their uptake by the students. This will be explored with a second data set where the sequence was switched to a 'script-first condition'. Another interesting question to be explored is: To what extent has a 'learning by doing' effect played a role, in addition to the instructional support, in the development of the collaborative process? In order to answer this question, at least two consecutive sessions without support in a control group would be required. Though interesting from a scholarly point of view, this seems to be daunting in terms of motivation. In some groups, the lack of support in session 1 led to disbelief and even bewilderment, as a chat contribution in one triad evidences: 'How are we supposed to go along with this? This is just impossible'. As the task in itself is complex, and students are usually not familiar with online collaboration, it is strongly recommended to provide them maximum support in this process.

The most noticeable, and important, effect of the training sequence was the enduring increase of attention to content throughout the intervention. Students realized that content elaboration plays a key role in synthesis writing. The following statement by a student in the post-hoc focus group interview is a case in point: 'I have learnt to focus more on content. First read the whole text and don't start to summarize the first paragraph directly. First find out what's it all about actually.' This is the stepping stone needed to move away from unreflecting verbatim copying of original source texts to elaboration and subsequent reformulation of content.

REFERENCES

- Baurmann, J. (2002). *Schreiben - Überarbeiten - Beurteilen. Ein Arbeitsbuch zur Schreibdidaktik* [Writing - Revising - Assessing. A handbook for writing pedagogy]. Seelze: Kallmeyer.
- Blin, F. & Appel, C. (2011). Computer supported collaborative writing in practice: An activity theoretical study. *CALICO Journal*, 28(2), 473-497. doi:10.11139/cj.28.2.473-497
- De la Colina, A. A. & García Mayo, M. d. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. d. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91-116). Clevedon NY Ontario: Multilingual Matters Ltd.
- Dillenbourg, P. & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl, & J.M. Haake (Eds.), *Scripting computer-supported collaborative learning. Cognitive,*

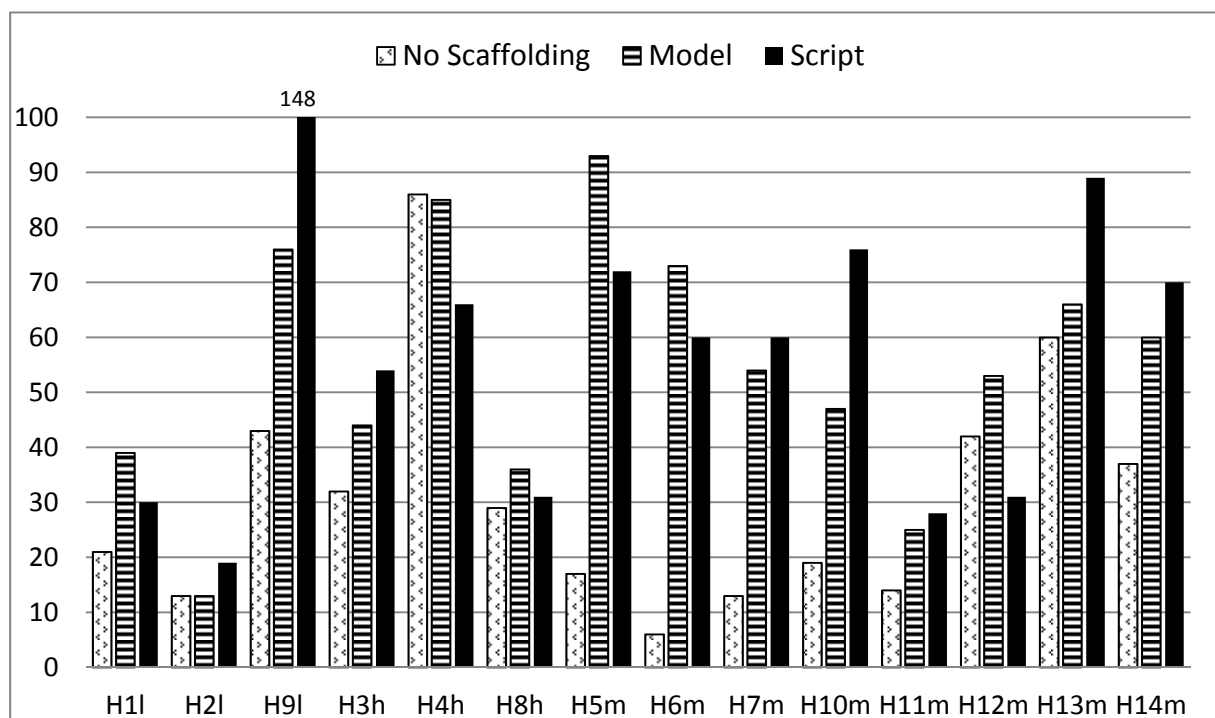
- computational and educational perspectives* (pp. 275-301). New York, NY: Springer US. doi:10.1007/978-0-387-36949-5_16
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61-92). Heerlen: Open Universiteit Nederland.
- Elola, I. & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning and Technology*, 14(3): 51-71.
- Fischer, F., Kollar, I., Mandl, H., & Haake, J.M. (Eds.) (2007). *Scripting computer-supported collaborative learning. Cognitive, computational and educational perspectives*. New York, NY: Springer US.
- Hämäläinen, R. & Arvaja, M. (2009). Scripted collaboration and group-based variations in a higher education CSCL context. *Scandinavian Journal of Educational Research*, 53(1), 1-16. doi: 10.1080/00313830802628281
- Kessler, G., Bikowski, D., & Boggs, J. (2012). Collaborative writing among second language learners in academic web-based projects. *Language Learning & Technology*, 12(1), 91-109.
- Kost, C. (2011). Investigating writing strategies and revision behavior in collaborative wiki projects. *CALICO Journal*, 28(3), 606-620. doi:10.11139/cj.28.3.606-620
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning* 2(1), 63-86. doi: 10.1007/s11412-006-9005-x
- Min, H. (2005). Training students to become successful peer reviewers. *System* 33, 293-308. doi:10.1016/j.system.2004.11.003
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265-289. doi:10.1016/s1060-3743(99)80117-9
- Raedts, M., Rijlaarsdam, G., van Waes, L., & Daems, F. (2007). Observational learning through video-based models: impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In G. Rijlaarsdam (series Ed.), P. Boscolo & S. Hidi (volume Eds.): *Studies in Writing. Volume 19. Writing and Motivation* (pp.142-165). Oxford: Elsevier.
- Rummel, N. & Spada, H. (2007): Can people learn computer-mediated collaboration by following a script? In Fischer, F., Kollar, I., Mandl, H., & J.M. Haake (Eds.) (2007), *Scripting computer-supported collaborative learning. Cognitive, computational and educational perspectives* (pp. 39-55). New York, NY: Springer US. doi:10.1007/978-0-387-36949-5_3
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4, 69-92. doi:10.1007/s11412-008-9054-4
- Schellens, T. & Valcke, M. (2000). Re-engineering conventional university education: Implications for students' learning styles. *Distance Education*, 21(2), 361-384. doi:10.1080/0158791000210210
- Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz, & N. Miller (Eds.),

Interaction in cooperative groups: The theoretical anatomy of group learning (pp. 145–173). New York, NY: Cambridge University Press.

- Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating information: An analysis of the processes involved and the products generated in a written synthesis task. *Written Communication* 30(1), 63-90. doi:10.1177/0741088312466532
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, 14(3), 153-173. doi:10.1016/j.jslw.2005.05.002
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20(4), 316-327. doi:10.1016/j.learninstruc.2009.08.009
- Würffel, N. (2008). Kooperatives Schreiben im Fremdsprachenunterricht: Potentiale des Einsatzes von *Social-Software-Anwendungen* am Beispiel kooperativer Online-Editoren [Collaborative writing in foreign language education. Affordances of *social software* using the example of collaborative online editors]. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [Online] 13(1), [http://zif.spz.tu-darmstadt.de/jg-13-1/beitrag/Wuerffel1.htm].

Appendix

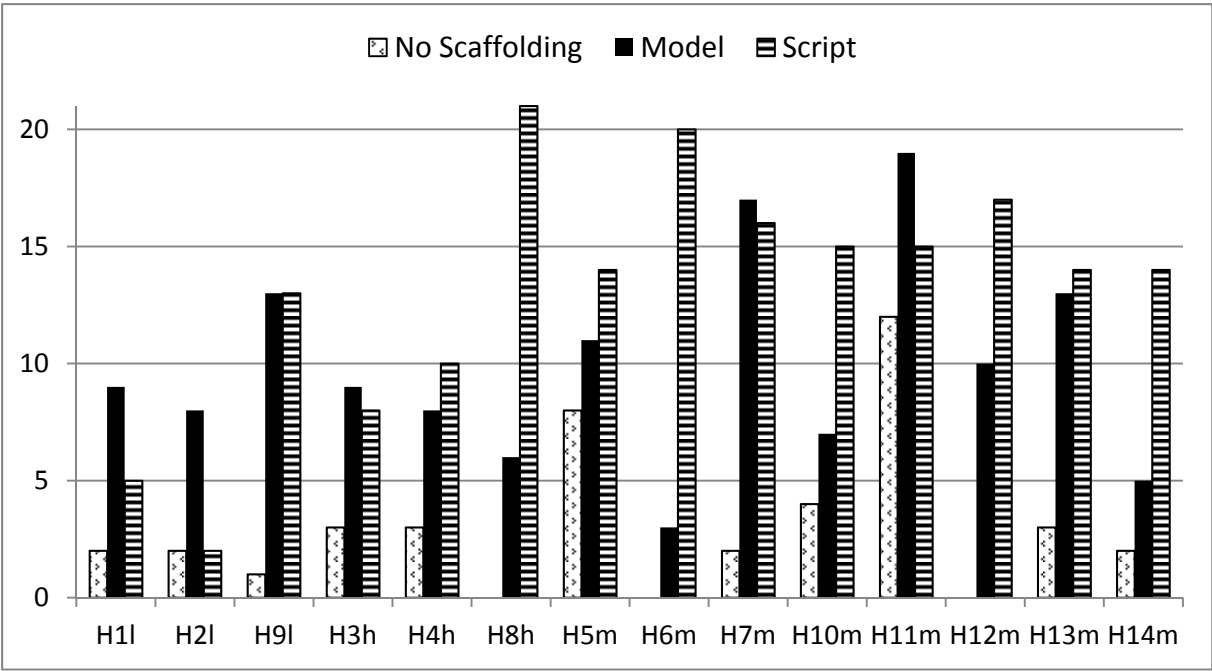
1 Chat usage per session per group (number of turns)



The chart provides insight into the difference in chat usage between the triads. The triad ID numbers are followed by a letter indicating their composition regarding

proficiency (l = low-proficiency group, h = high-proficiency group, m = mixed-proficiency group). While in some triads, chat usage remained relatively stable throughout the three sessions (H2l, H8h), in others, observing the model literally led to a communication explosion in the chat that sometimes even progressed in the script session (H9l, H10m, H13m, H14m), and sometimes was followed by a decrease (H5m, H6m). Data inspection reveals that the decrease of chat usage in the scripted session in six triads (H1l, H4h, H8h, H5m, H6m, H12m) is mainly caused by diminishing conversations about work flow.

2 Comment usage per session per group (number of original threads)



The chart provides insight into the difference in comment usage between the triads. The triad ID numbers are followed by a letter indicating their composition regarding proficiency (l = low-proficiency group, h = high-proficiency group, m = mixed-proficiency group). Some triads did not use comments (nearly) at all in the no scaffold session (H9l, H8h, H6m, H12m). All but one (H6m) of these triads, however, did communicate lively via chat in this session. This indicates that they had not discovered the advantage of using the two channels for different communication purposes. After watching the model, all triads engaged in discussions via comment *and* chat. Again, in line with chat usage, some triads still intensified the exchange of comments in the scripted session (H8a, H6m, H12m), whereas in other triads, the amount of comments remained stable (H9l, H3h, H4h, H7m, H13m) or dropped to a noticeably lower level (H1l, H2l).

3 Episodes from Interaction Protocols

The first column contains the unique reference code of the communication turn in the interaction protocol, starting either with “C” for “chat” or “K” for “comment”. The code in the second column refers to the participant’s identity, also used between brackets to anonymize the names in the conversation. The third column contains the episode in the original language (mostly Dutch³² interspersed with German and occasionally English), followed by the translation to English in the fourth column (in italics, apart from the fragments originally written in German and English).

3.1 Episode 1: High-proficiency group gaining confidence in workflow management through observational learning

Part 1a

No scaffold session, after reading the individual source texts

C12	4b	Ik heb een ideetje	<i>I've got an idea</i>
C13	4a	zeg maar :)	<i>shoot :)</i>
C14	4b	Als we nu es allemaal gewoon per alinea heel kort zeggen waarover het gaat, in een paar stichwörter, in googledocs, maar heel kort hé. En dan kijken we wat we kunnen pakken van elkaar en wat we aan elkaar kunnen linken en wat we kunnen weglaten... wat denken jullie? of is er een andere en rappere manier?	<i>If we now just all of us would say what's the gist of each paragraph, in a few stichwörter [keywords], in googledocs, but really short alright. And then we'll see what we can take from each other's sources and what can be linked and what can be left out... what do you think? or is there any other or faster way to do it?</i>
C15	4a	ja is goed	<i>yeah seems ok</i>

Part 1b

Model session, several episodes throughout phase 2

C01	4b	Hey! 'k Stel voor dat we het doen zoals in de video?	<i>Hey! I suggest we do it like in the video?</i>
-----	----	--	---

³² The linguistic standard variety quite often is blended with – or even replaced by – Flemish youth sociolect typically used in online communication, displaying a lot of dialectal influences, characteristic abbreviations and emoticons

C02	4c	Ja best zeker ;)	<i>Yep seems the best ;)</i>
C03	4a	Ja natuurlijk :) Dus, waarover gaat jullie tekst? (...)	<i>Yes of course :) So, what's your texts about? (...)</i>
C13	4b	ok; dus nu beschrijft iedereen z'n puntjes tamelijk uitgebreid maar toch niet te ;)	<i>ok; now everyone describes her main ideas in detail but not too ;)</i>
C16	4c	Oké in Word en dan plakken hierin [in the Google Doc] (...)	<i>Okay in Word and then we paste it here [in the Google Doc] (...)</i>
C18	4b	Misschien moet ieder nu voor zich eerst eens uitdenken hoe we de tekst gaan verdelen, welke 3 punten we gaan nemen zoals in de video? En dan kunnen we gaan kijken welke punten er bij passen? (...)	<i>Perhaps everyone has to come up with an idea about how we could divide the text, what 3 main topics we will have, like in the video? And then we can see which propositions fit in? (...)</i>
C28	4b	ok dus nu alle puntjes van onze eigen tekst onder de hoofdpunten slepen? Of hoe deden ze dat in de video?	<i>ok so now we drag all the items from our own text under the respective headers? Or how did they do it in the video?</i>

3.2 Episode sequences 2a and 2b: Using the script as backing for time management

2a

Low-proficiency group, starting phase 2 in script session

C22	2a	zijn jullie ook bijna klaar? Volgens het schema zijn we al 7 min over tijd :p	<i>are you nearly done as well? According to the script we're running 7 min late :p</i>
-----	----	---	---

2b

Several reminders by a time guardian of a mixed-proficiency group growing desperate in the course of the script session

C32	12b	[12c], hoe lang nog? :) (...)	<i>[12c], how long to go? :) (...)</i>
-----	-----	--------------------------------------	---

C35	12b	Je bent al bijna een kwartier over de tijd [12c]...	<i>you are running nearly 15 min late [12c]...</i>
		(...)	(...)
C39	12b	Timemanagement, [12c]... :p	<i>time management, [12c]... :p</i>
		(...)	(...)
C44	12b	Om 14u25, binnen een minuut dus, zouden we al moeten het skelett verdeeld hebben en schrijven aan ons stukje ik wil het echt eens proberen klaar hebben, 't is niet onmogelijk, alleen moet er doorgewerkt worden	<i>at 14.25, i.e. in one minute, we should have divided the skelett [enriched outline] and start writing on our individual fragments. I really would like to finish the task this time, it's not impossible but everybody has to stay focused on the task</i>

3.3 Episode 3: Joint information processing in a high-proficiency group in order to get the gist of source texts³³

C01	4a	Hey :) In mijn tekst staan veel voorbeelden uit verschillende thema's (bv. groente), maar volgens mij zijn die niet zo belangrijk en kunnen die weggelaten worden. Het is zo bij alinea 1, 2, 3, 5, en 6. Heb ze soms niet in mijn samenvatting vermeld, maar bij vb. 1 kon ik niet anders want behalve voorbeelden staat er niets anders in	<i>Hey :) My text contains a lot of examples from different topics (e.g. vegetables), but I don't think they are important and we can just leave them out. This is the case for paragraphs 1, 2, 3, 5, and 6. I sometimes did not mention them in my summary, but couldn't avoid it in 1 because it consists of only examples</i>
C02	4b	Ok dus [4a]: jouw tekst gaat er (heel algemeen) over dat het Oostenrijks Duits verdwijnt? Mijn tekst gaat erover dat Oostenrijkse woorden wel	<i>ok so [4a]: your text is (mainly) about the disappearance of the Austrian German language? My text says that Austrian terms were conserved and did</i>

³³ The overarching topic of the three source texts is the development of the Austrian German language variety in the light of globalization, more specifically, the political integration in the EU and the influence of German media.

- behouden worden en niet moesten veranderen toen ze bij de EU kwamen (dus in tegenstelling tot jouw tekst) en dat het Oostenrijks geen aparte taal is maar gewoon een variant, een deel van een Sprachraum maar dat het wel zijn eigen identiteit heeft
- C03 4a Ja. Er zijn veel voorbeelden en er komen enkele “wetenschappers” aan het woord, vooral die Wiesinger is belangrijk (komt twee keer voor) en ook Sedlaczek die dat boek geschreven heeft
- C04 4c Sorry, maar ik begrijp echt niet veel van mijn tekst...
- C05 4a Ja bij mij staat er ook dat het een Variant is, maar door de invloed van de media en boeken (die allemaal in het Hochdeutsch vertaald worden) verdwijnen de typisch Oostenrijkse woorden
- C06 4c Bij mij gaat het er dus over dat het Österreichische Deutsch verdrängt wird en dat sommige daar niets tegen hebben en andere wel
- C07 4c dat men niet alleen moet blijven bij zijn eigen taal en dus volledig gesloten moet zijn, maar dat men ook moet open staan om andere talen te leren
- C08 4a ja je interpreteert het goed hoor [4c]!
- C09 4c Maar er zijn precies allemaal verschillende onderwerpen en ik zie geen samenhang
- C10 4b Zoals ik kan zijn uit jouw samenvatting gaat het erover dat ze het Oostenrijks Duits wat willen
- not change with the integration into the EU (so this opposes what’s said in your text) and that Austrian German is not a language in its own right but just a variety, a part of a Sprachraum [language area] but that it has its own identity*
- Yes. Lots of examples, and some quotes of “researchers”, especially that Wiesinger seems to be important (is mentioned twice) and also Sedlaczek who wrote that book*
- sorry, but I really don’t get the gist of my text...*
- my text also says that it’s a variety, but the typical Austrian terms disappear due to the influence of the media and books (that are all translated into Hochdeutsch [standard German])*
- my text says that Österreichische Deutsch verdrängt wird [Austrian German is being replaced], and that some don’t mind it but others do*
- that you shouldn’t stick to your own language and seclude yourself, but you should be open to learn other languages*
- yes your interpretation is good, [4c]!*
- but they are all different topics, and I don’t see how they are related*
- what I can get from your summary is that they want to somehow preserve Austrian German because it is also an*

beschermen omdat het ook een identiteit , een stuk “eigen” is, maar dat je ook moet openstaan voor nieuwe dingen en dat je Oostenrijks niet als EEN taal mag zien, maar je ook moet openstaan voor andere dingen

identity, something “of their own”, but at the same time one should be open to new things as well, and don’t regard Austrian German as ONE language, but you have to be open to other things

C11 4c ja idd

yes that’s right

3.4 Episode 4: Joint information processing in a mixed-proficiency group to create a synthesis outline

C41 11c ok :)

ok :)

dus we hebben 1. inleiding, 2. waar het OD goed voor is en wat er gedaan is om het OD te beschermen en 3. waardoor het OD verdwijnt?

so we have 1. introduction, 2. what Austrian German is good for and what has been done to preserve it and 3. what are the reasons for it disappearing?

C42 11b fine by me

fine by me

C43 11a ok spreken we af wie wat uitschrijft?

ok shall we distribute the parts?

C44 11c zijn er nog dingen die erbij horen en die nog niet in die 3 puntjes passen?

any items left that are important and that don’t fit under these 3 headers?

C45 11b + we moeten nog weten wat we als slot gaan zeggen

+ we still need to know what to put in the conclusion

(...)

C48 11a als slot kunnen we zeggen dat elke dialect voor elk volk belangrijk is en dat het ins ons geval dus van belang is dat od niet zomaar verdwijnt kunnen we dat zeggen?

to conclude we could say that every dialect is important for its speakers and that in our case it is therefore important that Austrian German does not just disappear can we say that?

C49 11b staat dat in je tekst?

does this come from your text?

C50 11a nee, daarom dat ik aan het twijfelen ben :p

no, that’s why I am in doubt :p

C51 11c Misschien kunnen we een zin van mij

Perhaps we could transfer one of my

daarin steken in die conclusie en weglaten uit de middentekst? Das Eigene = was die Leute sprechen, beiseitegeschieben [sic]= einen Teil der Bevölkerung [sic] verdrängen.

propositions to the conclusion, and omit it from the text body? Das Eigene = was die Leute sprechen, beiseitegeschieben = einen Teil der Bevölkerung verdrängen. [The idiosyncratic = how people speak, eliminate = crowd out a part of the population]

C52 11b kversta de zin niet :D in mijn slot staat er dat niet alle woorden al verduitsd zijn dat er ook nog woorden zijn die niet wegraken

I don't understand this sentence :D in my conclusion it says that not all words have been Germanized that there are still some words that haven't been replaced

C53 11a Zusammenfassend sei festgestellt, dass, wie schon zuvor erwähnt, Sprache Identität verleiht dat is mijn slot

Zusammenfassend sei festgestellt, dass, wie schon zuvor erwähnt, Sprache Identität verleiht [To conclude, one can state that, as mentioned before, language provides identity]

that's my conclusion

C54 11c und wenn man die Sprache verdrängt, verdrängt man auch einen Teil der Bevölkerung

und wenn man die Sprache verdrängt, verdrängt man auch einen Teil der Bevölkerung [and when you eliminate the language, you also crowd out a part of the population]

C55 11b schöne [sic]

schöne [nice one]

3.5 Episode 5: Learning to quote correctly through peer assistance in a low-proficiency group

K11 2a (waarom hier een konjunktiv?)

(why a konjunktiv here?)

K12 2c Omdat het de hele tijd Rudolf is die spreekt, die wetenschapper. Mss niet nodig hier maar kweetniet, kvind da nog leuk om te gebruiken :p

Because it's Rudolf, this researcher, who talks the whole time. Perhaps it's not really necessary here but I don't know I just like to use it :p

K13 2a Oké maar waar staat die zijn naam ergens?

Okay but did you mention his name anywhere?

K14 2c Die zeg ik nergens... hebben jullie ook wetenschappers die dat allemaal

No I didn't ... do you two also have researchers who say all that stuff?

zeggen?

- K15 2a Bij mij soms eens ene ma dat gaat dan om een onderzoek dat niet belangrijk is. Ik zou zijn naam er zeker bijzetten! Anders gebruik je die konjunktiv zomaar... *In my text sometime there is one but then it's about an investigation that doesn't matter. I really think you should mention his name! Otherwise you use the konjunktiv here out of the blue...*

3.6 Episode 6: Synchronous use of chat and comment to decide upon elimination of examples in a high-proficiency group

This episode also illustrates the synchronous use of two different channels for different communicative goals, yet related to the same topic: The chat (=turn codes starting with C) is used for workflow organization, whereas the content is discussed in the comments section (=turn codes starting with K).

- C60 4b vlug eerst es kijken hoeveel woorden we al hebben *let's first have a quick look at the word count*
- C63 4b damn 536 :s *damn 536 :s*
- C66 4b DUS: iedereen leest nu alles na en geeft een opmerking als er een fout staat, of als er iets weg kan! *SO: everyone reads through everything and comments on errors or things that can be left out!*
- K18* 4b mss kunnen we dit weglaten? ** *perhaps we can cut this one? ***
- K19 4b of samen met vorige zin *or join it with the previous sentence*
- K20 4a neen da's redelijk belangrijk, da's zo'n Pseudoanglizismus *no, this is quite important, it's a pseudo-Anglicism*
- K21 4b ahja juist ok hmmm:p *oh yes right ok hmmm:p*
- K22 4c misschien dit ook inkorten? Het zijn voorbeelden... *** *perhaps shorten also these? They are examples... ****
- K23 4b ja kan ik wegdoen *yes I can cut them*
- K24 4a Ja idd, ik zou ook alle voorbeelden weglaten *Yes indeed, I'd also leave out all examples*
- K25 4b mss deze voorbeelden ook weg? ** *perhaps these examples can also go? ***
- K26 4a Ik heb die erbij gezet omdat het volgens mij anders niet duidelijk is waarover het gaat (ik verstond het alleszins niet zonder). Maar wat *I put them in there because otherwise I think it's not clear what it is about (at least I didn't understand it without). But what do you think? Do you*

		denken jullie? Verstaan jullie het zonder voorbeelden? Of mss eentje van de twee weglaten?	<i>understand the point without the examples? Or perhaps cut one of the two?</i>
K27	4b	ja alles is natuurlijk duidelijker met voorbeelden, dat had ik in mijn tekst ook :s We zullen ze er voorlopig laten staan en kijken naar het aantal woorden en als we er nog te veel hebben, ze op het einde weg	<i>Yes of course everything is clearer with examples, I felt the same way with my text :s We shall leave them in by now and count the words again and if the're still too many in the end, we'll cut them</i>
C68*	4b	492 woorden	<i>492 words</i>
C69	4b	[4a], jouw voorbeelden gaan wegmoeten :(sorry... Maar kdenk dat dat het meeste woorden al kan schrappen	<i>[4a], your examples have to go :(sorry... but I think that this will eliminate the biggest chunk</i>

* switch of communication channel

** highlighting a fragment in [4a]'s contribution

*** highlighting a fragment in [4b]'s contribution

5.3 Assessing the quality of computer-supported collaborative writing in a foreign language (Article submitted to *Journal of Writing Research*)³⁴

Computer-supported collaborative writing (CSCW) is a complex activity. This is even more true for novice writers in a foreign language (L2). Therefore, it is important to scaffold this activity with adequate instructions. In order to qualitatively analyse the impact of such scaffolding on the writing process, we developed an assessment tool based on the assessment instruments created by Meier, Spada, and Rummel (2007) and Kahrmanis et al. (2009). In this article, we present the development of this assessment tool and its implementation in a case study. The empirical data for the development and implementation stem from a classroom-based intervention study. Students of L2 German collaborated in groups of three to produce syntheses of three source texts using Google Docs. They were supported by a script and a video model in successive collaboration sessions. The quality of collaboration in the successive sessions was rated using our scheme which incorporates eight quality dimensions. The results suggest that the use of both scaffold instruments resulted in good collaboration, yet with a different impact on the writing process. Furthermore, the high inter-rater reliability of our assessment tool indicates a high operationability based on clearly defined dimensions.

Keywords: Computer-supported Collaborative Writing, L2 Writing, Instructional Support, Assessment, Rating Scheme.

1 Introduction

1.1 Collaborative Writing in an L2

The advantages of collaborative learning can be linked to the social constructivist theory of learning developed by Vygotskiĭ & Cole (1978). Building on Vygotskiĭ & Cole (1978), Kost (2001) attributes a positive learning effect to collaboration in an L2:

Vygotsky posits that a novice's cognitive development arises in social interaction with a more capable peer (expert), by helping novices move from their actual

³⁴ Hoste, L. and Strobl, C. (submitted): Assessing the quality of computer-supported collaborative writing in a foreign language. *Journal of Writing Research* (XX), p. XX-XX

level of development (i.e., what they can do by themselves) to their potential level of development (i.e., what they can do with assistance) which is also called scaffolding. Similarly, a more knowledgeable L2 learner can help out a less knowledgeable L2 learner when engaging in classroom tasks (Kost, 2011, p. 606).

In an L2 writing task, this means that students pool not only their cognitive resources to elaborate on content, but also their linguistic resources to formulate it (Storch, 2011). Previous research has pointed out multiple advantages of peer collaboration for L2 writing in instructional contexts. Storch (2013) states that collaborative writing fosters the sense of shared text ownership and, consequently, the sense of shared responsibility for it. Together, students can call on each other's help and knowledge of e.g. orthography, grammar, lexico-semantics, text type and/or culture, which considerably reduces the complexity of the writing process (Würffel, 2008). The co-authorship also lessens writers' anxiety regarding the quality of the outcome and fosters reader orientation, because students write for a specific audience (Würffel, 2008). They also co-construct knowledge and engage in collaborative critical thinking (CCT) which Ghodrati (2013) defines as follows: "The overt and tacit interaction between two or more individuals which involves collectively questioning, analyzing, synthesizing, evaluating and making decisions in order to build the collective knowledge of the group and the knowledge of the individuals in the group" (Ghodrati, 2013, p. 154). Research also suggests that working in pairs or groups affords students the opportunity to give immediate feedback on language and to pool ideas (Storch, 2005), which led in the study of Storch (2005) to shorter collaboratively written texts, but better texts in terms of task fulfilment, grammatical accuracy and complexity. Collaborative peer revision activities have also been found to be effective, since students can mutually scaffold each other when revising the text (De Guerrero & Villamil, 2000; Hanjani & Li, 2014).

However, the complexity of a collaborative writing task needs to be taken into account and catered for in instructional settings. Students encounter difficulties when writing together, on both content level and language level (Hanjani & Li, 2014; Storch, 2005, 2011; Strobl, 2015; Wigglesworth & Storch, 2012). On a language level, Van Steendam, Rijlaarsdam, Sercu, and Van den Bergh (2010) argue that, without training or scaffolding, L2 students tend to concentrate mainly on word and sentence level during revision. These superficial revisions are related to lower-order concerns (LOCs), whereas higher-order concerns (HOCs) include text structure, development of ideas, coherence, etc. (Keh, 1990). A similar tendency was revealed by the study of Kost (2011). The author investigated collaborative writing strategies and revision behaviour of L2 learners when they use a wiki for an essay assignment. 89% of the corrections done by fourth- and sixth-semester German students in her study concerned formal (surface) changes against 11% of meaning-preserving (stylistic) changes (Kost, 2011). Also on a content level, collaborative writing is challenging. Therefore, previous research states that instructional support is needed to improve various aspects of peer collaboration

(Rummel & Spada, 2005; Rummel, Spada, & Hauser, 2009; Strobl, 2015; Van Steendam et al., 2010). Previous research in CSCL has emphasized the positive effect of scripting (following instructions of a script) and observational learning (watching a video model). By observing others, students can copy successful strategies. Raedts, Daems, Van Waes, and Rijlaarsdam (2009) investigated the effect of learning-by-observing against learning-by-doing. In their study, 144 university freshmen were at random assigned to the experimental learning-by-observing condition or to the learning-by-doing control group. The participants had to complete a complex academic writing task, i.e. synthesizing research results of existing studies. The researchers found that learning-by-observing led to better introductions and conclusions, more thematically-organized content and broader task knowledge (Raedts et al., 2009). Rummel et al. (2009) also examined the effect of observational learning. The authors conducted an experimental study involving 40 dyads, of which each dyad was composed of one psychology student and one medical student. The dyads collaborated via a desktop-videoconferencing system to develop a diagnosis for a patient case. Five conditions were implemented: model-plus, model, script-plus, script and control condition (plus meaning with elaboration support). Their results revealed that learning from observing a model is a powerful pedagogic instrument for a computer-mediated setting. The model especially fostered the planning and coordination of the collaboration. Regarding the advantages of scripting, Rummel et al. (2009) pointed out that explicit instructions foster collaboration, in general, and help structure the different collaborative steps. However, Rummel et al. (2009) question the beneficial effects of scripts, because the authors did not find consistent results regarding its effectiveness. In fact, they caution for a possible discouraging effect of scripts on students' motivation. Strobl's (2015) study revealed that both scripting and observational learning are useful scaffold instruments for online collaborative synthesis writing in higher education. She found that the scaffold methods had a positive effect on collaboration efficiency, that is, a better work flow, and on collaboration intensity, which means that students discussed more content issues and that joint information processing was intensified through scaffolding.

The collaborative writing process is influenced by many factors that should be considered wisely. Würffel (2008) distinguished three main categories that influence the collaborative writing process: individual learner characteristics, group characteristics and the learning context (Würffel, 2008). The learning context includes artefacts used to facilitate collaboration. Computer-mediated tools, collaborative Web 2.0 platforms and online editors have been largely implemented in CSCW (e.g. Kost, 2011) and in CSCL (e.g. Rummel et al., 2009) settings. Chu and Kennedy (2011) analyzed the perception of Google Docs and MediaWiki as an online collaboration tool. Twenty-two undergraduate students at the University of Hong Kong reported on their use of the tools in questionnaires and interviews. In general, the results of Chu and Kennedy (2011) study pointed out that both tools were effective for the students to co-construct text and for

the teachers to monitor the students' work progress and to give them feedback. Chu and Kennedy (2011) emphasize in their study that Google Docs was welcomed by their students who highlighted the user-friendliness of Google Docs. Regarding group characteristics, it is important to take group size and group composition into consideration. Watanabe and Swain (2007) investigated the effect of proficiency differences by analyzing collaborative dialogue of 12 Japanese ESL learners with different L2 proficiency levels. The authors found that the higher the overall proficiency of the pair, the higher the production of language-related episodes (LREs). Furthermore, they found that patterns of interaction seemed to have a greater influence on the frequency of LRE than proficiency differences (Watanabe & Swain, 2007). Storch (2011) reviews different studies about collaborative writing conducted in L2 contexts and posits that low-proficiency L2 learners may not gain from collaborative tasks, unless they are paired up with high-proficiency learners.

1.2 Assessment of Collaborative Learning Processes

An important contribution towards assessment methods in CSCL research is the rating scheme developed by Meier et al. (2007). The tool was created to investigate the effect of evaluation support in addition to instructional measures for computer-supported interdisciplinary problem-solving. The authors combined a theory-based approach (top-down) with a data-driven qualitative analysis of transcribed collaboration dialogue (bottom-up) in order to develop the assessment instrument. The authors identified five broad aspects of good collaboration: 1. communication, 2. joint information processing, 3. coordination, 4. interpersonal relationship and 5. motivation. A further refinement of those categories resulted in seven quality dimensions that cover the collaboration aspects: 1. sustaining mutual understanding, 2. dialogue management, 3. information pooling, 4. reaching consensus, 5. task division, 6. time management, 7. technical coordination, 8. reciprocal interaction, 9. individual task orientation. The result is a rating scheme that enables researchers to assess the quality of collaboration processes. Then, they evaluated the scheme by applying it to a data set of 40 dyads. Each dyad consisted of one medical and one psychology student who, collaboratively, had to solve a patient case via a desktop-videoconferencing system. The dyads were grouped in one of the five conditions: a model, a model plus, a script, a script plus and a control condition. The plus conditions consisted of elaboration support in addition to the scaffold instrument. Rummel et al. (2009) report on the effects the model and the script had on the quality dimensions. Meier et al. (2007) claim that the scheme can be applied to video-recordings of other computer-supported, synchronous and interdisciplinary collaboration processes and suggest to tailor the rating instructions as well as the rating scales to the specific collaborative setting. Indeed, this original rating scheme of Meier et al. (2007) was adapted by Kahrmanis et al. (2009) to suit a very different CSCL setting: chat-based problem solving in computer-science classes. The rating dimensions were

redefined and are the following: 1. sustaining mutual understanding, 2. collaboration flow, 3. knowledge exchange, 4. argumentation, 5. structuring the problem solving process, 6. cooperative orientation and 7. individual task orientation. Their participants had to work in pairs through Synergo to build together a flow-chart of an algorithm. For the qualitative analysis, Kahrmanis et al. (2009) used the software ActivityLens that enabled them to select relevant parts of log data, audio recordings and screen captures.

2 Development of an assessment tool for collaborative writing

2.1 Adaptation process

In order to analyse the effect of scaffold instruments on the collaboration quality, we developed a rating scheme based upon a previous scholarship in L2 collaborative writing and the existing assessment tool of Meier et al. (2007) and its adaptation by Kahrmanis et al. (2009). First, we tracked the main differences between the settings and adjusted this rating scheme for CSCL to the specific needs of CSCW in an L2. Then, we applied the rating scheme first to a data set of three triads and fine-tuned it to another data set of three triads. Below, we set out the main differences between the settings and the resulting adaptations to the process dimensions.

Table 1. Main differences between settings of Meier et al. (2007), Kahrmanis et al. (2009) and this study

	Meier et al. (2007)	Kahrmanis et al. (2009)	This study
Domain	Medical decision making (diagnosing patients)	Computer programming (implementing algorithms)	Advanced L2 writing (producing syntheses)
Setting and communication medium	CSCL Desktop-videoconferencing system with shared text editor	CSCL Synergo: shared whiteboard and chat tool in one window	CSCW Google Docs: shared text document, chat tool and comment function in one window
Necessary knowledge resources	Knowledge-intensive task, collaborators with complementary knowledge	Task requires only basic knowledge; collaborators do not differ systematically	Task requires L2 writing skills, collaborators with comparable L2

	in their prior knowledge	proficiency, but complementary knowledge about content
--	--------------------------	--

Based on Rummel, Deiglmayr, Spada, Kahrmanis, and Avouris (2011, p. 376) with addition of the fourth column 'This study'.

Table 1 contains an overview of the main differences between the studies of Meier et al. (2007), Kahrmanis et al. (2009) and this study. Table 2 is identical to Table 17.1 of Rummel et al. (2011, p. 376), except that we added a fourth column with the main features of this study to highlight the differences. A first difference can be found in the collaborative setting. The studies of Meier et al. (2007) and Kahrmanis et al. (2009) investigate problem solving activities in a CSCL setting, more specifically in the domain of interdisciplinary decision-making between medical and psychology students (Meier et al., 2007) and computer programming by students of computer science (Kahrmanis et al., 2009), respectively. This study, on the other hand, investigates CSCW for L2 learning. Therefore, theoretical considerations about L2 writing and collaboration were taken into due account for the redefinition of the quality dimensions. The dimension *structuring the problem solving process* of Kahrmanis et al. (2009) was thus redefined into *structuring the writing process*.

Furthermore, this study overlaps more with the study of Meier et al. (2007) regarding the necessary knowledge resources. In our study, students have to complete a highly structured task, where they were assigned to write one synthesis together from three different source texts, based on the *jigsaw* concept. This means that each student was responsible for summarizing its main ideas on behalf of the peers, which increased their extrinsic motivation to elaborate on content. In the study of Meier et al. (2007), students also held complementary knowledge: a student of psychology and a medical student had to exchange expertise-specific information in order to diagnose a patient. Meier et al. (2007) covered this aspect by the dimension *information pooling*. For our rating scheme, given the fact that students did not only have to pool information from the source texts, but also linguistic resources about the L2, we divided this dimension into two: *language-related information pooling* and *content-related information pooling*.

A third important difference between the studies is the communication medium used for collaboration. In this respect, this study bears more of a resemblance to the one by Kahrmanis et al. (2009). Meier et al. (2007) used a desktop-videoconferencing system, where students could hear and see each other during collaboration. Therefore, the authors created a dimension called *dialog management* to coordinate the turn-taking. In the study of Kahrmanis et al. (2009), however, Synergo was used by the students, where collaboration occurred through chat messages and a shared whiteboard. This difference in communication style required Kahrmanis et al. (2009) to change the dimension *dialog management* (Meier et al., 2007) to *collaboration flow*, which refers to “a coherent

sequence of messages, both verbally and conveyed through actions” (Rummel et al., 2011, p. 377). In our study, students communicated using the online interaction facilities of Google Docs for their writing assignment. We restricted the dimension *collaboration flow* (Kahrimanis et al., 2009) to the more narrow dimension *communication flow*, because our sixth dimension *structuring the writing process* already covers the aspect of collaboration flow and workflow organization. Thus, our second dimension *communication flow* is restricted to the rating of the linguistic interactions and the use of communication channels available.

2.2 Rating procedure

For the in-depth qualitative analysis of the collaborative writing processes, the screens of the students were video-recorded using the screen-capturing software Morae. All interactions, including chat messages and comments, and actions in the joint text document, i.e. revisions, deletions, and insertions, were transcribed in interaction protocols and served as a fall-back during the rating procedure. Then, we applied the rating scheme to all (inter)actions of the collaborative writing processes and attributed a score on a 5-point scale for each quality dimension in each collaboration session. The eighth dimension, *individual task orientation*, was assessed for each student individually.

To assess the robustness of our scales, two coders independently rated the collaboration processes in three sessions of three triads (i.e. nine collaboration sessions). Inter-rater reliability was calculated using a two-way mixed, consistency, average-measures intra-class correlation coefficient ICC (McGraw & Wong, 1996) to assess the degree that coders provided consistency in their ratings of collaboration quality according to the established dimensions. The resulting ICC was in the excellent range, ICC = 0.948 (Cicchetti, 1994), indicating that coders had a high degree of agreement and suggesting that collaboration quality was rated similarly across coders. Quality ratings based on the rating scheme were, therefore, deemed suitable for the purposes of this study.

2.3 Proposed rating scheme for CSCW in an L2

The end result of the adaptation process is a rating scheme, presented in Table 2, to assess collaborative writing processes in an L2. In Appendix A, a more detailed description of two dimensions, *communication flow* and *language-related information pooling* can be found, with the five score levels and two examples to illustrate each dimension from the interaction protocols. We selected these two dimensions, because we wanted to illustrate how the workflow progress is rated (*communication flow*) and how we rate the aspect of collaborative writing in an L2 (*language-related information pooling*). The eight quality dimensions we identified as important aspects of CSCW in L2 are: 1. sustaining mutual understanding, 2. communication flow, 3. content-related

information pooling, 4. language-related information pooling, 5. argumentation, 6. structuring the writing process, 7. cooperative orientation, 8. individual task orientation.

Table 2. Rating scheme for CSCW in an L2

Dimension	Writing phase¹	Inherent aspects
1. Sustaining mutual understanding	Planning	<ul style="list-style-type: none"> - Exchanging information about main ideas and/or source texts - Giving and asking for feedback to generate a common base of knowledge - [Providing transparent, well-structured summaries of source texts, elucidating difficult concepts and ideas²]
2. Communication flow	All	<ul style="list-style-type: none"> - Noticing peers' actions - Reacting to peers' queries and suggestions without undue delay - Reporting on one's own progress in individual work phases - Making appropriate use of communication channels
3. Content-related information pooling	Planning and Translating	<ul style="list-style-type: none"> - Information pooling [based on source texts] for content elaboration - Selecting and organizing content for the joint text
4. Language-related information pooling	Translating and Revising	<ul style="list-style-type: none"> - Number of peer-induced revisions - Range in focus of revisions (LOC, HOC) - Success of collaboratively solved revisions
5. Argumentation	All	<ul style="list-style-type: none"> - Exchanging arguments during decision-making processes (i.e. proposing or defending an idea regarding the outline or the content of the text) - Quality of argumentation, including reference to external sources
6. Structuring the writing process	All	<ul style="list-style-type: none"> - Good time management - Adapting a stepwise procedure (if applicable, following a script and/or model, and respecting the word limit) - Distributing tasks and work load to avoid overlap - Technical coordination (incl. trouble shooting)
7. Cooperative orientation	All	<ul style="list-style-type: none"> - Creating a good working atmosphere, based on mutual respect and willingness to help - Cooperative tone of conversation

		<ul style="list-style-type: none"> - Using emoticons and pep-talk to create team spirit - Time spent in collaborative work phases (in relation to individual work phases)
8. Individual task orientation	All	<ul style="list-style-type: none"> - Showing commitment to, and motivation for, task and joint product - Number of individual contributions in different communication channels (e.g. chat, comments, revisions)

The eight quality dimensions can be found below.

Dimension 1: Sustaining mutual understanding

An essential aspect for successful peer collaboration is to *sustain mutual understanding*, i.e. create a common base of knowledge before the actual writing takes place. These grounding processes build a necessary basis for the students to be on the same page. When using a *jigsaw* concept for the writing assignment, students should inform each other about their own text piece in a transparent way and provide clear explanations. The students ensure everybody understands the main goal of the writing task and the direction the text is headed in future. To do so, students ask, and give, each other for feedback (e.g. "do you understand?" and "I understand"). Possible misunderstandings are avoided by clarifying difficult concepts or by translating them into their mother tongues.

Dimension 2: Communication flow

Students maintain a steady communication flow by using the interaction facilities of the online collaboration tool appropriately. They notice each other's actions, react to each other's queries and suggestions without undue delay. When the communication flow is interrupted or neglected, the group members try to bring it back to life. They should also report on their individual progress, e.g. by sharing how long the individual writing will take, and ask each other questions about their work progress to avoid chaos and duplication. Questions about content or language concerns should be answered.

Dimension 3: Content-related information pooling

When writing a text collaboratively, students should pool their content knowledge in order to create the joint text. When using a *jigsaw* concept, each co-author should pool the details given in his/her source text. Together, they should elaborate on the content and draft a text outline. Every group member should add the content-related information given by the other students to their own in order to create the final text. They search for similarities and differences in the available content.

Dimension 4: Language-related information pooling

When writing in an L2, beside information related to the writing subject, students also should pool their linguistic resources about L2 writing. Students can help and support each other during the whole writing process, but especially during the revision phase, for various linguistic topics, on word- and sentence-level (orthographic, syntax, lexicon, morphology), as well on textual level (content, coherence and cohesion).

Dimension 5: Argumentation

When students write a text together, they also have to make decisions together. This dimension rates the quality of the arguments exchanged for or against these decisions regarding the writing product and process. These may concern a certain wording, addition, deletion, interpretation, the text's structure, etc. To prove their arguments, students may refer to external sources to substantiate their case.

Dimension 6: Structuring the writing process

Collaborative writing is a complex, highly structured process which can be doomed without coordination. This dimension measures the degree of organization and structure during the writing process. Students have to manage the time, divide the tasks, handle technical problems, respect the word limit, and, if applicable, follow certain guidelines offered by instructional guidance.

Dimension 7: Cooperative orientation

Collaborative writing, logically, requires a cooperative attitude of all group members. This dimension measures the (un)cooperative atmosphere during the whole collaborative writing process. Students should help each other, treat each other as equals, show mutual respect and avoid working on their own islands too much. The use of emoticons or pep-talk contributes to the team spirit, and, thus, to the collaboration.

Dimension 8: Individual task orientation

The individual task orientation measures the degree of commitment to, and motivation for, the task and joint product. This dimension is rated for each student individually.

3 Implementation of the assessment tool in a case study

3.1 *Background information about the case study*

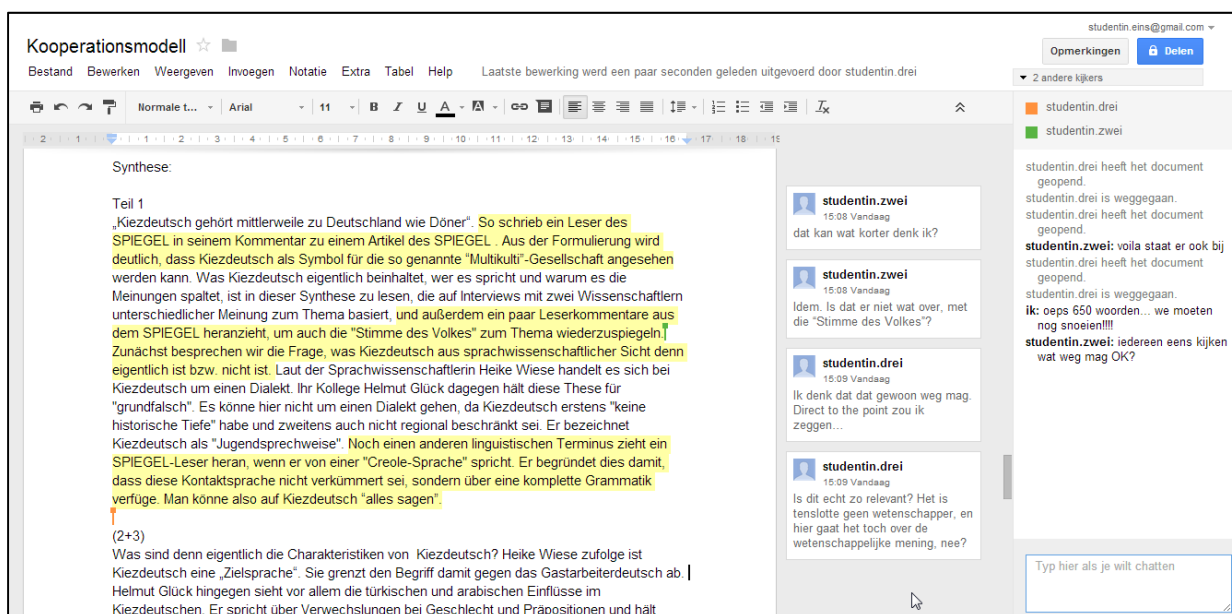
In order to implement our assessment tool in a case study, data was used from a classroom-based intervention study on collaborative writing in German L2. This study investigated the effect of scaffolding on the quality of online collaborative synthesis writing, using the online editor Google Docs for writing and communication

(synchronous CMC). Third-year bachelor students (n=76) of two study programmes in higher education institutions ("Linguistics and Literature" at Ghent University and "Applied Language Studies" at Ghent University College) were grouped in triads to produce collaborative syntheses of three source texts. All participants were Dutch native speakers and studied German as an L2. They had an advanced proficiency level of German (B2-C1 of the Common European Framework³). To investigate the impact of group composition on task performance, low-proficiency, high-proficiency, and mixed-proficiency groups were composed based on a pre-test score. In three consecutive weekly 90-minute sessions, the participants wrote each week in collaborative groups of three a synthesis in German based on three popular scientific texts, using the online editor Google Docs for writing and communication (synchronous CMC). In the second and third week of collaboration, they were given instructional support to structure the task and the workflow in the form of a script and a model video, pointing out and exemplifying the desirable steps and strategies to be implemented. In the collaborative planning phase, a *jigsaw* concept was applied (Dillenbourg & Jermann, 2007), in that each student only read one source text and was responsible for summarizing its salient points to their peers. Besides fostering positive interdependence between the peers (Johnson, Johnson, & Holubec, 1998), this information gap also served as extrinsic motivation to elaborate on the content of the source texts by negotiating meaning.

During the task assignment, the students communicated using the online interaction facilities of Google Docs: a chat space, a comment tool, and the emerging text itself. Students were recommended in script and model to use the chat for coordination and work-flow discussions, and the comments that can be linked to highlighted text sections for in-depth discussions about content and language issues. The students were also encouraged to write the collaborative text in the shared Google document and to prepare individual summaries in MS Word. Google Docs offers an easy tool to manage and track students' actions.

The script contained guidelines on how, where and when to plan, collaborate and coordinate in Google Docs. Students were prompted to follow six process steps, which are listed in Appendix B.

Figure 1. Screenshot of video model during step 5



The video model was shown to the students just before the actual collaboration, expecting this would lead to emulation of good practices and avoidance of pitfalls shown in the 13-minutes video. It illustrated the six process steps of the script. As Figure 1, a screenshot of the video model during step 5, exemplifies, students could observe in the model how to approach the writing task in Google Docs and how to appropriately use the three communication channels. The resulting text appears in the middle of the screenshot, and was written in German, the L2 of the students. The model demonstrated how they could structure the text in parts for each main idea. To the right of the text, students saw how they could open comments to ask specific questions on highlighted sections. The chat space is located on the far right, where students could talk about the text and the work flow. The comments and chat messages are written in their mother tongue, Dutch, to stimulate the students to formulate ideas in their own words and to facilitate the task assignment.

Table 3. Crossed conditions applied in the case study

	S1 triads (n=3) BA Language Studies & Literature	M1 triads (n=3) BA Applied Language Studies
Session 1	No scaffold	No scaffold
Session 2	Script	Video Model
Session 3	Video Model	Script

The two scaffold instruments were implemented using a crossed-conditions study design, as shown in Table 3. This design was adopted to investigate a possible impact of the order of scaffolding. In what follows, we distinguish between script-first triads (S1) and model-first triads (M1). For the case study, we randomly selected three triads of each condition according to proficiency level. As a result, our assessment tool was developed and fine-tuned on the basis of the qualitative analysis of six groups, viz. three S1 triads and three M1 triads, two of which were high-proficiency, two mixed-proficiency and two low-proficiency groups.

3.2 Results of the qualitative analysis applying the new rating scheme

The results of the ratings are presented per condition, i.e. M1 and S1 triads. Since all three sessions were executed consecutively, it is possible that the students learned how to collaborate over time and, thus, that the ratings show a learning-by-doing effect. Furthermore, the collaboration quality in the third session could still gain from the scaffold instrument presented in the second session. Thus, the results of the M1 triads can, first and foremost, afford an insight into the effect of the video model, and the results of the S1 triads into the effect of the script. The key changes that became evident over time through the rating are described in the section that follows.

3.2.1 Results of M1 triads

Table 4. Ratings for M1 triads

	Session 1: no scaffold			Session 2: video model			Session 3: script		
	Low-pr*	High-pr	Mixed-pr	Low-pr	High-pr	Mixed-pr	Low-pr	High-pr	Mixed-pr
1 Sustaining mutual understanding	3	5	2	4	5	4	3	5	5
2 Communication flow	2	3	2	4	5	4	3	5	4
3 Content-related information pooling	2	3	3	3	4	4	4	5	5
4 Language-related information pooling	1	1	1	2	3	3	3	4	4
5 Critical argumentation	2	3	4	3	4	5	4	5	5
6 Structuring the writing process	2	3	2	3	5	3	4	5	4
7 Cooperative orientation	2	4	4	4	5	4	4	5	5
8a** Individual task orientation	4	4	2	5	5	3	5	4	5
8b Individual task orientation	3	5	4	3	5	5	4	5	5
8c Individual task orientation	2	3	4	3	4	4	4	4	4

*low-pr = low-proficiency triad, high-pr = high-proficiency triad, mixed-pr = mixed-proficiency triad

** a, b, c = anonymized student code, different per triad

The ratings showed that the unscaffolded triads yielded the lowest ratings in comparison with the scaffolded sessions. Regarding the dimension *language-related information pooling*, all triads obtained a score of 1 without scaffolding, because they ran out of time to revise the synthesis and to pool their linguistic resources. This poor time management was reflected in the ratings: the low- and mixed-proficiency triads obtained a score of 2 and the high-proficiency triad a score of 3 on the dimension *structuring the writing process*, which covers the aspects of time management, task division, respect to the word limit, technical coordination and workflow organization.

In the video model session, the scores of all quality dimensions improved. One of the main quality improvements is the one of *communication flow*: in all triads, the scores increased by two points, going from a score of 2 to 4 for the low- and mixed-proficiency triad and from 3 to 5 for the high-proficiency triad. The dimension *content-related information pooling*, which captures the aspect of content elaboration and structuralization, improved with one point in all triads. The qualitative analysis revealed that students referred to the video model for the imitation of strategies shown in it. For example, students established a text skeleton (outline) with three main bullet points, as visualized in the model, or they introduced conjunctions and prepositions in their individual summary for a more structured representation. The ratings suggested also that the video model benefited the dimension *structuring the writing process*: workflow was organized more efficiently. However, time management still was not successful with the model. For example, the low-proficiency triad worked against the clock and did not allow for revision, resulting in a score of 2 on the dimension *language-related information pooling*. In the low- proficiency triad , the scores for *sustaining mutual understanding* rose by one point and in the mixed-proficiency triad by two. On this dimension, the high-proficiency triad already obtained an excellent score of 5 without scaffolding, and maintained this score throughout the sessions. The ratings also revealed a rise by one point in all triads in terms of *argumentation*. Furthermore, the *cooperative orientation* of the low-proficiency triad went up by two points, going from a score of 2 to 4 thanks to the model. Students' *individual task orientation* either increased or stagnated.

In the script session, all scores improved by one point or remained the same. However, in the low-proficiency triad, the scores for *sustaining mutual understanding* and *communication flow* dropped by one point, going from 4 to 3, in the script session.

3.2.2 Results of S1 triads

Table 5. Ratings for S1 triads

	Session 1: no scaffold			Session 2: script			Session: video model		
	Low-pr*	High-pr	Mixed-pr	Low-pr	High-pr	Mixed-pr	Low-pr	High-pr	Mixed-pr
1 Sustaining mutual understanding	3	4	4	2	4	4	3	4	4
2 Communication flow	4	2	3	4	3	5	5	4	4
3 Content-related information pooling	2	3	2	2	3	3	3	4	5
4 Language-related information pooling	1	1	1	3	4	3	3	2	1
5 Argumentation	4	3	3	3	3	4	3	4	5
6 Structuring the writing process	2	2	2	3	4	3	4	4	4
7 Cooperative orientation	4	3	4	4	4	5	4	4	4
8a** Individual task orientation	3	2	2	3	3	Absent	3	3	3
8b Individual task orientation	5	4	3	5	4	4	5	5	5
8c Individual task orientation	4	3	5	4	4	5	4	4	4

* low-pr = low-proficiency triad, high-pr = high-proficiency triad, mixed-pr = mixed-proficiency triad

** a, b, c = anonymized student code, different per triad

In the unscaffolded session, all triads obtained a score of 1 for the dimension *language-related information pooling* due to a lack of time. On *structuring the writing process*, which covers e.g. the aspect of time management, all triads were rated at 2. The ratings revealed an unexpected trend, namely a better performance by the low-proficiency triad in comparison with the high-proficiency triad on various dimensions. The results also suggested relatively high scores in all triads, even without scaffolding, for the dimensions *sustaining mutual understanding*, *argumentation* and *cooperative orientation*.

In the script session, the scores for *structuring the writing process* increased by one point for the low- and mixed-proficiency triads and by two points for the high-proficiency triad. Especially regarding its aspect of time management, students used the script as a fall-back during the collaboration and referred to the time indications within each step (see Appendix B). As a result, all triads had more time for revision and *language-related information pooling*, as was illustrated by scores going from 3 to 4 on this dimension. The qualitative analysis revealed that the focus of these revisions were predominantly LOCs. Another distinct pattern indicated by the ratings is that the scores for *content-related information pooling* remained generally stationary. Indeed, the qualitative introspection showed that the script did not stimulate content elaboration and that the students still encountered difficulties when trying to integrate their individual summaries into one synthesis. Furthermore, the ratings revealed a one-point drop, going from a score of 3 to 2, regarding the dimension *sustaining mutual understanding* in the low-proficiency triad. The *communication flow* of the mixed- and high-proficiency triads increased in the script session, by two points and one point, respectively. The scores for *argumentation* are not straight-forward: the low-proficiency triad lost one point, the mixed-proficiency triad gained one and the high-proficiency triad remained level on this dimension. The ratings also showed that the students' *cooperative orientation* and *individual task orientation* generally increased.

The ratings of the video model session generally increased or remained as they were. When we compare the script session with the consecutive video model session, an interesting pattern became apparent, namely, *content-related information pooling* improved by one point in the low- and high-proficiency triads and by two points in the mixed-proficiency triad.

3.2.3 Discussion

Figure 2. Spider charts with ratings for M1 and S1 triads

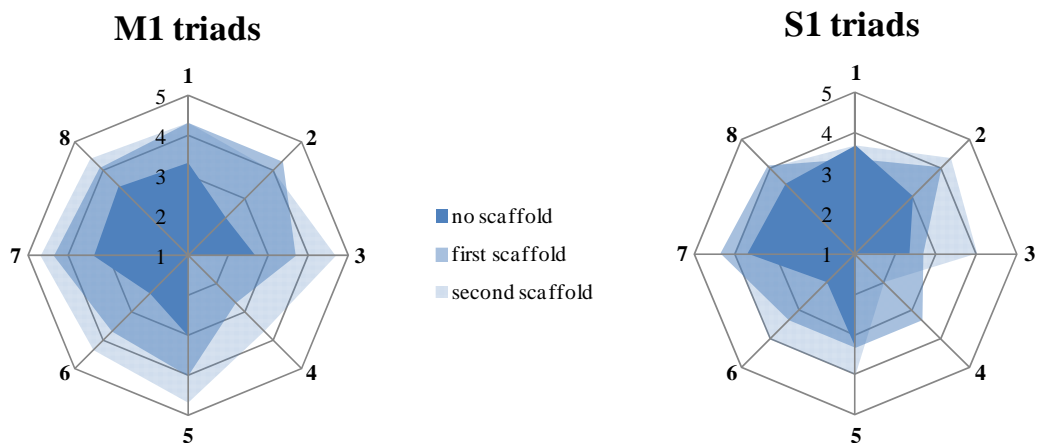


Figure 2 visualizes the development of the overall collaboration quality in the M1 and S1 triads over the course of the three consecutive sessions. The five rating scores are represented on the vertical axis, while the eight quality dimensions are introduced as the spokes of the spider chart. At first sight, it is clear from Figure 2 that collaboration quality improved over time in both conditions on multiple collaboration aspects. The dark blue area (unscaffolded session) expands with the first scaffold (session 2) and enlarges even more with the second scaffold (session 3). Figure 2 also indicates that the M1 triads experienced a much more transparent development of the collaboration quality, and, thus, gained more from scaffolding. In the following section, our most important findings are summarized and illustrated with chat sequences.

In general, in the unscaffolded sessions, the ratings were lower compared to the scaffolded sessions. This suggests that collaboration quality improves with instructional support (Van Steendam *et al.*, 2010; Rummel *et al.*, 2009; Rummel & Spada, 2005; Strobl, 2015). M1 as well S1 triads in the unscaffolded sessions made insufficient organizational agreements, had long individual phases, did not communicate well and expressed their dissatisfaction during the chat section. First of all, low scores on the dimension *structuring the writing process* were found: students did not know how to approach the complex writing task, which often resulted in chaotic moments, frustration and time pressure (*Example 1*). As a consequence, students were unable to revise, or finish, their synthesis. This lack of time resulted in low scores on the dimension *language-related information pooling*. Furthermore, untrained students nearly elaborated on the content: they copied-pasted their individual summaries and made small adjustments, but failed to integrate them into one coherent text. This was reflected by low scores on the dimension *content-related information pooling*. Low- and mixed-proficiency triads focused – without training – mainly on word and sentence level when revising the text, with coincides with the study of Van Steendam *et al.* (2010).

Example 1. Structuring the writing process in a high-proficiency S1 triad (score: 2)

Students encountered time pressure and weren't able to finish their synthesis.

C34	3b	[...] hebben wij nu nog maar 25 minuten? (...)	[...] we've only got 25 minutes left?
C38	3a	Idd we zullen eens moeten beginnen schrijven (...)	Indeed we should start writing
C49	3c	Oei oei	Oops oops
C50	3b	Nicht ganz!	Nicht ganz [not fully]
C51	3a	Das niet erg, twas te veel om helemaal af te werken in zo korte tijd	Doesn't matter, it was too much to complete in such a short time

In the scaffolded sessions, we found both scaffold instruments enhanced collaboration quality, but with another impact on the quality dimensions. The most distinct tendency identified by the ratings was that the video model stimulated *content-related information pooling* and that the script predominantly fostered time management and workflow coordination, covered by the dimension *structuring the writing process*. Students often referred to the script and used it as a fall-back for time management and word limit (*Example 2*), but it did not contribute to content elaboration. This absence of content-related integration resulted in incoherent texts, where identical ideas are repeated whereas they should have been fused, and where connections between text parts (contradiction, concession, etc.) were ignored (*Example 3*). A possible explanation for that could be the following: content elaboration in a synthesis writing task is complex and requires a high level of comprehension (Solé, Miras, Castells, Espino, Minguela, 2013), and thus script guidelines are too abstract and too general to meet the high cognitive skills required for content elaboration. A video model on the other hand departs from observational learning, where students visually see, learn and imitate strategies illustrated in the model (Van Steendam et al., 2010). The ratings revealed that, based on the model, students searched for similarities and differences in their source texts, discussed difficult concepts, decided which sections to eliminate and which to keep, and they imitated strategies shown in the model (*Example 4*). For instance, they drafted an outline with three bullet points and inserted conjunctions in the text skeleton.

Example 2. Using the script as a fall-back for time management in a high-proficiency S1 triad, script session (score: 4). The students followed the time indications and respected the word limit as instructed in the script.

C14	3b	dus om 9.40 zetten we onze	So at 9.40 we put our summaries in
-----	----	----------------------------	------------------------------------

		samenvattingen in dit document [...] 450	<i>this document [...] 450</i>
C15	3c	ich habe 133 wörter	<i>I've got 133 words</i>
C16	3b	ich 103 om 9.55 moeten we naar stap 3 dan	<i>me 103 at 9.55 we have to go to step 3 then</i>
C17	3a	125 woorden	<i>125 words</i>

Example 3. Content-related information pooling in a high-proficiency S1 triad, script session (score: 3)

Student B suggested to integrate the text parts more, but student C rejected this proposal by stating it would be better that every group member stays with his/her own text.

C27	3b	uh-kee moeten we die teksten niet wat mischen? of wordt dat te ingewikkeld?	<i>uh okay, shouldn't we mix the texts more? Or would that be too difficult?</i>
C28	3c	denk dat het beter is om elk bij zijn tekst te blijven	<i>think it's better to just stay with your own text</i>

Example 4. Content-related information pooling in a high-proficiency M1 triad, model session (score: 4)

Students referred to the video model and made great efforts to integrate the content.

C01	4b	Hey! 'k Stel voor dat we het doen zoals in de video? (...)	<i>Hey! I suggest we do it like in the video?</i>
C18	4b	Misschien moeten we ieder nu voor zich eerst eens uitdenken hoe we de tekst gaan verdelen, welke 3 punten we gaan nemen zoals in de video? En dan kunnen we gaan kijken welke punten er bij passen? (...)	<i>Perhaps everyone has to come up with an idea about how we could divide the text, what 3 main topics we will have, like in the video? And then we can see which propositions fit in?</i>
C24	4c	b bij jou gaat het er ook over dat ze die Engelse woorden willen vermeiden	<i>b in your text it's also about avoiding those English words</i>
C25	4c	en bij mij is het ook zo dat ze de Engelse taal in de Werbung beter wat meer zouden vermeiden omdat het te ingewikkeld wordt, de mensen begrijpen die Engelse woorden niet	<i>and in my text they also say they should avoid English in advertisements because it becomes too complicated, the people don't understand those English words</i>
C26	4b	de nadruk ligt bij mij niet echt op	<i>the emphasis isn't laid on the</i>

het
vermeiden maar op die actie zelf. 't
Is niet
echt dat ze er super tegen zijn ofzo

*avoidance, but rather on the action
itself. It's not like they are really dead
set against it*

4 Conclusions and Limitations

This paper describes the development of a rating scheme to assess the quality of CSCW processes in an L2 and reports on the results of its implementation in a case study. Our assessment tool was based on the rating scheme of Meier et al. (2007) and Kahrimanis et al. (2009), which were both developed to assess computer-supported problem-solving with different settings. We adapted the rating scheme to fit the needs of qualitative inspection into CSCW processes in L2 synthesis writing. The result of the adaptation process are eight quality dimensions to assess web-based peer collaboration in an L2. To test its effectiveness, we adopted the new rating scheme for a qualitative analysis of the processes of six triads who produced syntheses in German L2 in three consecutive sessions, being supported in the process by a script and a model video as scaffolding instruments. The ratings of the case study revealed that the assessment tool is reliable in terms of inter-rater agreement and sensitive to both the development of collaboration quality over time and to the effect of instructional support. Therefore, we suggest that our rating scheme can be valuable for both researchers and teachers who wish to assess the quality of L2 online collaborative writing. We recommend considering an adaptation and redefinition of the dimensions to the specific task and setting. The tool can also be used for formative assessment in order to raise students' awareness about the aspects of good collaboration. Further research is necessary to test the rating scheme in other CSCW in L2 scenarios.

The ratings enabled us to identify quality improvements in the triads' collaboration development, which can be traced back to the beneficial effect of the scaffold instruments implemented in the case study. Higher ratings in the video model session on the dimension *content-related information pooling* and qualitative introspection into their collaboration revealed a tendency towards the importance of observational learning for content elaboration. In academic writing, content elaboration plays a pivotal role. Thus, we would suggest to scaffold students for such a task using a model. The script appeared to be very useful when implementing a collaborative task that requires strict time management and that holds possible organizational pitfalls, which was covered by the dimension *structuring the writing process*.

The rating scheme has proved to be effective to detect progress in collaborative processes. The reason for these quality improvements can also, beside the effect of instructional support, be attributed to a learning-by-doing effect. Thus, it would be interesting to further analyse the effect of learning-by-doing, as opposed to learning-by-observing, by involving a control group. Furthermore, the ratings showed a different

collaboration quality development in the two conditions, i.e. M1 and S1 triads: M1 triads had higher ratings and a clearer positive development in the scaffolded sessions. A possible explanation for the higher scores for M1 triads can be found in the order of the consecutively presented scaffold instruments. Taking in *first* visually how the complex writing task can be tackled, i.e. by learning through observing, seems a necessary basis for a successful collaboration. Therefore, presenting the script *after* having seen a collaboration model might be a more promising sequence of scaffolding than vice-versa. Therefore, as recommended by Strobl (2015), we would suggest to support students with a video model before engaging in a complex collaborative writing task, and to offer the script as an optional fall-back during their actual collaboration. In the present study, next to scaffolding sequence, another variable might have had an impact on this between-groups difference. Although both groups major in German studies and are comparable with regard to their language proficiency, they have a different study background. S1 triads are students of a Linguistics and Literature programme at a University, whereas M1 triads follow a bachelor programme in Applied Language Studies at a University College. There is a difference in the instructional approach of both institutions of higher education, in that University College students are used to being scaffolded while engaged in study tasks, whereas University students are used to managing on their own. This could in part explain the higher uptake of the support provided in the scaffolding instruments by the M1 group, resulting in the more transparent collaboration quality development in the M1 triads.

Footnotes

1. Phases based on the writing model of Flower and Hayes (1981).
2. This aspects applies to the specific task of synthesis writing.
3. Correspond with ACTFL Proficiency Guidelines levels 'Advanced mid' and 'Advanced high'.

References

- Chu, S. K. W., & Kennedy, D. M. (2011). Using online collaborative tools for groups to co-construct knowledge. *Online Information Review*, 35(4), 581-597. doi: 10.1108/14684521111161945
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4), 284-290.
- De Guerrero, M., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *The Modern Language Journal*, 84(1), 51-68.
- Dillenbourg, P., & Jermann, P. (2007). Designing Integrative Scripts. In F. Fischer, I. Kollar, H. Mandl, & J. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (Vol. 6, pp. 275-301). New York, NY: Springer US. doi:10.1007/978-0-387-36949-5_16

- Ghodrati, N. (2013). Collaborative critical thinking on asynchronous discussion forums in a multicultural learning context in higher education. Retrieved from <http://hdl.handle.net/11343/39752>
- Hanjani, A. M., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System*, 44, 101-114. doi: 10.1016/j.system.2014.03.004
- Johnson, D. W., Johnson, R. T., & Holubec, E. (1998). *Cooperation in the classroom*. Boston: Allyn and Bacon.
- Kahrimanis, G., Meier, A., Chounta, I.-A., Voyiatzaki, E., Spada, H., Rummel, N., & Avouris, N. (2009). Assessing collaboration quality in synchronous CSCL problem-solving activities: Adaptation and empirical evaluation of a rating scheme. In U. Cress, V. Dimitrova, & M. Specht (Eds.), *Learning in the Synergy of Multiple Disciplines* (Vol. 5794, pp. 267-272): Springer Berlin Heidelberg. doi: 10.1007/978-3-642-04636-0_25
- Keh, C. L. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44(4), 294-304.
- Kost, C. (2011). Investigating writing strategies and revision behavior in collaborative wiki projects. *CALICO Journal*, 28(3), 606-620. doi: 10.11139/cj.28.3.606-620
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, 1(1), 30-46.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 63-86. doi: 10.1007/s11412-006-9005-x
- Raedts, M., Daems, F., Van Waes, L., & Rijlaarsdam, G. (2009). Observerend leren van peer models bij een complexe schrijftaak. [Observational learning through peer models for a complex writing task]. *Tijdschrift voor Taalbeheersing*, 31(2), 142-165. doi: 10.5117/TVT2009.2.OBSE357
- Rummel, N., Deiglmayr, A., Spada, H., Kahrimanis, G., & Avouris, N. (2011). Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (Vol. 12, pp. 367-390): Springer US. doi: 10.1007/978-1-4419-7710-6_17
- Rummel, N., & Spada, H. (2005). Instructional Support for Collaboration in Desktop Videoconference Settings. In R. Bromme, F. Hesse, & H. Spada (Eds.), *Barriers and Biases in Computer-Mediated Knowledge Communication* (Vol. 5, pp. 59-88): Springer US. doi: 10.1007/0-387-24319-4_4
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4(1), 69-92. doi: 10.1007/s11412-008-9054-4
- Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating Information An Analysis of the Processes Involved and the Products Generated in a Written Synthesis Task. *Written Communication*, 30(1), 63-90. doi: 10.1177/0741088312466532
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of second language writing*, 14(3), 153-173. doi: 10.1016/j.jslw.2005.05.002
- Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, 31, 275-288. doi: 10.1017/S0267190511000079
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Bristol Buffalo: Multilingual matters.
- Strobl, C. (2015). Learning to Think and Write Together: Collaborative Synthesis Writing Supported by a Script and a Video-based Model. In T. Guasch & M. Dean (Eds.), *Learning and teaching writing online: Strategies for success*. Leiden: BRILL.
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order

- peer feedback in EFL. *Learning and Instruction*, 20(4), 316-327. doi: 10.1016/j.learninstruc.2009.08.009
- Vygotskiĭ, L. & Cole, M. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, 11(2), 121-142. Retrieved from <http://search.proquest.com/docview/85678931?accountid=11077>
- Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of second language writing*, 21(4), 364-374. doi: 10.1016/j.jslw.2012.09.005
- Würffel, N. (2008). Kooperatives Schreiben im Fremdsprachenunterricht: Potentiale des Einsatzes von Social-Software-Anwendungen am Beispiel kooperativer Online-Editoren. [Collaborative writing in foreign language education. Affordances of social software using the example of collaborative online editors]. *Zeitschrift für interkulturellen Fremdsprachenunterricht*, 13(1), 26.

Appendices

Appendix A: Illustration of rating procedure by two dimensions

The rating procedure is exemplified by two dimensions: *communication flow* (2) and *language-related information pooling* (4). Examples from the interaction protocols are used to illustrate the dimensions and the rating procedure, with in the first column a reference code to the chat (C), comment (K) or revision (R), in the second column a reference to the student code (a, b or c) and in the third column the transcription (in Dutch, often with German words and occasionally English), followed by a translation to English (in italics).

C: chat, K: comment, R: revision, a/b/c: anonymized student code

[a], [b], [c]: reference to a student in the chat

[...]: part of chat or comment transcription omitted

[x]: explanation, x/[x]: translation

*: switch of communication channel (chat – comment – revision)

Dimension 2: Communication flow

Description: Students maintain a steady communication flow by using the interaction facilities of the online collaboration tool appropriately. They notice each other's actions and react to each other's queries and suggestions without undue delay. When the communication flow is interrupted or neglected, the group members try to breathe new life into it. They should also report about their individual work, e.g. by sharing how long the individual writing will take, and ask each other questions about their work progress to avoid chaos and duplication. Questions about content or language concerns should be answered.

Aspects:

- Noticing peers' actions
- Answering peers' questions and/or suggestions
- Reporting on individual work progress
- Intensity of the collaboration
- Appropriate use of the interaction facilities

When: in all writing phases (planning, translating and revising)

Quality levels:

1. The communication flow is very poor. The group members do not notice each other's actions and they have to wait a very long time before their questions or suggestions are answered. The group members do not report on their individual work. There are very few interaction sequences and the communication channels are used in a very inefficient way.
2. The communication flow is poor. The group members mostly do not notice each other's actions and they have to wait a long time before their questions or

suggestions are answered. The group members hardly report on their individual work. There are few interaction sequences and the communication channels are used in an inefficient way.

3. The communication flow is intermediate. The group members notice each other's actions to a certain extent and their questions or suggestions are answered within a reasonably short time. The group members report on their individual work more or less. There are a certain amount of interaction sequences and the communication channels are used mostly used efficiently.
4. The communication flow is good. The group members notice each other's actions and their questions or suggestions are answered quickly. The group members report on their individual work. Many interaction sequences can be found and the communication channels are used in an efficient way.
5. The communication flow is very good. The group members always notice each other's actions and their questions or suggestions are answered promptly. The group members constantly report on their individual work. Many extensive interaction sequences can be found and the communication channels are used in a very efficient way.

Example 1. Reporting on individual work progress and in a high-proficiency M1 triad, model session (score: 5)

C19	4b	ben al beetje bezig hoe we de tekst kunnen indelen maar help maar hoor ;)	<i>I'm already looking at a possible structure for the text but you guys can help if you want ;)</i>
C20	4c	Ok	<i>Ok</i>
C21	4a	Ja ik ben ook bezig 😊 zeg anders wat je al hebt	<i>Yes I'm also busy doing that 😊 maybe you can say what you already have</i>
C22	4b	Ben bezig in docs dus je kan het zien ;) vanonder [...]	<i>I'm working on it in docs [Google Docs] so you can see it ;) at the bottom</i>
C43	4a	Zijn we klaar met de stukjes te plakken onder het juiste kopje?	<i>Are we ready pasting the pieces under the right heading?</i>
C44	4b	ja:D	<i>yes:D</i>
C45	4c	ja [...]	<i>yes</i>
C56	4a	Zo, ik ben klaar. Mijn tekst staat helemaal vanonder. Hoe zit het bij jullie?	<i>Ok, I'm ready. My text is at the very bottom. How are you doing?</i>
C57	4b	ik ben bijna klaar, juist nog jouw stukje, maar zal hem al plakken zodat je al wat kan nalezen ;)	<i>I'm almost finished, just your part, but I'll paste it so you can read it through already ;)</i>
C58	4a	Ok 😊	<i>Ok 😊</i>
C59	4b	ziezo, 'k ga nu nog jouw stukje doen, schrijf maar opmerkingen als er iets fout is	<i>there, now i'll do your part, don't hesitate to make comments if something is wrong</i>

Example 2. Commenting one's actions in the text document in a low-proficiency S1 triad, script session (score: 4)

C19	4c	ja maar dan zou ik wel die eerste zin van het 2e deel een beetje [sic] veranderen, zodat het vlotter leest als een vloeiende tekst	<i>Yes but in that case I would change the first sentence from the second part a bit, so it reads more fluently like a fluent text</i>
C20	4a	ja inderdaad, en dan ook zeggen dat het niet meer over de reclame gaat	<i>Yes exactly, and then we should say it's not about advertising anymore</i>
C21	4c	Zoiets?	<i>Something like that?</i>
C22	4c	Ja is goed	<i>Yes that's good</i>
C23	4b	Vree goed! Ik ben [sic] tekst aan het bewerken hoor 😊	<i>Very good! I'm editing the text 😊</i>

Dimension 4: Language-related information pooling

Description: When writing in an L2, beside information related to the writing subject, students also should pool their linguistic resources about L2 writing. Students can help and support each other during the writing process, but especially during the revision phase, for various linguistic topics, on word- and sentence level (orthographic, syntax, lexicon, morphology) as well on textual level (content, coherence and cohesion).

Aspects:

- Number of peer-induced revisions
- Degree of assistance each other with L2 issues
- Focus of the revisions (LOC, HOC)
- Success of collaboratively solved revisions

When: in the planning and translating phase

Quality levels:

1. Language-related information pooling is very poor. No peer-induced revisions or a very small number of them are made (between 0 and 3). Linguistic knowledge is not pooled.
2. Language-related information pooling is poor. A small number of peer-induced revisions are made (between 4 and 8). The focus of the revisions is on LOCs. Some revisions are unsuccessful. Little linguistic knowledge is pooled.
3. Language-related information pooling is intermediate. There is a certain number of peer-induced revisions (between 9-15). The focus of the revisions is on LOCs, some HOCs are being discussed. Most of the revisions are successful. Linguistic knowledge pooled to a certain extent.
4. Language-related information pooling is good. Many peer-induced revisions are made (between 16-24). The focus of the revisions is balanced between LOCs and HOCs. Almost all revisions are successful. Linguistic knowledge is pooled.

5. Language-related information pooling is very good. The number of peer-induced revisions is very high (more than 25). The focus of the revisions is on HOCs. All revisions are successful. A great deal of linguistic knowledge is pooled.

Example 3. Pooling of linguistic knowledge and peer revision in a high-proficiency S1 triad, script session (score: 4)

This episode exemplifies how students can successfully call on each other when writing in an L2. Student C is unsure how to decline a substantive (das Deutsch or das Deutsche) and asks this question to her group members in the chat (C50). Then, student A corrects this word (R21: im Deutschen), at which point student C is again in doubt, and she unsuccessfully revises it by changing the upper case into a lower case (R22: im deutschen). Student A comments on this incorrect revision in the chat (C51). Student C gives in, changes it into upper case (R23) and saves face (C52, C54).

C50	3c	[...] gebruik je das Deutsch of das Deutsche?	[...] <i>do you use</i> das Deutsch <i>or</i> das Deutsche [lexical question concerning declension, correct answer: both correct]?
R21*	3a	<i>changes the case: das Deutsch > im Deutschen</i> [no change, both correct]	
R22	3c	<i>changes the capital letter in a small letter : im Deutschen > im deutschen</i> [unsuccessfull revision]	
C51*	3a	im Deutschen is met hoofdletter hoor :)	<i>im Deutschen is with a capital letter you know</i> [correct remark]
C52	3c	das dan raar, want in dtv 1 zeiden ze van ni ofwel auf Deutsch, ofwel im deutschen	<i>well that's weird, because in dtv 1 [course] they said it isn't, either auf Deutsch or im deutschen</i>
C53	3a	ahso	<i>ah owkay</i>
C54	3c	google zegt ook hoofdletter die van dtv 1 zijn dan ook wel ontslagen	<i>google also says with a capital letter, those from dtv 1 are fired then</i>
R23*	3c	<i>changes the small letter in a capital letter: im deutschen > im Deutschen</i> [successfull revision]	
C55	3b	jup, google zegt im Deutschen	<i>Jep, google says im Deutschen</i>
C56	3a	ja ik ben het eigenlijk vrij zeker	<i>I'm actually pretty sure</i>
C57	3b	:P	<i>:P</i>
C58	3a	want in mijn hausarbeit in oostenrijk heb ik dat 120 keer gebruikt	<i>Because in my hausarbeit [assignment] in Austria I used that 120 times</i>
C59	3c	ok, dan weet ik nu het juiste [...]	<i>Ok, well now I know the correct form [...]</i>

Example 4. Pooling of knowledge about reading fluency and suggesting peer revision in a high-proficiency M1 triad, script session (score: 4)

- | | | | |
|-----|----|---|--|
| K43 | 4a | dit leest een beetje moeilijk vind ik, zo twee keer zo + inf. Mss zo beter: die Ängste der Österreicher zu überwinden, dass sie ihre Sprache... Verlieren würden? | <i>this is a bit difficult to read I think, two times + inf [infinitive]. Maybe better like this: die Ängste der Österreicher zu überwinden, dass sie ihre Sprache... Verlieren würden? [to allay the fears of the Austrians, so they... lose their language?]</i> |
| K44 | 4b | ja khad het ook moeilijk [sic] toen ik het zelf las. Idd goed alternatief! Wat denk jij [c]? | <i>yes I also had difficulties when reading it. Indeed, that's a good alternative! What do you think [c]?</i> |
| K45 | 4c | Ja goed! | <i>Yes good!</i> |

Appendix B: Script used to support the collaborative writing process

Work step	What to do	Where to do it	Approx. starting t	Approx. duration
1 (ind)	Read your source text and summarise the main thoughts in bullets. Use your own words as much as possible. Mark literal textual borrowing as quotation and add the source (who said / wrote it?) Copy-paste your summary to the common Google Doc.	MS Word -> Copy-paste to Google Doc	10h40	15 min
2 (coll)	Read all summaries and discuss the content. Don't focus too much to language right now.	Comment in Google Doc	10h55	15 min
3 (coll)	Select the main propositions for your joint synthesis . Compose a text skeleton ¹ . Distribute the text to be written into three parts.	Chat + comment in Google Doc	11h10	15 min
4 (ind)	Write a first version of your synthesis part. Respect the word limit.	Word -> Google Doc	11h25	15 min
5 (coll)	Carefully read all parts of the synthesis. Discuss in group: Is everything clear and correct? Do the text parts match with each other? Does the ensemble represent a coherent text that reads fluently? Add transition passages and delete repetitions. Change the proposition order if appropriate. Focus on content and language alike.	Comment and text in Google Doc	11h40	20 min
6 (coll)	Re-read your synthesis a last time and edit until you are all satisfied with the result. Respect the word limit.	Google Doc	12u00	10 min

¹ skeleton = bulleted outline enriched with cohesive ties

5.4 Learning to write syntheses: Changes in individual writing after collaboration (Article submitted to *Journal of Second Language Writing*)³⁵

Abstract

Synthesis writing from multiple sources places a high cognitive load on the writer, especially so in the case of L2 writers, who struggle with limited linguistic resources and a natural propensity to focus on the local textual level. Peer collaboration can help novice writers, as it allows them to pool their cognitive and linguistic resources. Furthermore, collaborative writing fosters content elaboration and reader-orientation. This study investigates whether online collaboration could benefit *individual* writing development. The participants (n=76), collaborating synchronously in Google Docs, wrote three syntheses in triads that were homogeneous or heterogeneous in terms of their proficiency in German L2. Individual pre- and post-tests were compared, whereby both writing process and product were analysed. This revealed an increase in planning strategies and in global revision, as well as an improvement in content selection and overall structure. Proficiency grouping during collaboration had no significant impact on these changes from pre- to post-test. Furthermore, no change was observed regarding overall coherence and rephrasing behaviour. On the contrary, the amount of textual borrowing from sources even increased. This *patch-writing* strategy in students' academic text production and general implications for instruction are discussed.

Keywords

Advanced foreign language instruction, Higher education, Synthesis writing, Strategy development, Collaborative writing

Highlights

- Students, collaborating synchronously in triads via Google Docs, wrote three syntheses
- Pre- and post-tests reveal changes in individual writing after collaboration
- Increased fluency, planning, and content elaboration in post-test

³⁵ Strobl, C., Raes, A., & Schellens, T. (submitted): Learning to write syntheses: Changes in individual writing after collaboration. *Journal of Second Language Writing*

- Coherence and textual borrowing remain problematic issues in post-test
- A significant impact was found for individual proficiency, but not for proficiency grouping

Introduction

Challenges for novice writers engaging in (L2) synthesis writing

Summarisation stands out as a highly effective strategy in learning-to-write, as Graham and Perin (2007) stated in their meta-analysis of writing instruction for adolescent learners. Moreover, summary and synthesis are important genres that students in tertiary education need to master in order to write literature reviews. In spite of their recognised high learning potential and importance, summarisation strategies are seldom explicitly taught in higher education (Doolittle, Hicks, & Triplett, 2006). Students are expected to know how to summarise when they enter university. However, according to writing development theories, this is highly unlikely. Summary and synthesis writing are subgenres of academic writing and require processes of *knowledge crafting* which can usually only be mastered by the age of early adulthood (Kellogg, 2008, p. 1). Knowledge crafting implies a threefold mental representation of the emerging text in the writer's working memory, being the author's intention, the actual text, and the reader's interpretation. Allocating attention to these three representations simultaneously places a lot of strain on working memory capacity. Therefore, other cognitive processes, like planning and sentence generation, need to be automated in order to overcome processing constraints (Torrance & Galbraith, 2006).

Synthesis writing from sources requires several processes to be carried out by the writer. Spivey (1990) identified *selecting*, *organising* and *connecting* source content as three essential processes. As synthesis writing includes the sub-task of summarising the different source texts, first, their *macrostructure* has to be established by the writer (Kintsch & van Dijk, 1978). Together, these four processes represent the necessary *planning* strategies required for synthesis writing. During the actual synthesis *composition*, two major processes are at play: textual appropriation and establishing text cohesion. With regard to the former, adequate strategies are needed to integrate and rephrase the source texts. Text cohesion is an important feature that helps guide the reader. When composing a synthesis, a "superproposition" (Segev-Miller, 2004, p. 6), i.e., the writer's own macroproposition of the synthesis, has to be built, where the different representations of the text are joggled, and, above all, where the imagined reader is considered. Therefore, appropriate cohesive ties have to be chosen in order to represent the underlying textual coherence. Finally, when *reviewing* the text, again, the mental representation of the reader has to be taken into account, which means that coherence and cohesion, i.e., *global* aspects of the text, need special attention. Table 1 represents

the different subprocesses required for synthesis writing, mapping them to the three major writing stages of the Hayes and Flower (1980) model of composition writing.

Writing stage	Synthesis subprocess
Planning	(1) Establish the macrostructure of source texts
	(2) Select main propositions
	(3) Merge the main propositions of the different source texts
	(4) Establish writer's own coherent structure
Translating	(5) Rephrase in writer's own words
	(6) Insert cohesive ties for reader-orientation
Reviewing	(7) Revise text on both local and global levels

Table 1 Seven subprocesses required for synthesis writing, mapped to the three stages of the writing process proposed by Hayes and Flower (Hayes & Flower, 1980)

Table 1 illustrates that synthesis writing requires extended planning. The first major challenge for novice writers, therefore, is to gain awareness of the important role that planning and elaboration play in summarising processes. Students need to *elaborate* the content of the source texts in order to gain deep understanding (Solé, Miras, Castells, Espino, & Minguela, 2013). The second challenge they encounter is that, during translating and reviewing, they need to broaden their focus of linguistic attention from word and sentence level to include text level (Paulus, 1999).

According to Kellogg (2008, p. 7), college freshmen have a natural propensity to limit their reviewing attention to local aspects, e.g. lexical changes, while knowledge crafting requires attention to global aspects, which leads to structural changes. This natural propensity is aggravated when writing in an L2, where linguistic accuracy is a major concern. Scholarship in L2 writing has highlighted the limited lexical variety, specificity, and sophistication as compared to L1 (native language) writing, with important implications for cohesion (Crossley & McNamara, 2009; Hinkel, 2002). Plakans (2009) found that vocabulary size also had an impact on the discourse synthesis process of ESL (English as a Second Language) students, both during reading and writing. Moreover, the retrieval of linguistic knowledge in L2 writing requires additional memory capacity. Chenoweth and Hayes (2001) found that writing fluency as measured in production burst length, was considerably lower in L1 than in L2. Galbraith (2009) points to the importance of writing fluency in knowledge-constituting processes, and concludes that "the writer finds it harder to articulate their [sic] personal understanding in L2" (*op. cit.*, p. 19). This underpins the difficulties L2 writers face to

introduce their own voice through rephrasing while summarising in order to avoid literal reproduction.

A specific characteristic of synthesis writing is that it is "heavily intertextual" (Pecorari, 2013) in that the writer has to incorporate external textual sources into his/her own text. Finding an academically acceptable balance between integration and appropriation of the other's voice in the writer's own text also requires apprenticeship. Untrained student writers tend to *patchwrite* their literature reviews, a strategy that often reflects a lack of elaboration and can cause plagiarism (Keck, 2014). Drawing on Howard et al. (2010), Pecorari defines *patch-writing* as a form of textual borrowing that is situated between paraphrase and quotation, "stitching together elements from one text with elements from another and making some superficial changes to the language" (Pecorari, 2013, pp. 70-71). A broad range of studies have examined the problem in depth, investigating copy-pasting in L2 academic writing from viewpoints of limited attentional and linguistic resources (McDonough, Crawford, & De Vleeschauwer, 2014), international/interdisciplinary differences between academic cultures (Pecorari, 2013), or students' attitudes and perceptions (Shi, 2012). Whatever the underlying reason, it is important to sensitise students to the boundaries between acceptable and unacceptable textual borrowing (Pecorari, 2013).

Promising routes for L2 synthesis instruction

Since research on the effect of writing instruction based on quality parameters of students' written *products* in (L2) academic writing has led to diverging conclusions (Nicolás-Conesa, Roca de Larios, & Coyle, 2014), *process* approaches have been adopted to investigate academic writing development. Plakans (2009) investigated the composing processes of ESL writers during both the reading and writing processes in synthesis production. She found that the degree to which the participants employed discourse synthesis processes of organising, selecting, and connecting during source elaboration had a direct impact on the number of ideas from the source texts they integrated in their syntheses. Rooted in the cognitive paradigm, the development of students' mental models of the academic writing task (or task representations) and/or their writing goals have been a focus of research (Manchón, 2012; Nicolás-Conesa et al., 2014; Wolfersberger, 2007). Nicolás-Conesa et al. (2014) found that a process-oriented approach to writing prompts students to adopt more problem-solving behaviour, strive for "more sophisticated sub-goals when composing and thus [producing] better texts" (Nicolás-Conesa et al., 2014, p. 16). Based on their findings, they recommend that "teachers might foster the conception of writing in terms of problem-solving behaviour through instruction and recursive writing practices oriented towards writers' goals" (Nicolás-Conesa et al., 2014, p. 16).

In the same vein, Seidlhofer (1995) argues that task representation plays an important role in the strategy use of students engaged in academic writing. She differentiates between two task types (summary vs. account) and respective students' roles as *animator* (merely reproducing another author's text intention) vs. *author* (interpreting another author's text as a reader). Seidlhofer points to the need of raising students' awareness of both writer's roles, and the different strategies they require, and advocates to explicitly train both writer's roles. An important goal of academic writing instruction, in her view, is to sensitise students to "what the conditions for membership are in the relevant (foreign language) discourse communities, and what the leeway is for individual action within those" (Seidlhofer, 1995, p. 218).

Wette (2010) reported on the success of explicit strategy instruction for textual appropriation, i.e., paraphrasing, summarising, and adequate and correct quotation, which had a beneficial effect both on the declarative knowledge of the participants regarding adequate textual borrowing and on the amount of fully acceptable source citations in their post-intervention writing from sources. Zhang (2013) linked explicit instruction on general discourse synthesis strategies with in-class writing practice in an experimental study with ESL students that involved a control group who did not receive instructions specifically related to synthesis writing. The results of pre-post-test holistic scores comparison revealed significant differences between groups, suggesting that explicit instruction had a clearly beneficial impact on discourse synthesis writing ability.

To conclude, scholarship on (L2) synthesis instruction has provided evidence that stimulating deliberate practice and strategy use can help improve adequate textual appropriation as well as general synthesis writing ability. Based on sociocultural theory of language learning, such deliberate practice can also be fostered in collaborative settings.

Benefits of collaborative writing

The benefits of collaborative settings, both online and face-to-face, for the development of writing skills have been largely investigated, highlighting a positive impact on the product in terms of accuracy (Davoli, Monari, & Severinson Eklundh, 2009; De la Colina & García Mayo, 2007; Fernández Dobao, 2012; Lund, 2008), as well as on reader-orientation (Blin & Appel, 2011; Kuteeva, 2011). The qualitative inspection of collaborative writing and revision processes has provided evidence for successful peer scaffolding relating to both focus-on-form (Guerrero & Villamil, 2000; Hanjani & Li, 2014; Strobl, 2015; Wigglesworth & Storch, 2012) and focus on higher-order concerns, such as textual coherence and content selection (Strobl, 2015; Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010). According to Storch (2013), the outcome of a collaborative writing activity is not limited to the jointly produced text, but also concerns *collective cognition*, which she defines as "learning new vocabulary, improved

ways of expressing ideas, gaining a greater understanding of certain grammatical conventions or greater control over the use of a particular grammatical structure" (Storch, 2013, p. 4). Indeed, regarding the impact of collaboration on individual L2 learning, it has been shown that lower-order issues, e.g. grammar and vocabulary, solved in collaborative language-related episodes, were recalled by individuals in delayed post-tests (Kim, 2008). This confirms the hypothesis that collaborative dialogue fosters second language learning (Storch, 2013; Swain, 1997; Swain, Brooks, & Tocalli-Beller, 2002). The question arises whether a positive transfer also occurs with regard to strategies related to higher-order issues, e.g. outlining, acquired in a collaborative writing experience. Within the sociocultural paradigm, strategy development is regarded as highly interactive (Lei, 2009). Thus, collaborative learner interaction should offer potential for writing strategy development.

However, good collaboration that leads to the intended interactions does not come by itself. Especially when the collaborative task is complex in nature, it is necessary to support the collaboration process through instructional means (Kirschner, Martens, & Strijbos, 2004). Scripting and modelling are two forms of effective instructional scaffolding that have been well-researched in CSCL settings (Dillenbourg & Jermann, 2007; Hämmäläinen & Arvaja, 2009; Rummel & Spada, 2007; Rummel, Spada, & Hauser, 2009). Especially *observational learning* through coping models has been reported to yield effect on the development of writing and reviewing strategies (Couzijn, 1999; Raedts, Rijlaarsdam, van Waes, & Daems, 2007; Van Steendam et al., 2010). Pre-task modelling was also proposed by Storch (2013) to encourage learners to adopt collaborative patterns in face-to-face interaction during writing tasks.

The main research question of the present article therefore is: Does students' individual synthesis writing change after a collaborative writing experience in which they were supported by scaffolding instruments? More specifically, can we observe an increase in individual planning strategies, content and linguistic elaboration, and focus on higher-order concerns in revision (RQ1)? Secondly, if so, (how) does this change in *process* translate to *product* quality (RQ2)? In addition, a possible impact of individual L2 proficiency on the transfer from collaborative to individual writing should be investigated. The influence of group composition on task achievement and collaboration quality has been largely discussed in CSCL scholarship (Kirschner et al., 2004; Schellens, van Keer, Valcke, & De Wever, 2007; Wang & Lin, 2007). Studies in L2 collaborative writing have investigated the impact of proficiency pairing (homogeneous vs. heterogeneous level of L2 proficiency) on negotiation leading to focus-on-form in meaning-oriented tasks. Storch (2013, pp. 57-60) provides an overview of studies that investigated the number, focus, and resolution of LREs (language-related episodes) in face-to-face talk of pairs while engaged in collaborative writing activities. Leeser (2004) found that significantly fewer LREs were initiated and correctly resolved by low-low pairs than by high-high or high-low pairs in a dictogloss task, and that in the

heterogeneous pairs, only 8% of the LREs were correctly resolved by the low-proficiency peer. This finding led him to question whether low proficiency students could benefit from a collaborative setting in a task without information gap. Storch and Aldosari (2013) reported similar findings for a composition task and added to the discussion about the benefit of heterogeneous pairing for low-proficiency students by pointing to the negative impact of dominant/passive interaction patterns. However, when collaborative interaction patterns prevail, heterogeneous pairing has been shown to be fruitful for the low proficiency peer (Ohta, 1995; Storch & Aldosari, 2013).

To our knowledge, the aspect of proficiency pairing has not been studied with regard to strategy development in advanced L2 writing instruction. Therefore, we tackle this aspect in a third research question, viz., whether group composition according to proficiency level plays a modifying role in the strategy transfer from collaborative to individual writing (RQ3).

Method

Context of the study and participants

The participants of our study (n=76) were enrolled in the third study year of two disciplines with a strong focus on second language (L2) proficiency, being Applied Language Studies at a university college, and Linguistics and Literature at a university. All participants are Dutch native speakers with an advanced language proficiency for German (B2-C1 of the Common European Framework)³⁶, which was the targeted L2. Students were assigned to collaborative groups of three according to proficiency level based on pre-test scores. In order to investigate the impact of group composition, low-proficiency, high-proficiency, and mixed-proficiency groups were composed.

Study design

The intervention study was carried out inside the classroom as part of a curricular course. In three consecutive weekly sessions of 90 minutes, the participants each time wrote a synthesis based on three popular scientific texts in collaborative triads, using the online editor Google Docs for writing and communication. Although they were seated in the same PC classroom to carry out this synchronous collaboration task, they communicated exclusively online. In the first week of collaboration, they only received general guidelines with regard to online collaboration via Google Docs. In the second and third week of collaboration, they received instructional support to structure the task and the workflow in the form of a script (see Appendix A) and a model video of an

³⁶ Correspond with ACTFL Proficiency Guidelines levels 'Advanced mid' and 'Advanced high'.

online writing collaboration, which pointed out or exemplified the desirable work steps and strategies to be implemented. In the collaborative planning phase, a *jigsaw* concept was applied, in that each student only read one source text and was held responsible for summarising its main ideas for the peers (see Figure 1). Beside fostering positive interdependence between the peers (Johnson, Johnson, & Holubec, 1998), this information gap also served as extrinsic motivation to elaborate the content of the source texts through negotiating meaning.

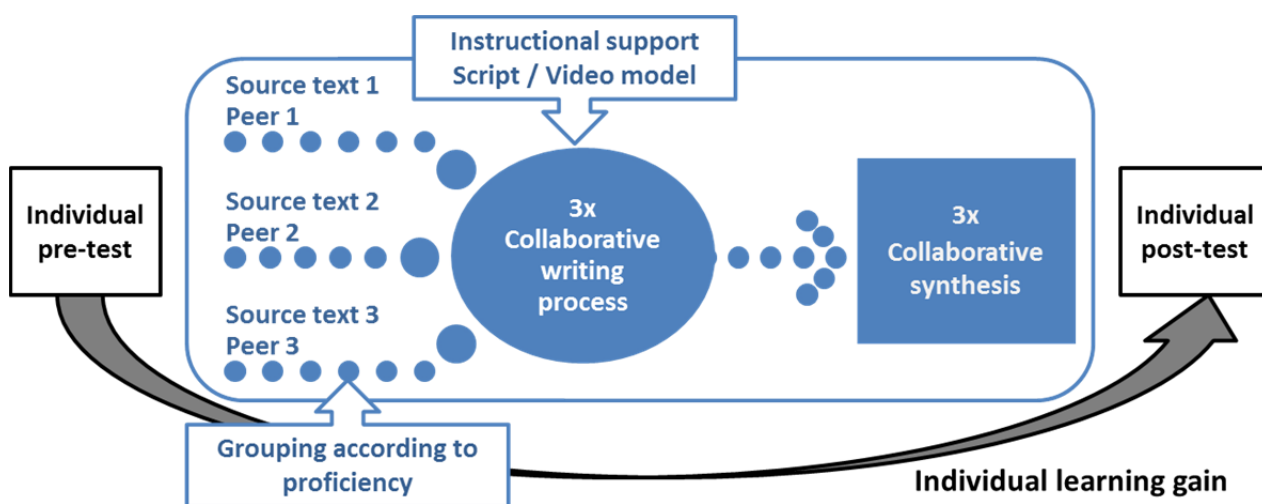


Figure 1. Study design including collaborative writing task and individual pre-post-tests

The impact of this collaborative writing intervention on individual synthesis writing was measured by comparing an individual pre- and a post-test that consisted of summarising the content of two popular scientific texts of about 450 words each in one synthesis of maximum 250 words. The time allotted for this individual task to be carried out in Word was 60 minutes. All texts used in the intervention and the pre- and post-tests were comparable in terms of topic (linguistic variations of German) and difficulty (authentic texts taken from German quality press). The pre- and post-test composition processes were recorded with *Inputlog* (Leijten & Van Waes, 2013), a keystroke logging tool that allows the visualisation of the writing process for qualitative analysis, and automated retrieval of process and product measures for the quantitative analysis.

Process measures

Both the writing process and the product of the individual pre- and post-tests were analysed. The individual process analysis was based on four different components, combining analytic and holistic measures (see Table 2).

Component	Analysis unit	Rating	Scale
Fluency	Characters/minute	analytic	n.a.
	Active writing time / pause time	analytic	n.a.

Time management	Time spent reading sources	analytic	min.
Elaboration while planning	Summarisation of sources	analytic	0-4
	Outlining for synthesis	holistic	0-2
Writing and reviewing behaviour	Product/process -Ratio	analytic	0-1
	Type of revisions	holistic	0-2

Table 2. Process measures

The three measures needed for fluency calculation, being characters per minute, active writing time, and pause time, were retrieved from the output file "summary analysis" generated by *Inputlog*. Another output file of *Inputlog* is the writing process graph which provides a visual overview of the process. From this graph (see Figure 2 below), we derived the measure for time spent reading sources, being the number of minutes without writing activity at the beginning of the task.

The introspection of the writing process in play-back mode allowed us to track planning behaviour, which was subdivided into two quantitatively coded analysis units. The coding for elaboration while planning is based on the following strategies that were recommended or modelled, as the case may be, during the collaboration sessions:

- a) For further elaboration and integration of the source content, (elaborated) bullets are more favourable than a full text.
- b) The summary bullets should be reshuffled and linked in an enriched outline to create coherence before writing the first draft.

To quantify the variable "summarisation", five categories were established (0=no elaboration, 1=full text, 2=paragraphs, 3=bullets, 4=elaborated bullets). "Outlining" was coded in the three categories "no outline", "raw outline" (i.e., only basic titles for the parts of the text to be written), and "full outline".

Finally, two more measures were used to analyse overall writing style and reviewing behaviour. The product/process-ratio can be retrieved from the "summary analysis" output file of *Inputlog* (total number of characters in the final text / total number of characters produced during the writing process). The closer its value is to 1, the more linear the overall writing process, i.e., the fewer revisions took place. However, this value has to be interpreted with caution, and a visual verification of the writing process in play-back mode is needed. After all, students follow different strategies when elaborating on the bullets or text fragments they produced in the planning stage. Some start to write the first draft afresh, step-by-step deleting the planning fragments after their elaboration, or just leaving them in the document, while others use these bits of

text as a kind of a skeleton, building their text around them. The latter two strategies lead to a higher value for product/process-ratio, as in total, less text is deleted during the process, while they can certainly not be interpreted as more (or less) linear than the first strategy. A visual analysis of the process graph in combination with the play back mode helps to analyse the process more in depth.

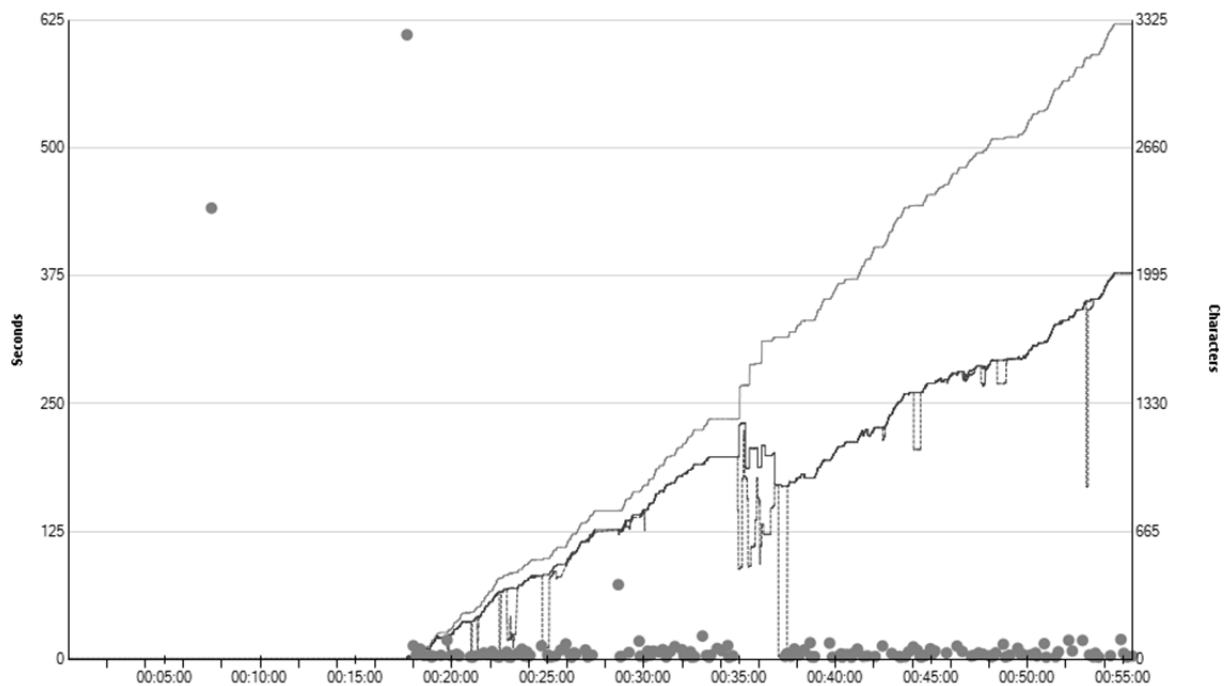


Figure 2. Example writing process graph output file from *Inputlog*

Figure 2 showcases a writing process graph. The upper, grey line depicts the number of (typed or copied) characters produced. The lower, black line represents the actual number of characters in the document. The distance between these two lines indicates the number of deletions in the text already produced. The black dotted line displays the cursor position. Where it does not overlap with the continuous black text production line, dropping in direction of the x-axis, the cursor has been moved to other positions than the end of the text produced so far, which indicates a revision of already written text. The grey dots represent pausing time which has not been used in the present analysis. In this case, the student spent the first 18 minutes reading the sources without writing. Then, he started producing text (here: bullets summarising the source texts) in a fairly linear way (the black and the grey line movements coincide rather well), occasionally revising and/or deleting already written text. Between minute 35 and 38 of the writing process, major portions of text already produced are being deleted, because the distance between the grey and the black line increases abruptly. Introspection of the writing process in play-back mode shows that the student actually started to write his first draft at this point, deleting the summary fragments considered insignificant for the synthesis. During the production of the first (and only) draft, only occasional reviewing

activities took place (e.g. in minutes 44, 48, and 53). In order to gain an insight into the prevailing revision type ("0=no revisions", "1=only local", "2=local and global"), a visual analysis of the writing process in play-back mode is required.

Strategy questionnaires

In order to triangulate the observational data concerning the writing process, pre-post-questionnaire responses regarding writing strategy use are compared. These questionnaires were administered before the pre-test and after the post-test on synthesis writing. They comprised 20 writing strategies that were formulated following the example of Oxford's (1990) SILL (Strategy Inventory for Language Learning) as statements in the first person. Students had to indicate how frequently they used the respective strategy on a 5-points-Likert scale. Contrary to the SILL, which covers with its 49 items a broad range of strategies to be used in all kinds of language-based interactions, our questionnaire was adapted to the writing task under research. It includes 16 general strategies for planning, composing, and reviewing (e.g., "During writing, I reread the text I have already produced so far"), as well as four specific strategies for writing syntheses (e.g., "I change the structure of the source texts when writing a synthesis if this helps to improve coherence").

Product measures

To measure changes in product quality, a CAF (complexity, accuracy, and fluency) analysis commonly applied in the analysis of L2 writing development (Wolfe-Quintero, Inagaki, & Kim, 1998) was not deemed adequate to detect a possible impact of the intervention. After all, it targeted strategies concerning mainly higher-order textual aspects. Therefore, analysis units related to content and coherence were chosen, which are a mixture between a holistic and an analytic rating system (see Table 3).

Component	Analysis unit	Rating	Scale
Elaboration of content	Content selection	analytic	0-5
	Structure	analytic	0-2
Coherence	Overall coherence	holistic	0-3
	Density of cohesive ties	analytic	0-1
Textual appropriation	Ferret similarity rate	analytic	0-1

Table 3. Product measures

Elaboration of content was measured in two analytical scores. With respect to content selection, the text was compared to a pre-established list of main propositions, and .25 points were deducted for each missing item. Regarding structure, a division of the text into three minimal sections, namely introduction, body, and conclusion, was attributed

the score 1(2), and an additional subdivision of the body in paragraphs according to main propositions led to an increase with .5 and 1 of the score. Coherence is used here as an umbrella term, as it was measured based on two scores that actually combine features of coherence (i.e., the underlying logical structure of propositions) with cohesion (i.e., the representation of coherence on a textual level). To calculate the analytical "cohesive density" score, all cohesive ties according to the five categories proposed by Halliday and Hasan (1976), being reference, substitution, conjunction, lexical cohesion, and ellipsis, were counted and divided by the total number of written words. The density of cohesive ties, however, does not necessarily reflect a high coherence, as the choice of cohesive ties might be inadequate, or even misleading, especially in L2 writing. Therefore, a second, holistic score for 'overall coherence' was added to measure overall reading fluency. Inter-rater reliability (IRR) for this qualitative analysis was calculated based on the comparison of the scores attributed by two independent coders to one third of all texts (i.e., 46 texts). To calculate IRR, a two-way mixed, consistency, average-measures intra-class correlation (ICC) (McGraw & Wong, 1996) was used. The resulting ICC (.657) was in the medium range, indicating that coders had a moderate degree of agreement and suggesting that overall coherence was rated moderately similar across coders. A discussion of cases with divergent ratings revealed that, in spite of clear guidelines and benchmark texts used in the procedure, interpretative bias was induced in the rating of overall coherence (see discussion in limitations section). Finally, the degree of textual appropriation was measured indirectly by computing the similarity rate between the source texts and the synthesis. To this aim, the plagiarism detection tool *Ferret* was used that calculates a similarity rate based on the amount of identic trigrams (Lyon, Barrett, & Malcolm, 2004). Next to the quantitative result, the tool also provides an output file highlighting all copy-pasted fragments from the source texts in the synthesis (see Figures 3a and 3b, showing one example of *patch-writing*, and one example of a synthesis produced in the student's own words).

Schweizerdeutsch **ist eine Sammelbezeichnung für die in der Deutschschweiz gesprochenen alemannischen Dialekte und wird als Umgangssprache in allen Bereichen des öffentlichen Lebens angewandt. Hochdeutsch dient als Schriftsprache oder wird in gewissen Orten bzw. Situationen auch mündlich benutzt. Seit einiger Zeit gibt es eine "Mundartwelle": der Dialekt ist allgegenwärtig und beliebt. Jugendliche verwenden es sogar als geschriebene Sprache.**

Die **Schule ist verantwortlich für schlechte Hochdeutschkompetenzen, weil in den Schulen Hochdeutsch nur als Schriftsprache verwendet wird. Die Menschen haben nicht wirklich gelernt, Hochdeutsch in alltäglichen Situationen mündlich anzuwenden.** Deshalb bevorzugen **viele Deutschschweizer im Kontakt mit Anderssprachigen, wenn Mundart keine Option sei, Englisch oder Französisch statt Hochdeutsch.** Vor 20 Jahren war das anders: **Damals hätte ein Deutschschweizer automatisch auf Hochdeutsch gewechselt, jetzt sprechen sie in der Mundart weiter.** Das erschwert auch **die Integration ausländischer Mitbürger in der Schweiz: Sie werden nur akzeptiert, wenn sie Schweizerdeutsch sprechen.**

Kinder aus der Deutschschweiz **brauchen beide Sprachen.** Als Letztes gibt es noch die Frage, wann Deutschschweizer Kinder am Besten **beginnen sollen, Hochdeutsch zu lernen.** Laut **Psychologieprofessor Allan Guggenbühl ist es besser zu warten, bis die Kinder sieben Jahre alt** seien, weil sie dann schon über eine Grundsprache verfügen. Es ist also eine gute Idee, wie eine Volksabstimmung im Mai 2011 herausgegeben hat, um im Kindergarten nur Dialekt zu reden.

Eine Studie hat gezeigt, dass es ein Bedürfnis **nach einer stärkeren lokalen Identität** gibt. Einerseits gilt dies **als Bereicherung für die sprachliche und kulturelle Vielfalt in der Schweiz,** andererseits muss gewarnt werden für offensichtliche Gefahren: es gibt Kommunikationsprobleme, was den nationalen Zusammenhalt schwächt und **der Zugang zum deutschen Sprachraum auf der kulturellen, politischen und wirtschaftlichen Ebene** wird erschwert. **Die Anwendung des Hochdeutschen muss wieder zu der Selbstverständlichkeit finden, die ihm als Landessprache zukommt.**

Die Sprachsituation in der heutigen Deutschschweiz kann als Diglossie bezeichnet werden. Das heißt, dass man je nach der Situation verschiedene Sprachvarietäten benutzt. **In der Deutschschweiz** gibt es zwei solche Varietäten: Hochdeutsch und Schweizerdeutsch, das eigentlich viele verschiedene alemannische Mundarten umfasst.

Traditionsgemäß wird Hochdeutsch, **die eigentliche Amtssprache,** als Schriftsprache benutzt, und in festgelegten Orten und Situationen auch als Umgangssprache. Die Dialekte wiederum bewältigen Alltagssituationen.

Diese Lage ist in letzter Zeit ein heißes Thema. So gab es einen Konflikt um den Sprachgebrauch in den Kindergärten: Seit 2008 förderte ein Gesetz **die Anwendung des Hochdeutschen,** 2011 wurde nach einer Volksabstimmung wieder auf Mundart gewechselt. Hier zeigt sich folgender Gegensatz: Einerseits ist nationale Kohäsion wichtig, weshalb man auf die Standardsprache setzen kann, andererseits wird aber auch auf **die lokale Identität** Wert gelegt und kann man Dialekt bevorzugen.

Der Psychologieprofessor Allan Guggenbühl befürwortet jedenfalls die heutigen Bestimmungen. Seiner Meinung nach muss ein Kind erst die Umgangssprache lernen, und erst auf dieser Grundlage kann mit dem Erwerb einer zweiten Sprache, des Hochdeutschen also, begonnen werden.

Viele sind da aber ganz anderer Ansicht: Die Beliebtheit **der Mundart und** deren allgegenwärtiger Gebrauch in Schulen, **SMS und E-Mails** würden zu mangelhaften Hochdeutschkenntnissen führen. So warnt der Sprachwissenschaftler Iwar Werlen, dass sich **viele Deutschschweizer im Kontakt mit**

Anderssprachigen nicht des Hochdeutschen bedienen, sondern z.B. des Englischen, weil sie Hochdeutsch nicht als alltägliche Umgangssprache betrachten.

Auf diese Weise droht nicht nur **der nationale Zusammenhalt** innerhalb der Schweiz schwächer zu werden, auch Deutschland wird viel schwerer zugänglich. Wichtig ist also, einen Kompromiss zwischen beiden Tendenzen zu finden.

Figures 3a and 3b. Output of the *Ferret* tool (copy-pasted fragments from sources are highlighted in bold), visualising applied *patch-writing* strategy in a post-test (above) and rephrasing of the same content in student's own words (below) respectively

Data analysis

In order to investigate a possible learning gain from the (collaborative writing) intervention measurable in the individual writing process (RQ1) and product (RQ2), we first created a basic model for multi-level-analysis, as the data are hierarchical in nature (measurements over time within individual students who, during the intervention, were part of a collaborative triad). The unconditional null model without any predictor variables provided both the overall pre-test score and the overall learning gain for all individual students across all triads. Moreover, by means of intraclass correlation, the model answered the question of whether the outcome measures varied among students, and across triads. The contribution of the triad level to pre-post-test variance proved to be insignificant for all variables (see Appendix B). Therefore, t-tests of paired samples were carried out for all analysis units, and the effect size was calculated for each comparison. To answer RQ3 on proficiency and the effect of group composition on the learning gain, we created a *merged-proficiency* variable based on individual proficiency ("high" versus "low" based on mean (6.0) split of holistic pre-test score) and group composition ("heterogeneous" vs. "homogeneous") with the four values "low in homogeneous triad", "low in heterogeneous triad", "high in homogeneous triad", and "high in heterogeneous triad". Then, we conducted one-way analyses of covariance (ANCOVA's) for all investigated features of process and product, with post-test scores as dependent variable, *merged proficiency* as independent factor, and pre-test scores as covariate to discover whether there are differences between merged proficiency levels on the post-test measure, after adjustment for the pre-test scores. The Bonferroni test, which corrects for the number of pairwise tests, was used to compare main effects. The significance level was 0.05 for all analyses.

Results

Below, first, the results for the individual process analysis are presented, followed by the results of the product analysis. At the end of each section, the results of the ANCOVA analysis regarding impact of proficiency are reported.

Process analysis

Process measures	pre	post	t	df*	p	Cohen's D	Effect size**
Characters/minute	45.65	64.66	-8.453	45	.000	-1.05994	large
Active writing time/ pause time	1.53	1.17	4.081	45	.000	0.530861	medium
Reading sources	16.68	12.48	5.288	27	.000	1.468531	very large
Summarising	1.71	2.08	-1.027	23	.315	-0.31116	

Outlining	.54	.33	1.310	23	.203	0.364302	
Product/process-ratio	.6274	.5867	1.316	30	.198	0.238012	
Type of revisions	1.44	1.63	-1.145	15	.270	-0.33654	
Mean all strategies	3.32	3.42	-3.192	71	.002	-0.39072	small

* missing data due to processing problems during file storage leading to recording errors in *Inputlog* that impeded play back, and in some occasions, automated analysis (see "limitations" section)

** effect size is only provided for significant differences of means

Table 4. Results of paired samples t-test for all dependent variables of writing process

The paired samples t-test for all dependent variables used to measure the writing process (see overview of results in Table 4) revealed a significant development of fluency, with a large effect³⁷ for typing fluency (characters/minute), and a medium effect for the ratio between active writing time and pause time, indicating that students paused less while producing their syntheses in the post-test. The pause threshold default is set at two seconds in *Inputlog*, which is a mean value for adult writers established in writing scholarship (Wengelin, 2006). Along the same lines, students spent considerably less time reading the source texts without typing notes or starting to summarise in their word documents in the post-test. In fact, seven of the 76 students failed to finish their syntheses in the pre-test which was, to a significant extent, due to a prolonged reading activity before starting to write (Pearson's correlation coefficient $r = -.513$, $p = .002$).

Regarding elaboration during the planning phase, mixed results were obtained: Whereas students in general summarised the source texts in a more favourable way for further elaboration in the post-tests (small, not significant effect), i.e., listing bullets instead of writing full text (see also Figure 4 for details), they transformed those bullets less frequently to an outline (small, not significant effect). As introspection revealed, they started to produce the first draft directly after summarising, using the bullets as a "content pool" to pick and choose from while writing.

³⁷ Cohen's (1992) rules of thumb were applied to determine the effect size for d : .2=small, .5=medium, .8=large

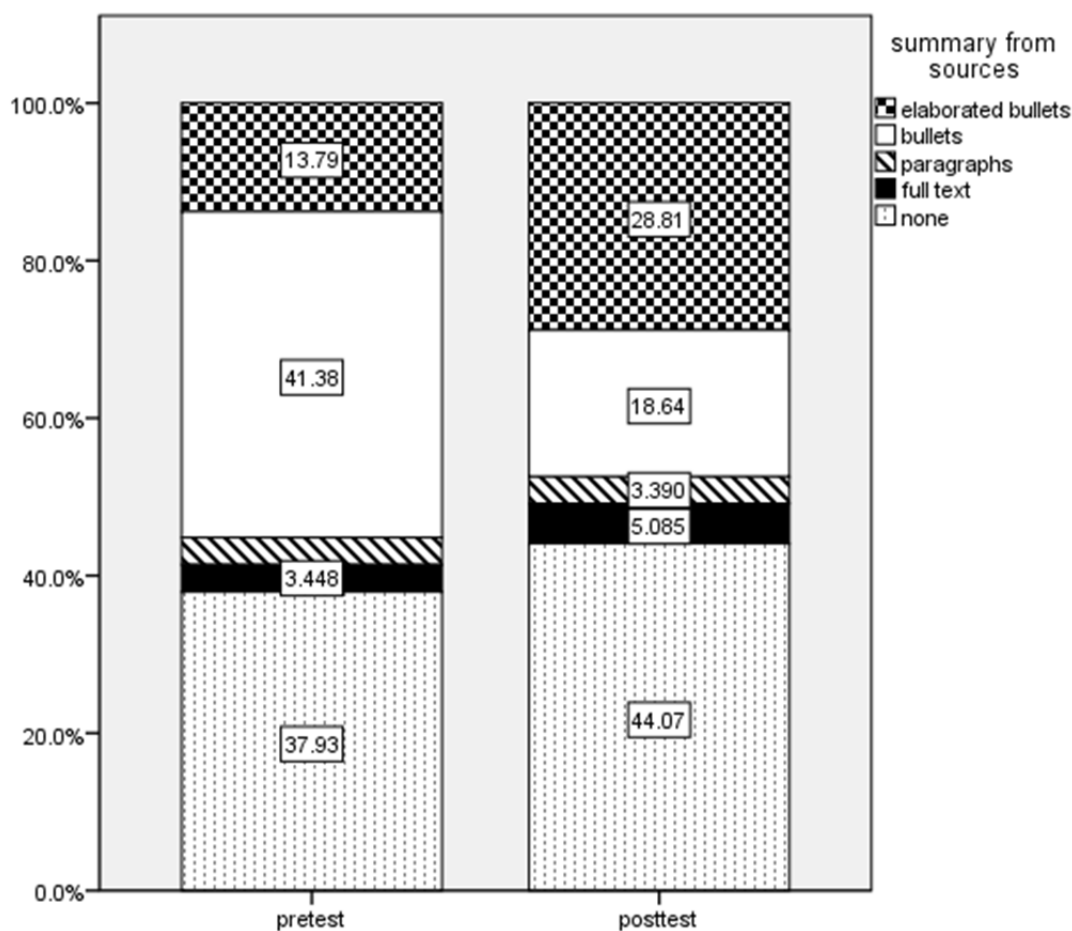


Figure 4. Summarising activities, details of the analysis (labels represent percentages)

The ANCOVA analysis revealed no significant impact of group composition according to proficiency on differences between pre- and post-tests in terms of process variables (see Table 5).

Process measures	F	p	Partial Eta squared
Characters/minute	1.260	.301	.084
Active writing time/ pause time	.690	.564	.048
Time reading sources	.978	.420	.113
Summarising	.358	.784	.054
Outlining	.796	.511	.112
Product/process-ratio	1.40	.265	1.39
Type of revisions	3.106	.071	.459
Mean all strategies	.372	.774	.016

Table 5. Results of ANCOVA analysis for impact of proficiency grouping during collaboration on all dependent individual process variables (post-test scores controlled for pre-test scores)

Also for revision behaviour, only non-significant tendencies were observed³⁸. As indicated above, the product/process-ratio allows for a first tentative interpretation of the overall writing style. The lower value in the post-test indicates that students tended to write in a more recursive way. This tendency was confirmed by a visual analysis of the writing process graphs. Regarding revision types, a small, non-significant tendency towards more attention to global aspects can be observed.

Students' self-reported writing strategies use ("mean all strategies") shows a small, but significant increase from pre- to post-hoc-questionnaire. The following four strategies stand out in terms of significant development, as paired t-tests revealed:

1. "I write my text in one flow and reread it only after I have finished writing."³⁹ (t=-2.915, p=.005, d=.34)
2. "If I cannot come up with a word in German, I make an educated guess based on Dutch or another German word." (t=-2.867, p=.005, d=.34)
3. "If I don't understand something in the source text, I don't mention it in my synthesis." (t=-4.417, p=.000, d=.39)
4. "I change the structure of my text while writing to improve coherence." (t=-2.668, p=.01, d=.59)

³⁸ The amount of available data for measures that needed introspection using the play back mode is considerably smaller than the purely quantitative measures (see limitations section). Therefore, statistical inference tests are less likely to reveal significant results.

³⁹ All strategy statements are translated from Dutch original

Product analysis

Product measures	pre	post	t	df*	p (2- tailed)	Cohen's d	Effect size
Content selection	2.847	3.379	-3.511	61	.001	-0.60006	medium
Structure	1.403	1.560	-2.468	61	.016	-0.38188	small
Overall coherence	1.734	1.827	-.960	61	.341	-0.14029	
Cohesive ties density Co/W	.1600	.1320	1.348	62	.182	0.279359	
Similarity rate	.0168517	.0368083	-8.348	58	.000	-0.98328	large

* missing data (n=76) due to students who were absent for either pre- or post-test

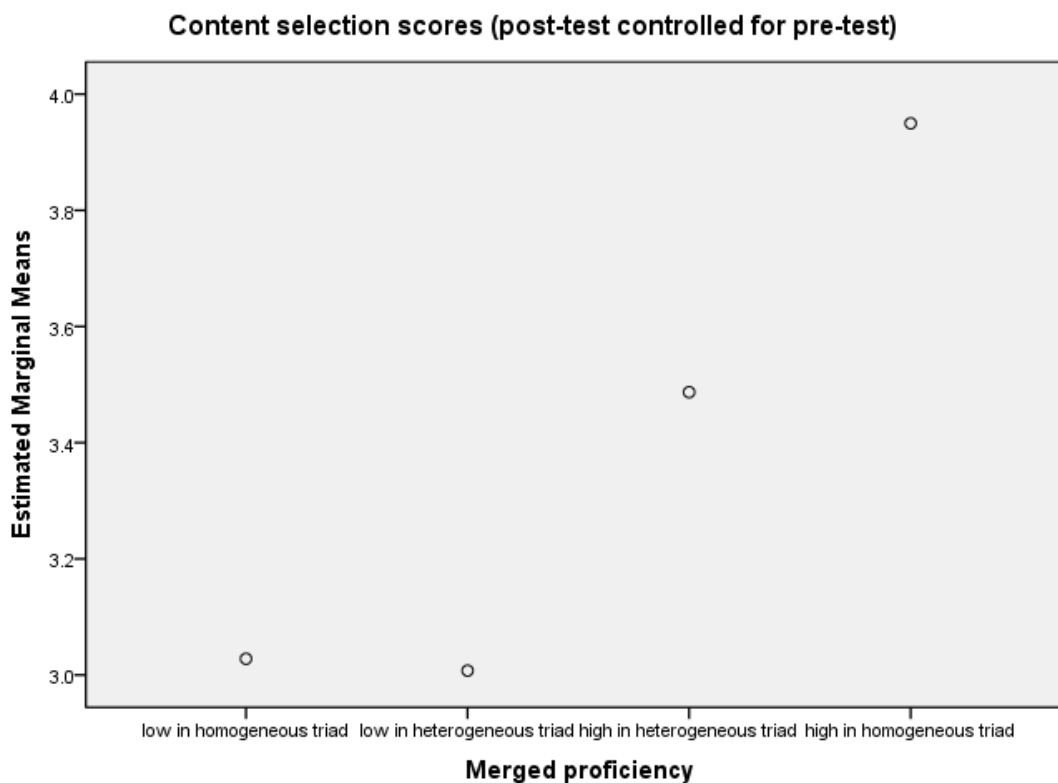
Table 6. Results of paired samples t-test for all dependent variables of writing product

Regarding the product analysis of the final texts, a significantly better content selection (medium effect size), as well as a better overall structure of the syntheses (small effect size) in the post-tests can be observed (see Table 6). The holistic measurement for overall coherence only displays a non-significant slight tendency towards improvement in the post-tests, and density of cohesive ties even decreased. The most noticeable and, at the same time, unexpected result, however, is a significant tendency towards more *patch-writing* behaviour in the post-test as measured in *Ferret's* similarity rate (large effect).

Regarding group composition according to proficiency, the ANCOVA analysis revealed a significant impact with a medium to large effect size on pre-post-difference regarding content selection ($F=5.456$, $p=.002$, $\eta^2=.229$), showing that the high-proficiency students in homogeneous groups improved most through the intervention (see Figure 5). Pairwise comparisons showed that there were significant differences between the conditions "low in homogeneous triad" and "high in homogeneous triad" (mean difference -0.922 , $p=0.016$), between "low in heterogeneous triad" and "high in homogeneous triad" (mean difference $-.942$, $p=0.004$). No significant differences were found between the other conditions (low in homogeneous vs. heterogeneous triads and high in homogeneous vs. heterogeneous triads). From these results, we can conclude that the performance of high-proficiency students regarding content selection increased significantly compared to low-proficiency students, independent of proficiency grouping. For the other measures, no significant impact was stated (see Table 7).

Product measures	F	p (2-tailed)	Partial Eta squared	Effect size
Content selection	5.456	.002	.229	medium
Structure	1.193	.321	.064	
Overall coherence	1.150	.337	.061	
Cohesive ties density	1.737	.170	.082	
Similarity rate	.426	.735	.023	

Table 7. Results of ANCOVA analysis for impact of proficiency grouping during collaboration on all dependent individual product variables (post-test scores controlled for pre-test scores)



Covariates appearing in the model are evaluated at the following values: content.0: holistic score content = 2.847

Figure 5. Estimated marginal means of the content selection (post-test scores controlled for pre-test scores) according to proficiency

Discussion

From the results, we can conclude that the most significant change in writing process can be observed in fluency. Students learnt to manage their time in order to fulfil the complex task of writing a synthesis in an L2 based on two popular scientific texts within the relatively short time frame of 60 minutes. They did so by spending less time reading the sources, and employing more efficient strategies for planning by writing bulleted summaries instead of full-texts. This beneficial strategy was clearly transferred from the collaborative setting where students, due to the jigsaw principle, had to inform each other about the content of their source texts. During the course of the collaboration, the majority of the triads adopted the bulleted form of summaries, as suggested in the model and the script (Strobl, 2015). Although students did not produce as many (full) outlines before the actual writing phase, seemingly, this did not impact adversely on content selection and structure, as the scores for both product variables increased. A tentative interpretation is that students learnt to write from (not-ordered) bulleted summaries, deciding on their relevance and logical order on-line (i.e., during composition). The higher fluency in the post-tests is also reflected in the self-reported changes in writing strategies.

Regarding the self-reported strategies development, it is interesting to note that the first three of the strategies that show significant increases in the post-hoc questionnaire can be related to writing speed in the translation phase, striving for maximum efficiency. Strategies 2 and 3 indicate a tendency towards adopting quick solutions (less content elaboration effort in case of understanding difficulties and less effort to use auxiliary means to look up vocabulary). Curiously, students reported to write more "in one flow" (strategy 1), whereas the observed process analysis points to the converse, revealing a slight tendency towards more recursive writing. While this seems contradictory at first sight, it might be rooted in a diverging interpretation of the statement (fluency vs. linearity). This explanation becomes more plausible if one considers that, at the same time, students also reported that they reorganised the structure of their text during writing (strategy 4) more.

Concerning the interpretation of the '*patch-writing* conundrum', a caveat is in order. The significantly higher rate of textual borrowing could originate in the source texts used in the pre- and the post-test. Although we strived for a comparable degree of linguistic difficulty and of content complexity, the source texts of the pre-test might lend themselves less to literal copying because one is written from a first-person's perspective, and the other contains several examples. Another plausible explanation is related to the above-stated increase in process fluency. Students 'learnt' through practising during the intervention that *patch-writing* saves time. The time limit, together with the desire to hand in finished texts, might have induced increased copy-pasting from sources as kind of trade-off. Indeed, trade-off strategies seem to play an important

role in the (unconscious) decisions about the focus in writing made by novice writers. In this sense, Kellogg stated that

students apparently know how to revise globally as well as locally. But they typically do not do so in their college writing assignments to avoid shortchanging the time and effort devoted to other necessary processes and representations during composition and subsequent revision (Kellogg, 2008, p. 8)

McDonough, Crawford, and De Vleeschauwer (2014) reported similar findings, i.e., increased use of textual borrowing in post-test L2 summaries after classroom-based instruction in content organisation and paraphrasing at a Thai University. However, they revealed through a fine-grained analysis that the overall increase of similarity rate was mainly caused by increased copying of shorter phrases (2-4 words), whereas the copying of longer phrases (5-8 words) actually decreased. McDonough et al. (2014) interpreted this selective copy-paste behaviour as a strategy to avoid textual borrowing, as it coincided with a higher rate of explicit reference. They suggest that

the path toward successful paraphrasing may not be a simple reduction in the overall quantity of copied strings, but may involve intermediary stages in which the overall number of copied strings remains constant or even increases, but the length of copied strings decreases. (McDonough et al., 2014, p. 27)

Concerning implications for L2 summarisation instruction, the opinions are divided. At one side of the spectrum, researchers-pedagogues stress the need to sensitise students to the danger of textual borrowing, investigating the impact of explicit teaching of rephrasing strategies (Lantolf, 2000; McDonough et al., 2014; Wette, 2010). This view is contested by others, including Seidlhofer, who argues that

the request [of L2 learners to use their own words in summarising tasks] is in urgent need of deconstruction: it seems to me to be paradoxical, even perverse, to ask language learners/users to summarise strictly someone else's ideas in strictly their own words. Particularly for foreign language students, the requirement appears to be tantamount to adding insult to injury, and to fly in the face of a methodology whose primary aim is, after all, to get learners to accept, adopt, use the vocabulary presented to them (Seidlhofer, 1995, p. 155).

In the same vein, Pecorari (2013) supports the idea that *patch-writing* should not necessarily be banned from L2 academic writing training, but could actually be used as a stepping stone towards the appropriation of target-like structures. She argues that paraphrasing requires sufficient linguistic resources to find an equivalent formulation and the writer's confidence in his/her ability to do so. Especially L2 writers often lack this confidence, and therefore (should be allowed to) rely on the language of their sources to a certain degree.

Although *patch-writing* might indeed have some potential for language learning, it implies the danger of poor elaboration. As the lack of linguistic elaboration often coincides with a lack of content elaboration, allowing students to *patchwrite* might encourage shallow processing of source text content. Therefore, rephrasing should be trained and encouraged in L2 academic writing, even if this means (temporarily) accepting stages where non-target-like structures occur as part of an ‘academic interlanguage system’.

To sum up, we can state that the answer to RQ1 is positive. There are, indeed, favourable changes in individual planning strategies from pre- to post-test after the collaborative writing experience. However, the interpretation of the results in the light of RQ2, i.e., whether a development in process is measurable in the product, is not that straightforward. Although there is a positive development regarding product quality concerning content selection and structure, our data do not show a direct relationship with the collaborative intervention (see section ‘limitations’). The change in quality might just as well point to a ‘learning-by-doing’ effect. However, in the collaborative setting, the triads were forced to elaborate their source texts in order to sustain mutual understanding, and jointly decide on the relevance of content for the common synthesis. A qualitative inspection into the collaborative writing processes revealed that these processes of joint information processing led indeed to the desired deep content elaboration, especially in homogeneous high-proficient groups (Strobl, 2015). The better progress of students belonging to those triads in individual content selection after collaboration that was unravelled in our data analysis makes a strong case for a causal relationship. Finally, we detected a significant impact of group composition according to proficiency (RQ3), indicating that high-proficiency learners in homogeneous groups showed the highest learning gain regarding content elaboration in their individual post-tests. Therefore, we have observed a similar impact of proficiency pairing as reported by previous scholars (Leeser, 2004; Ohta, 1995; Storch & Aldosari, 2013). These studies focused on lower-order concerns, i.e., attention to language in meaning-focused tasks shows that high-high pairs had a significantly higher number of (correctly resolved) LREs during collaboration. The present study provides evidence that it is also the students in high-homogeneous groups that improved most after a collaborative writing experience regarding a strategy related to higher-order concerns, being content elaboration.

Limitations and further research

The present study has some limitations regarding data collection. Due to technical problems, the keystroke logging files of approximately 30% of the pre-tests and 5% of the post-tests could not be used for graphical inspection nor play back of the writing process. This shortage of writing process data leads to a reduced statistical power, which might distort the possible impact of the proficiency and group composition on

process development. Secondly, the reliability of the holistic measure for cohesion turned out to be poor. A clear separation of coherence judgement from content, structure, and to some extent also text length, proved to be difficult, if not impossible, to the extent that some discrepancies between coders could not be resolved.

As for the source texts used in pre- and post-tests, they covered the same topic area (linguistic variations in contemporary German) and were chosen carefully, taking into account linguistic and cognitive levels of difficulty. However, as texts are never fully comparable, a cross-over design with a random distribution of texts over pre- and post-test could help rule out an effect of source text choice on post-test results. Last but not least, an experimental design with control groups could help research the specific differential impact of collaborative writing as compared to students who received individual training. However, given the ethical restrictions in the authentic instructional setting of this study, this option was not feasible. Therefore, we considered it to be more valuable to manipulate conditions within the collaborative set-up and to look at between-groups and within groups comparisons by adding group composition by proficiency level to our design.

To conclude, it can be stated that the collaborative writing experience in which students were supported by a script and a model had a positive bearing on their individual planning and, to a lesser extent, on their reviewing strategies. Furthermore, the post-tests showed better content selection and structure than the pre-tests. However, the intervention apparently did not help students to find their own voice in this highly intertextual task. The ultimate reason for their increased patch-writing behaviour could not be pinpointed with the data collected in this study. It is possible that the collaborative setting in this regard might even have had a counterproductive effect. After all, matching three different voices into one coherent writing style is far from evident. Qualitative analysis of the collaborative writing processes, including a thorough investigation of text genesis and the communication in the triads, is needed to unravel whether, and how, the intervention setting adversely affected students' mental task representation in this respect. In this sense, a cooperative task setting might be a valuable alternative, combining collaboration that proved beneficial for deep content elaboration in the planning phase and for linguistic elaboration in the reviewing phase (Strobl, 2015) with individual writing of the actual synthesis. Furthermore, to foster students' awareness of their textual misappropriation behaviour, it could be helpful to visually confront them with the output that the *Ferret* tool generates, based on their own synthesis, and to ask them to rephrase the highlighted text.

The growing scholarly interest in patch-writing as a typical problem of L2 academic writing underlines the need for instruction. (Students') writing nowadays occurs predominantly in digital environments with a constant connection to online (re)sources. This lack of physical boundaries between texts might contribute to a lack of

awareness of textual misappropriation. Digital writing has, to a large extent, influenced the way we write, and, therefore, also impacts on individual writing development.

Acknowledgement

We would like to thank Petra Campe for her kind cooperation as a second coder for the holistic product analysis unit regarding overall coherence.

References

- Blin, F., & Appel, C. (2011). Computer Supported Collaborative Writing in Practice: An Activity Theoretical Study. *CALICO Journal*, 28(2), 473-497.
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in Writing: Generating Text in L1 and L2. *Written Communication*, 18(1), 80-98. doi: 10.1177/0741088301018001004.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Couzijn, M. (1999). Learning to write by observation of writing and reading processes: effects on learning and transfer. *Learning and Instruction*, 9(2), 109-142. doi: 10.1016/S0959-4752(98)00040-1.
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18(2), 119-135. doi: 10.1016/j.jslw.2009.02.002.
- Davoli, P., Monari, M., & Severinson Eklundh, K. (2009). Peer activities on Web-learning platforms—Impact on collaborative writing and usability issues. *Education and Information Technologies*, 14(3), 229-254. doi: 10.1007/s10639-008-9080-x.
- De la Colina, A. A., & García Mayo, M. d. P. (2007). Attention to Form Across Collaborative Tasks by Low-Proficiency Learners in an EFL Setting. In M. d. P. García Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 91-116). Clevedon: Multilingual Matters Ltd.
- Dillenbourg, P., & Jermann, P. (2007). Designing Integrative Scripts. In F. Fischer, I. Kollar, H. Mandl & J. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (Vol. 6, pp. 275-301): Springer US.
- Doolittle, P. E., Hicks, D., & Triplett, C. F. (2006). Reciprocal teaching for reading comprehension in Higher Education: a strategy for fostering the deeper understanding of texts. *International Journal of Teaching and Learning in Higher Education*, 17(2), 106-118.
- Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, 21(1), 40-58. doi: 10.1016/j.jslw.2011.12.002.
- Galbraith, D. (2009). Cognitive models of writing. *German as a foreign language*(2-3), 7-22.
- Graham, S., & Perin, D. (2007). A Meta-Analysis of Writing Instruction for Adolescent Students. *Journal of Educational Psychology*, 99(3), 445-476.
- Guerrero, M. C. M. d., & Villamil, O. S. (2000). Activating the ZPD: Mutual Scaffolding in L2 Peer Revision. *The Modern Language Journal*, 84(1), 51-68.

- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hämäläinen, R., & Arvaja, M. (2009). Scripted Collaboration and Group-Based Variations in a Higher Education CSCL Context. *Scandinavian Journal of Educational Research*, 53(1), 1-16. doi: 10.1080/00313830802628281.
- Hanjani, A. M., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System*, 44(0), 101-114. doi: 10.1016/j.system.2014.03.004.
- Hayes, J. R., & Flower, L. (1980). Identifying the organization of the writing process. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.
- Hinkel, E. (Ed.). (2002). *Second Language Writers' Text. Linguistic and Rhetorical Features*. Mahwah, N.J.: Erlbaum.
- Howard, R. M., Serviss, T., & Rodrigue, T. K. (2010). Writing from sources, writing from sentences. *Writing and Pedagogy*, 2(2), 177-192.
- Johnson, D. W., Johnson, R. T., & Holubec, E. (1998). *Cooperation in the classroom*. Boston: Allyn and Bacon.
- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25(0), 4-22. doi: 10.1016/j.jslw.2014.05.005.
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1), 1-26.
- Kim, Y. (2008). The Contribution of Collaborative and Individual Tasks to the Acquisition of L2 Vocabulary. *The Modern Language Journal*, 92(1), 114-130. doi: 10.2307/25172996.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, 85(5), 363-394.
- Kirschner, P. A., Martens, R. L., & Strijbos, J.-W. (2004). CSCL in higher education? A framework for designing multiple collaborative environments. In J.-W. Strijbos, P. Kirschner & R. L. Martens (Eds.), *What We Know About CSCL: And Implementing It In Higher Education* (pp. 3-30). Dordrecht: Kluwer Academic Publishers Group.
- Kuteeva, M. (2011). Wikis and academic writing: Changing the writer-reader relationship. *English for Specific Purposes*, 30(1), 44-57. doi: 10.1016/j.esp.2010.04.007.
- Lantolf, J. P. (Ed.). (2000). *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.
- Leeser, M. J. (2004). Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research*, 8(1), 55-81. doi: 10.1191/1362168804lr1340a.
- Lei, X. (2009). *Understanding writing strategy use from a sociocultural perspective: a multiple-case study of Chinese EFL learners of different writing abilities*. (Unpublished doctoral dissertation), University of Hong Kong, Hong Kong.
- 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.
- Lund, A. (2008). Wikis: a collective approach to language production. *ReCALL*, 20(01), 35-54. doi: 10.1017/S0958344008000414.

- Lyon, C., Barrett, R., & Malcolm, J. (2004). *A theoretical basis to the automated detection of copying between texts, and its practical implementation in the Ferret plagiarism and collusion detector*. Paper presented at the Plagiarism: Prevention, Practice and Policies.
- Manchón, R. (2012). Multiple perspectives in the study of L2 writing development: An introduction. In R. Manchón (Ed.), *L2 writing development: Multiple perspectives*. (pp. 1-16). Boston / Berlin: Walter de Gruyter.
- McDonough, K., Crawford, W. J., & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context. *Journal of Second Language Writing*, 24(0), 20-32. doi: 10.1016/j.jslw.2014.03.001.
- Nicolás-Conesa, F., Roca de Larios, J., & Coyle, Y. (2014). Development of EFL students' mental models of writing and their effects on performance. *Journal of Second Language Writing*, 24(0), 1-19. doi: 10.1016/j.jslw.2014.02.004.
- Ohta, A. S. (1995). Applying Sociocultural Theory to an Analysis of Learner Discourse: Learner-Learner Collaborative Interaction in the Zone of Proximal Development. *Issues in Applied Linguistics*, 6(2).
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. (Vol. 1). New York: Newbury House.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265-289. doi: 10.1016/s1060-3743(99)80117-9.
- Pecorari, D. (2013). *Teaching to avoid plagiarism. How to promote good source use*. Berkshire: Open University Press.
- Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561-587. doi: 10.1177/0265532209340192.
- Raedts, M., Rijlaarsdam, G., van Waes, L., & Daems, F. (2007). Observational learning through video-based models: impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In S. Hidi & P. Boscolo (Eds.), *Writing and motivation* (pp. 219-238). New York: Springer.
- Rummel, N., & Spada, H. (2007). Can People Learn Computer-Mediated Collaboration by Following a Script? . In F. Fischer, I. Kollar, H. Mandl & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 39-55). Boston, MA: Springer Science+Business Media.
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4(1), 69-92. doi: 10.1007/s11412-008-9054-4.
- Schellens, T., van Keer, H., Valcke, M., & De Wever, B. (2007). Learning in asynchronous discussion groups: a multilevel approach to study the influence of student, group and task characteristics. *Behaviour & Information Technology*, 26(1), 55-71. doi: 10.1080/01449290600811578.
- Segev-Miller, R. (2004). Writing from Sources: The Effect of Explicit Instruction on College Students' Processes and Products. *L1-Educational Studies in Language and Literature*, 4(1), 5-33. doi: 10.1023/B:ESLL.0000033847.00732.af.

- Seidlhofer, B. (1995). *Approaches to summarization. Discourse analysis and language education*. Tübingen: Gunther Narr Verlag.
- Shi, L. (2012). Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing*, 21(2), 134-148. doi: 10.1016/j.jslw.2012.03.003.
- Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating Information: An Analysis of the Processes Involved and the Products Generated in a Written Synthesis Task. *Written Communication*, 30(1), 63-90. doi: 10.1177/0741088312466532.
- Spivey, N. N. (1990). Transforming Texts Constructive Processes in Reading and Writing. *Written Communication*, 7(2), 256-287.
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Bristol: Multilingual Matters.
- Storch, N., & Aldosari, A. (2013). Pairing learners in pair work activity. *Language Teaching Research*, 17(1), 31-48. doi: 10.1177/1362168812457530.
- Strobl, C. (2015). Learning to think and write together: Collaborative synthesis writing supported by a script and a video-based model. In T. Guasch & M. Deane (Eds.), *Learning and teaching writing online: Strategies for success* (pp. 69-95). Leiden: BRILL.
- Swain, M. (1997). Collaborative dialogue: Its contribution to second language learning. *Revista Canaria de Estudios Ingleses*, 34(115-32).
- Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 171.
- Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67-80). New York: Guilford Press.
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20(4), 316-327. doi: 10.1016/j.learninstruc.2009.08.009.
- Wang, S.-L., & Lin, S. S. J. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning. *Computers in Human Behavior*, 23(5), 2256-2268. doi: 10.1016/j.chb.2006.03.005.
- Wengelin, Å. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging and writing: methods and applications* (pp. 107-130). Oxford: Emerald.
- Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing*, 19(3), 158-177. doi: 10.1016/j.jslw.2010.06.002.
- Wigglesworth, G., & Storch, N. (2012). Feedback and writing development through collaboration: A socio-cultural approach. In R. Manchón (Ed.), *L2 writing development: Multiple perspectives*. (pp. 69-97). Boston / Berlin: Walter de Gruyter.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: measures of fluency, accuracy, & complexity*. Hawaii: Univ. of Hawaii Press.
- Wolfersberger, M. (2007). *Second Language Writing From Sources: An Ethnographic Study of an Argument Essay Task*. (phd), University of Auckland. Retrieved from

https://researchspace.auckland.ac.nz/bitstream/handle/2292/2364/02whole.pdf?sequence=19&origin=publication_detail.

Zhang, C. (2013). Effect of instruction on ESL students' synthesis writing. *Journal of Second Language Writing*, 22(1), 51-67. doi: 10.1016/j.jslw.2012.12.001.

Appendices

Appendix A

Script used to structure task and workflow during collaborative writing intervention, translated from Dutch original. The six recommended work steps include individual (ind) and collaborative (coll) steps.

Work step	What to do	Where to do it	Approx. starting t	Approx. duration
1 (ind)	Read your source text and summarise the main thoughts in bullets. Use your own words as much as possible. Mark literal textual borrowing as quotation and add the source (who said / wrote it?) Copy-paste your summary to the common Google Doc.	MS Word -> Copy-paste to Google Doc	10h40	15 min
2 (coll)	Read all summaries and discuss the content. Don't focus too much to language right now.	Comment in Google Doc	10h55	15 min
3 (coll)	Select the main propositions for your joint synthesis . Compose a text skeleton ¹ . Distribute the text to be written into three parts.	Chat + comment in Google Doc	11h10	15 min
4 (ind)	Write a first version of your synthesis part. Respect the word limit.	Word -> Google Doc	11h25	15 min
5 (coll)	Carefully read all parts of the synthesis. Discuss in group: Is everything clear and correct? Do the text parts match with each other? Does the ensemble represent a coherent text that reads fluently? Add transition passages and delete repetitions. Change the proposition order if appropriate. Focus on content and language alike.	Comment and text in Google Doc	11h40	20 min
6 (coll)	Re-read your synthesis a last time and edit until you are all satisfied with the result. Respect the word limit.	Google Doc	12u00	10 min

¹ Skeleton = bulleted outline enriched with cohesive ties

Appendix B

Total variance and variance distribution between the three levels occasion, student, and triad. Exemplary interpretation for one variable (content selection): The total variance of the pre-test scores for content selection was 1.027, which was the sum of the between-triads (Level 3) variance (.317) and the within-triads between-students (Level 2) variance (.71). After calculating the ICC, which revealed the correlation of the observations (cases) within each cluster on the different levels, we can state that 30.87% of total pre-test variance lay at the triad level but was not significant and the proportion of variance due to differences between individual students was 69.13% (significant). With regard to the learning gain, the total variance for this variable was 1.502, i.e., the sum of the between-triads (Level 3) variance (.319) and the within-triads between-students (Level 2) variance (1.183). ICC calculation revealed that the learning gain significantly varied on the individual students' level and accounted for 78.76% of the total variance, whereas the variance at the triad level (21.23%) again was not significant.

Variable		Variance			
		Total	Stud. level	Triad level	
PROCES MEASURES	Characters/minute	Pre-test	241.5	100%*	0%
		Learning gain	219.6	91.46%*	8.54%
	Time reading sources	Pre-test	15.042	100%	0%
		Learning gain	20.217	80.08%	19.92%
	Summarising	Pre-test	2.515	100%	0%
		Learning gain	3.186	100%*	0%
	Outlining	Pre-test	.544	83.19%*	16.91%
		Learning gain	1.23	78.29%*	21.71%
	Type of revisions	Pre-test	.406	100%*	0%
		Learning gain	.423	100%*	0%
	Mean all strategies	Pre-test	.062	100%*	0%
		Learning gain	.075	100%*	0%
PRODUCT MEASURES	Content selection	Pre-test	1.027	69.13%*	30.87%
		Learning gain	1.502	78.76%*	21.23%
	Structure	Pre-test	.18	93.89%*	6.11%
		Learning gain	.23	100%*	0%
	Overall coherence	Pre-test	.433	75.98%*	24.02%*
		Learning gain	.57	100%*	0%
	Cohesive ties density	Pre-test	.023	100%*	0%
		Learning gain	.025	100%*	0%
	Ferret similarity rate	Pre-test	.15	94.76%*	5.24%
		Learning gain	.202	80.35%*	19.65%

Note: * indicates $p < .05$

5.5 Further results and conclusions from the study

5.5.1 Impact of the scaffolding condition on collaboration quality

5.5.1.1 Results of quantitative analysis

As explained in chapter 5.2, the quantitative investigation of the collaboration was subdivided into two aspects. Fruitful collaboration requires a) an effective workflow, and b) a sufficient amount of goal-oriented communication between the group members. Table 9 provides an overview of the variables that were created to analyse these two aspects.

Table 9 Variables of quantitative comparative collaboration analysis

Aspect	Variable	Measurement	Scale
Process	Overall workflow quality	task steps present in workflow	1...13
	Content elaboration	individual summaries	1...4
		collaborative outline	1...3
	Linguistic elaboration	similarity rate between sources and synthesis	0...1
	Efficiency	final status	0...2
Communication	Overall interaction intensity	number of all chat turns and comments	- 2.39...5.83 ⁴⁰
	Text-related interaction	all comments and text-related chat	-2.20...6.57
	Content-related interaction	content-related comments and chat	-2.49...4.98

⁴⁰ All variables for communication vary around 0 because they represent sums of standardised values

In the following, the results of the quantitative comparative analyses of both quality aspects of collaborative writing will be presented and discussed. The results of paired samples t-tests to measure within-condition, between-sessions effects, and independent samples t-tests to measure between-conditions effects per session, and for all sessions together, are represented in tables (see Appendix 17). For significant changes, the effect size was calculated as Cohen's D, applying Cohen's rules of thumb to label the effect size (Cohen, 1992). The abbreviations used in the tables are "Model1" (model-first condition), i.e., the condition where the first scaffold instrument applied in session 2 was the video model, followed by the script as second scaffold in session 3, "Script1" (script-first condition), with an accordingly changed order of scaffold instruments, and "occ1", "occ2", and "occ3" for the respective collaboration sessions (measure occasions). The data for both conditions and all occasions are visualised as boxplots. Significant differences are highlighted with double-sided arrows between the respective medians, and labelled with the respective t-value preceded by asterisks indicating the degree of significance. Where appropriate, the results of the quantitative analysis are elucidated by references to a qualitative inspection of the collaboration processes.

Process effectiveness

The analysis of the collaboration process is based on four variables, viz. a) overall workflow of the collaborative activity, b) content elaboration, as overt in two planning activities, viz. individual summarising of the source texts, and collaborative outlining for syntheses, c) degree of linguistic elaboration, measured indirectly by quantifying the amount of textual borrowing from the source texts, i.e., the similarity rate between the latter and the final synthesis, and finally, d) efficiency, as measured in the final status of the joint texts at the end of the session. In combining these four measures, the main objectives of the intervention regarding the collaborative process were tackled. To collaboratively produce an effective synthesis, a thorough joint planning on the content level is needed (variable b), and textual misappropriation must be avoided through rephrasing and cohesion building (variable c). The appropriate task steps must be overt in an *effective* workflow (variable a), which at the same time needs to be *efficient* in order to meet the task requirement, i.e., deliver a finished text within the allotted time (variable d).

Overall workflow of the collaborative activity

Based on the number and the sequence of task steps that are clearly distinguishable in the workflow (see chapter 5.2), and on the completeness of the process, a variable was created to express the overall quality of the collaborative workflow on a scale of 13 for each student (see mapping table in Appendix 16). Of course, the workflow scores for students within the same triad was similar, but as not all students participated to the

same degree in all task steps, it was deemed necessary to apply a score per student rather than per triad.

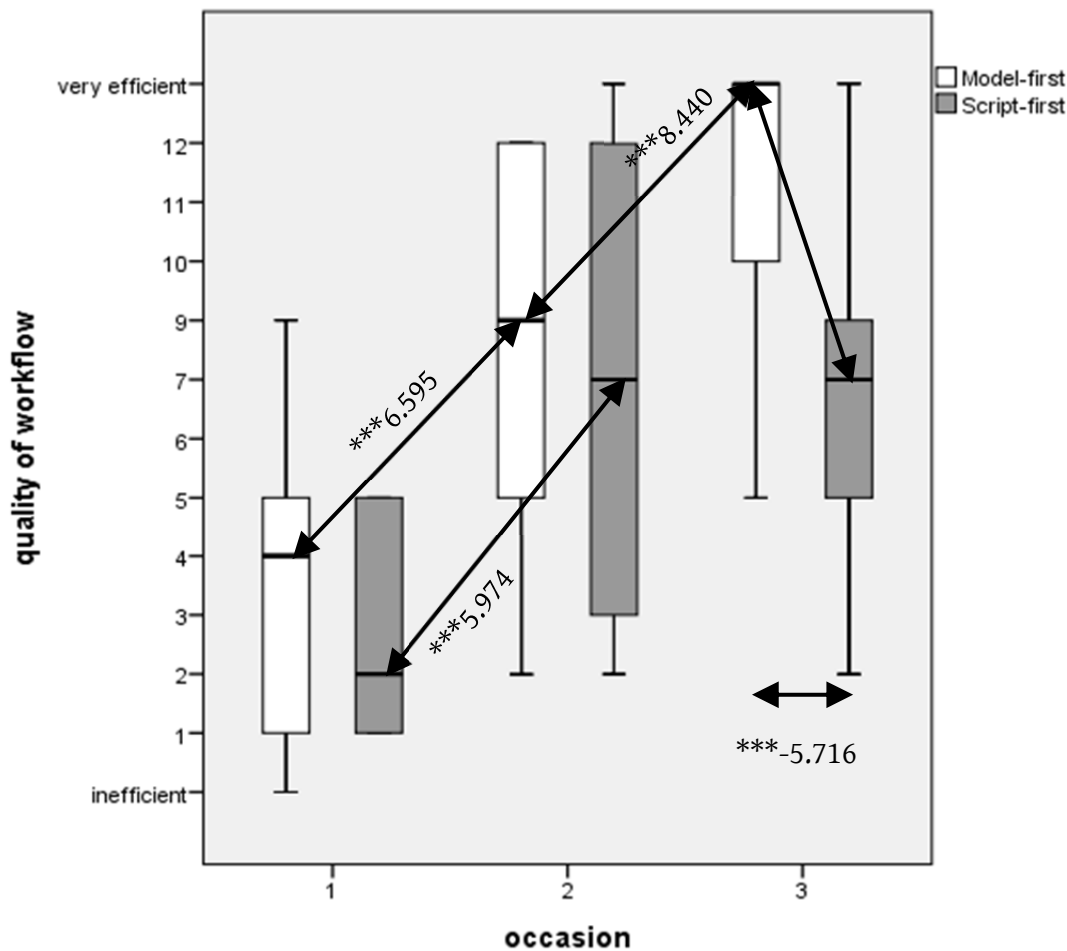


Figure 18 Workflow quality in IS3
 Note: *p < .05. **p < .01. ***p < .001

Figure 18 reveals that scaffolding had a clear impact on workflow, but that there were differences between the two groups. The overall workflow quality score was significantly higher for the triads in the model-first condition than in the script-first condition (see Appendix 17, Table 1). The discrepancy that was already present in the first, unscaffolded session, increased through scaffolding. This means that the model-first triads adhered to the task steps modelled in the video and recommended in the script. Their greater concern with workflow is also visible in the significantly higher number of chat turns related to workflow (see Figure 28). In the sequence "script after model", the script added to the process quality. Qualitative inspection revealed that the added value of the script is based on the time indications for the work steps that were provided, helping the triads to finish their joint text in time. On the contrary, the

sequence "model after script" did not seem to have the same incremental effect on workflow quality. Qualitative inspection revealed that the participants of this condition did not spend a lot of attention to the model video, to the extent that some resolved to fast forward it. Clearly, they did not perceive the need to watch the model after having already collaborated twice.

Content elaboration

The two overt planning activities for content elaboration, viz. summarising source texts and outlining the synthesis, were recommended in the F2F-session prior to the pre-tests, and re-emphasised in the scaffolding instruments. Best practice strategies for both activities were highlighted. Regarding summarisation of the source texts, it was recommended to produce bullets, if possible indicating their logic relationship in an outline, to facilitate an easy comparison of the source texts in view of content selection for the synthesis to be written. For the outlines of the synthesis, an "enriched" form of outline was suggested, by adding cohesive ties in order to indicate, again, the logic relations among the content items to be integrated. The quantitative results are represented in stacked bars to visualise the different forms of summarising and outlining employed by the students.

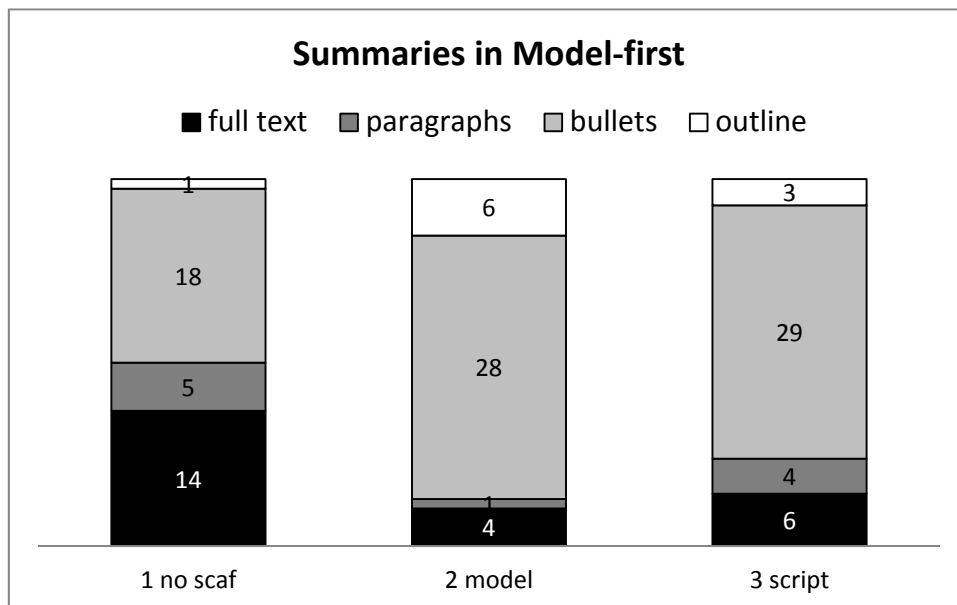


Figure 19 Form of individual summaries produced in the model-first group. The numbers in the stacked bars indicate the number of individual summaries produced in the respective form.

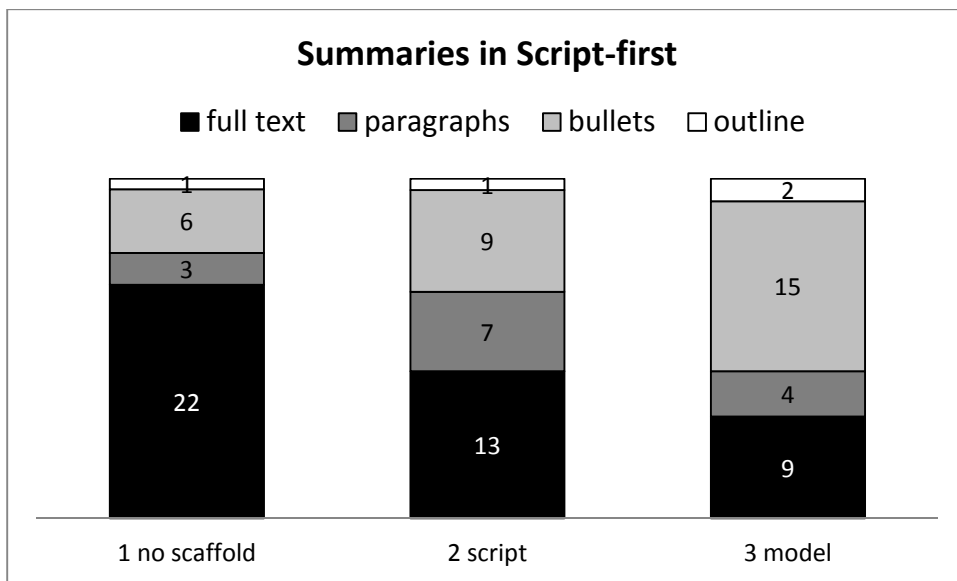


Figure 20 Form of individual summaries produced in the script-first group

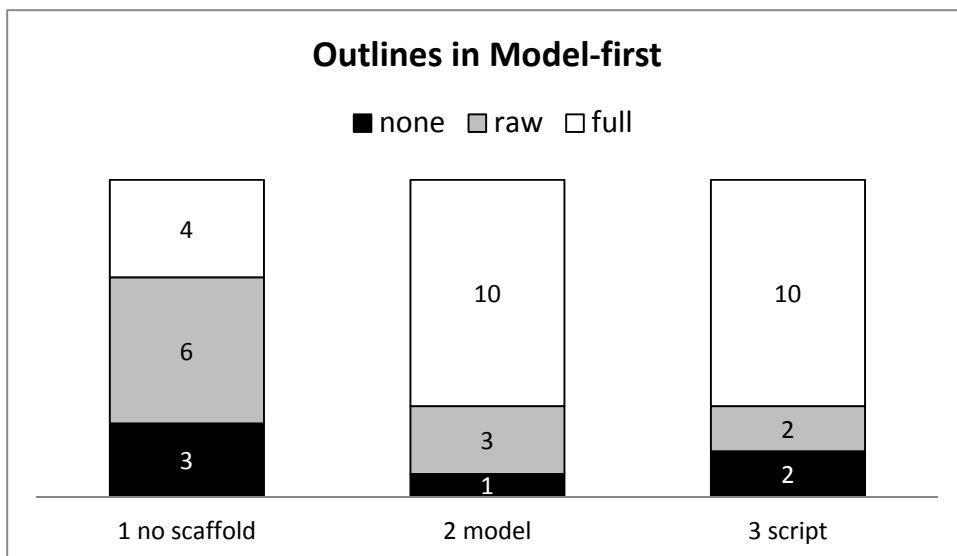


Figure 21 Form of collaborative outlines for syntheses in in model-first group. The numbers in the stacked bars indicate the number of collaborative outlines produced in the respective form.

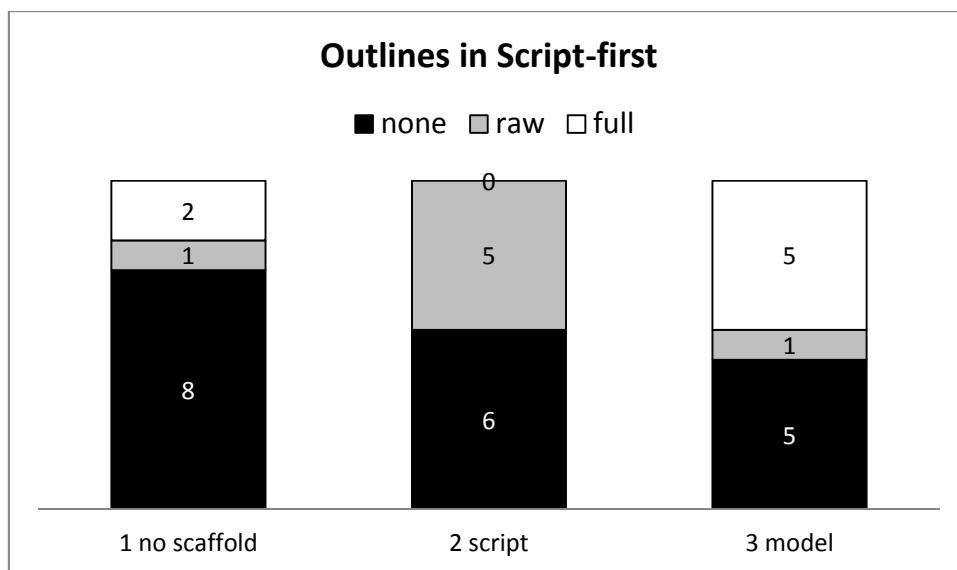


Figure 22 Form of collaborative outlines for syntheses in in script-first group

The results of these two units of analysis, viz. form of individual summaries and form of collaborative outlines, are related to the elaboration and restructuration of content and coincide to some degree with the findings for overall workflow quality. Students in the model-first condition adhered more to the planning strategies that were modelled and recommended, by producing (outlined) bullets or independent chunks rather than a fluent text, than the students in the script-first condition. They also wrote more often a full (i.e., enriched) outline of the synthesis before starting to produce it. Again, the differences between the two conditions that were already noticed in the first session, increased in the successive sessions with scaffolding. However, the model video seems to have influenced the students more than the script regarding the use of planning strategies: In both conditions, the most remarkable changes occurred in the "model session", as Figures 19 to 22 reveal. More specifically, regarding summarisation of source texts, after watching the model video, only five students of the model-first condition still produced text (fragments) instead of bullets (see Figure 19). In the script-first group, the amount of students who wrote bullets increased slightly in the scripted session, but their number still increased more after watching the model (see Figure 20). More or less the same pattern was observed in the strategies employed for outlining. To conclude, it can be stated that the model stimulated emulation of the observed planning strategies.

Linguistic elaboration

Beside content elaboration, one of the most difficult subtasks in synthesising from sources is linguistic elaboration. This term is used here as an umbrella term to define the linguistic changes that are required to avoid overly textual borrowing from source

texts, i.e., rephrasing, and to clarify the coherence of the new text by adding cohesive ties. As an indirect quantitative indication of linguistic elaboration, the similarity rate between the source texts and the synthesis was measured. To this aim, the plagiarism detection tool *Ferret* was used that calculates a similarity rate between two texts based on the amount of identic trigrams (Lyon, Barrett, & Malcolm, 2004), ranging from 0 (no similarity) to 1 (identical texts).

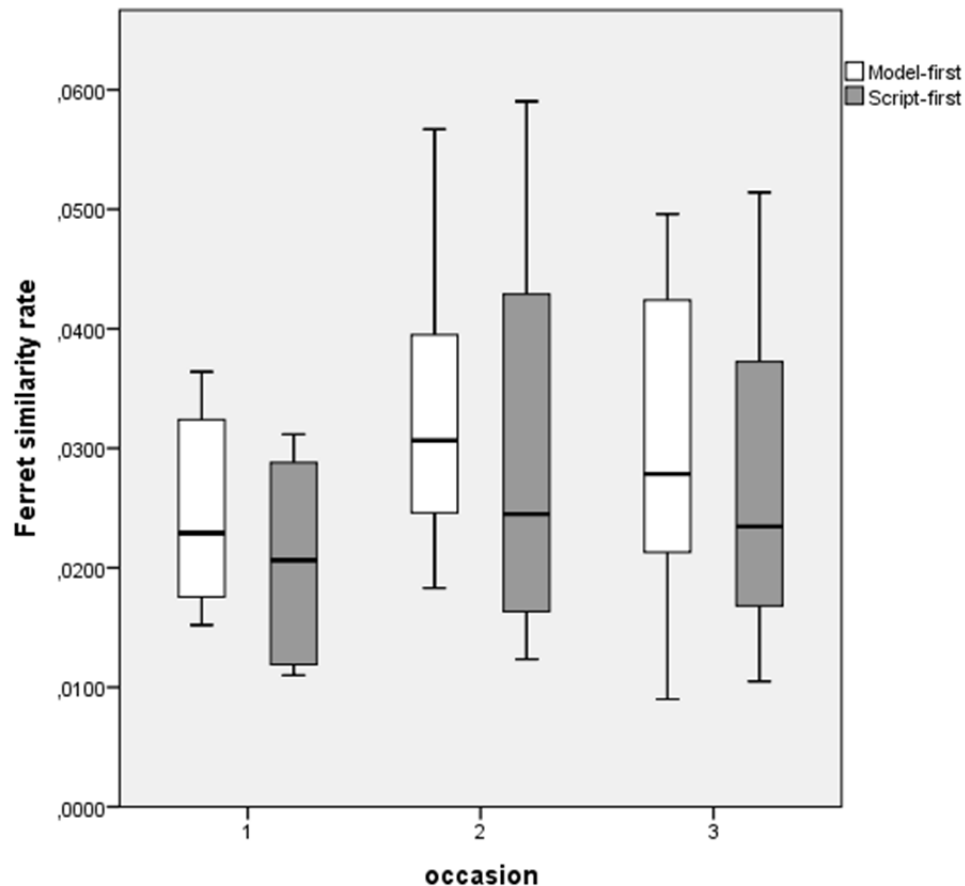


Figure 23 Ferret similarity rates in IS3

The results of t-tests revealed no significant differences between-conditions nor between-occasions. However, in both conditions, a tendency towards increased similarity rate, i.e., increase of textual borrowing from the source texts, was observed between session 1 and session 2 (first scaffold), followed by a slight decline between session 2 and session 3 (see Figure 23). Furthermore, the triads in the model-first condition displayed a higher rate of textual borrowing throughout the intervention than the triads in the script-first condition. Consequences of this finding will be discussed below (see 5.5.1.2).

Process efficiency

Process efficiency was operationalized as the ability of the triads to deliver a finished product in the allotted timeframe of approximately 90 minutes.

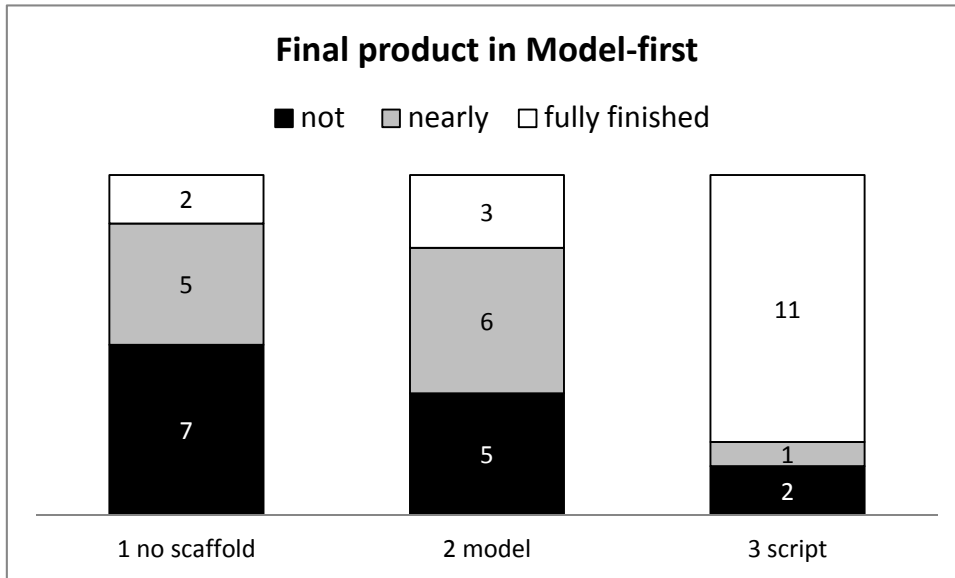


Figure 24 Status of final product of collaboration sessions in model-first group. The numbers in the stacked bars indicate the number of final products in the respective form.

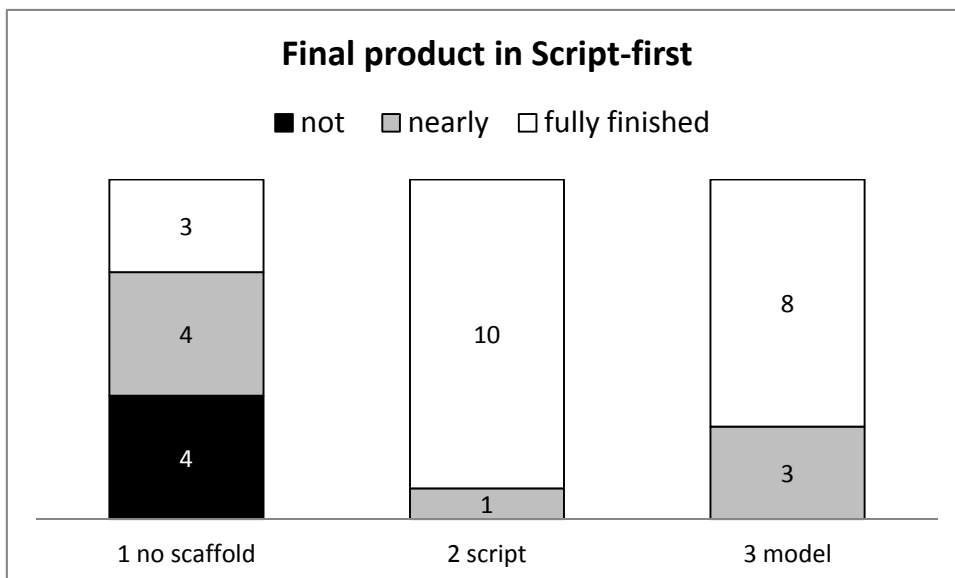


Figure 25 Status of final product of collaboration sessions in script-first group

Regarding process efficiency, significant differences were observed, both between-conditions, and within-conditions, between-occasions (see Appendix 17, Table 3). Overall, the triads in the script-first condition clearly displayed a higher process efficiency. All triads in this condition succeeded in (nearly) finishing their joint

syntheses in both scaffolded sessions, whereas in the model and the script sessions of the model-first condition, five and two triads respectively did not succeed in finishing their texts. This difference between conditions will be discussed below (see 5.5.1.2). In the model-first condition, there was also a significant growth in process efficiency throughout the intervention. The highest between-occasions effects in each group was observed in the sessions with script as a scaffold. This indicates that the script, containing time indications for the task steps, had a clear impact on process efficiency.

Communication

A *sine qua non* for good collaboration is good communication. This aspect was measured based on the three variables a) overall interaction intensity, b) text-related interaction, and c) content-related interaction. Google Docs provides two channels for communication, viz. a chat at the right hand side of the screen, which becomes visible as soon as at least two participants are logged into the document, and the comments space which is activated by clicking the respective icon in the toolbar (see screenshot in Figure 26). In the F2F instruction on the use of Google Docs prior to the collaboration, students were advised to use the chat rather for quick fixes and workflow issues because it "disappears", and therefore, no reference can be made to chat sequences written in an earlier stage of the collaboration. To discuss text-related issues, the use of the comment function was recommended, as comments can be linked to text fragments. In addition, they remain visible throughout the collaboration session until marked as "resolved", and reactions can be added in a later stage.

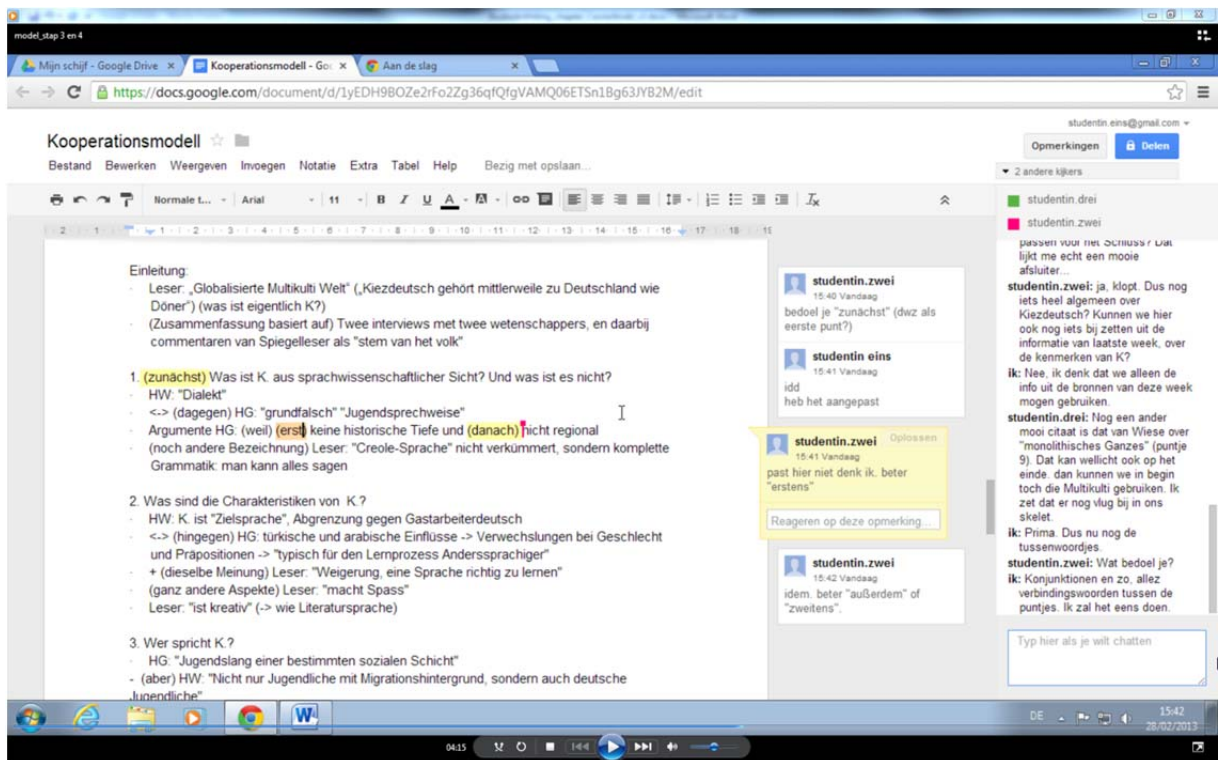


Figure 26 Screenshot of an interaction in Google Docs from the model video

To measure overall interaction intensity, the interactions in both communication channels were counted. To create a combined measure, both values first were standardised before they were added, as there were in total a lot more chat turns than comments. The chat turns were subdivided for further analysis according to their focus into the categories "workflow chat", "content chat", and "social chat". To calculate the variable "text-related interaction", the standardised values for amount of comments and for amount of content-related chat were added. The comments were further subdivided according to their main focus into the three categories LOC (e.g., issues related to spelling, grammar, lexical choice), HOC1 (content), and HOC2 (coherence and cohesion). LOC discussions appear "naturally" in L2 peer revision. Although they can have a positive impact on text quality (Benevento & Storch, 2011; Storch, 2005; Wigglesworth & Storch, 2009), in this specific intervention, the focus of interest lied on HOC discussions because they are decisive for the quality of syntheses. Therefore, in the video, HOC discussions were modelled, and in the script, special mention was made of HOCs in the reviewing step (see Appendix 21). It was expected that the amount of HOC discussions would increase in the sessions 2 and 3, as compared to the first, unscaffolded session. To verify this hypothesis, the variable "content-related communication" was created as a joint variable for standardised values of HOC1-comments and of content-related chat turns.

Overall interaction intensity

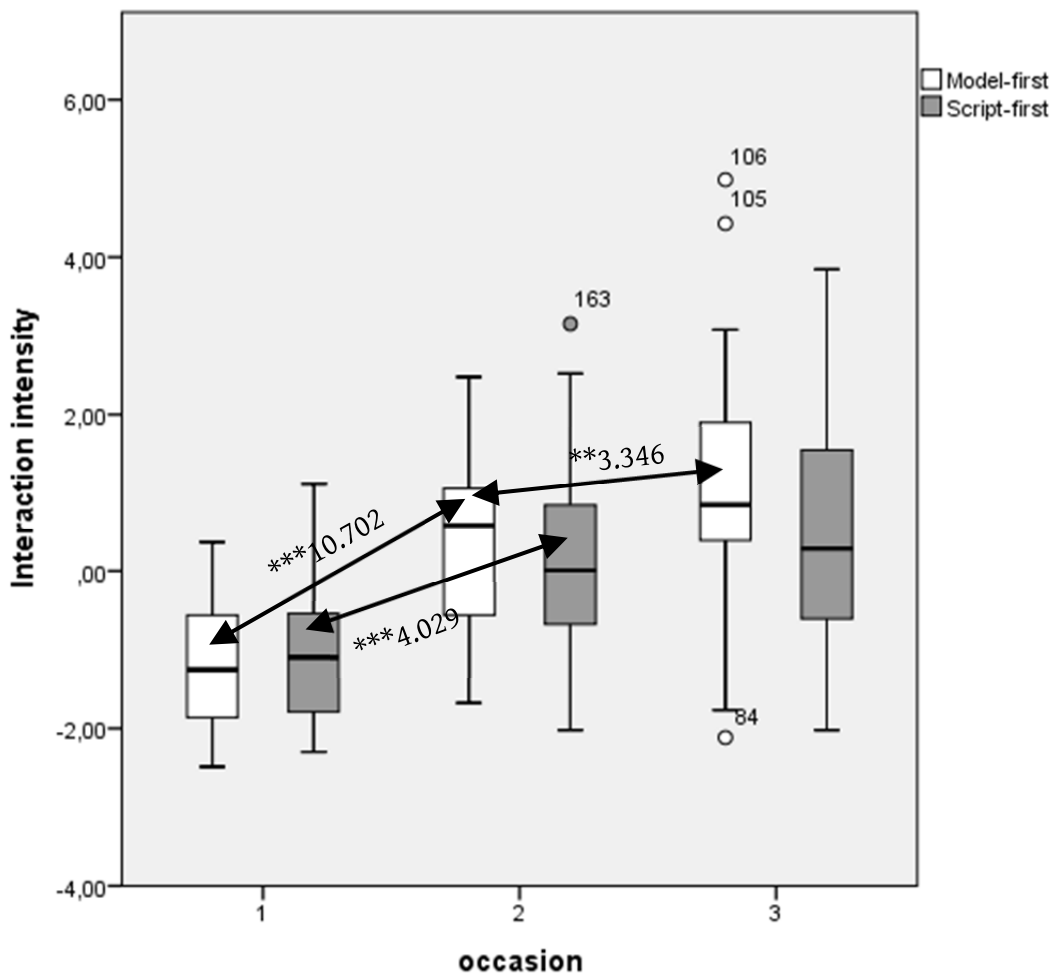


Figure 27 Overall interaction intensity in IS3

Note: **p < .01. ***p < .001

The data concerning overall interaction intensity revealed that the overall interaction intensity increased throughout the intervention in both conditions. The increase was significant between all occasions in the model-first triads, and between occasion 1 (no scaffold) and occasion 2 (script) in the script-first triads. There were no significant differences between-conditions regarding this variable. However, the predominant channel and the nature of the interaction differed between conditions, as the graphs in figures 28 and 29 visualise.

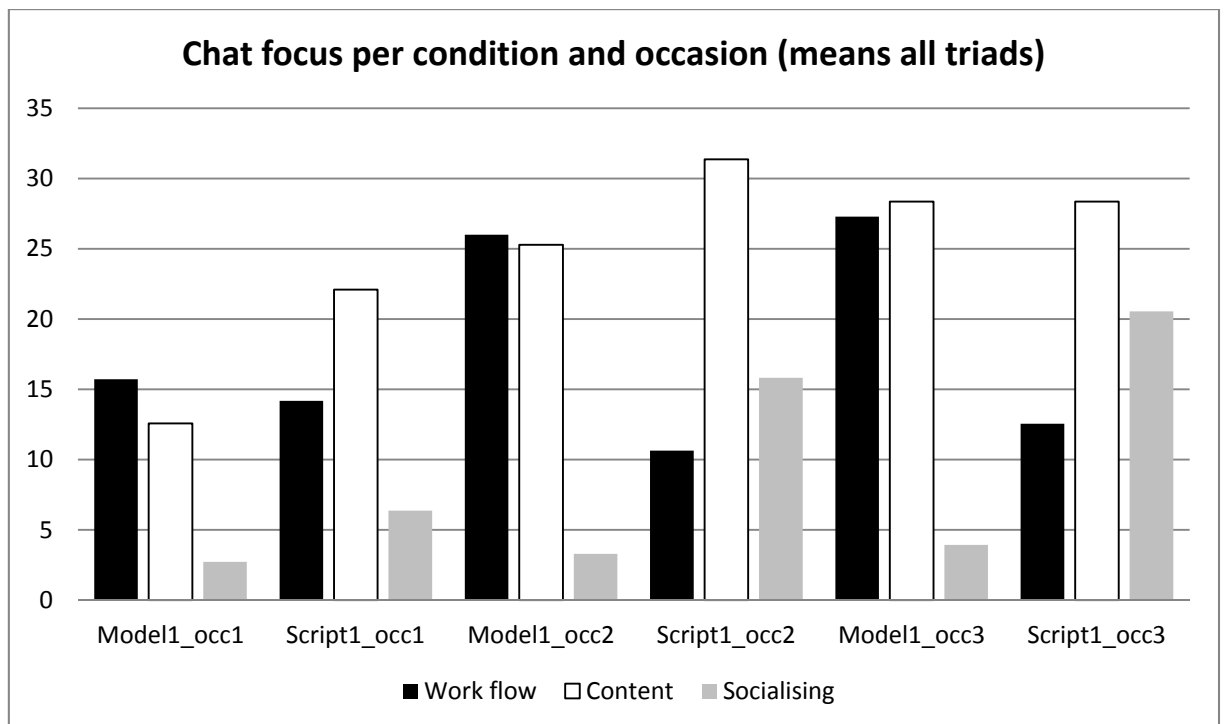


Figure 28 Number of chat turns according to focus in IS3

Regarding chat focus, it is interesting to note that the amount of workflow-related chat was noticeably higher than in the model-first condition than in the script-first condition. Furthermore, it increased throughout the sessions. This indicates the higher concern of the model-first triads with finding an effective collaborative process and coincides with the findings concerning overall workflow. Vice-versa, the script-first triads seemed to discover the chat as a socialising channel in the course of the intervention, as the respective values increased noticeably with each occasion, whereas they remained on a stable low level in the model-first condition. Both groups also made use of the chat to discuss content issues to a continuously high degree. Insights from qualitative analyses revealed that the chat was used for overarching content discussions that were not related to a specific text fragment.

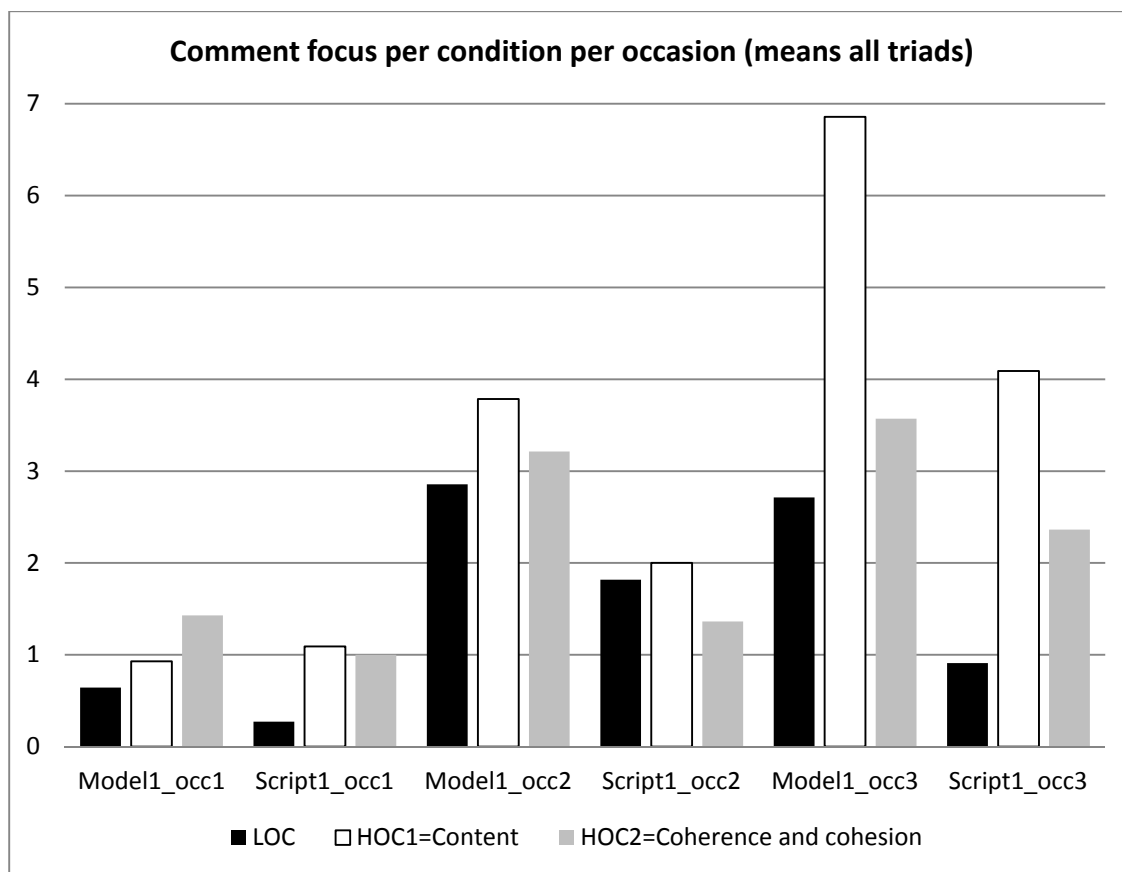


Figure 29 Number of comments according to focus in IS3

The mean number of comments posted is considerably lower than the chat turns. In the first, unscaffolded collaboration session, few triads made use of the comment function at all. The amount of comments increased considerably throughout the sessions in both conditions. It is especially interesting to note that the focus on HOC1 issues increased throughout the sessions, with a peak in the last collaborative session. In the same vein, but not to the same degree, HOC2-related comments increased continuously, whereas the amount of LOC-related comments increased only in the second session, remaining at a stable level (model-first condition) or decreasing considerably (script-first condition) in the last session. Qualitative inspection reveals that the lower mean number of comments in the script-first condition was caused by the fact that four triads in session 2 and two triads in session 3 did not use the comment function at all, and only communicated via the chat. Although the hypothesis that watching the *model* would instigate more HOC-related discussions cannot be verified through the results, it can be stated that HOC-related communication gained importance in the triads' interactions throughout the collaborative sessions.

Text-related communication intensity

The t-tests results regarding differences in text-related communication (see Appendix 17, Table 5), i.e., combining the values for content-related chat and all comments, confirmed the findings of the visual comparison presented in the last subchapter. Text-related communication increased significantly throughout the collaborative sessions in the model-first condition, and between session 1 and session 2 in the script-first session. Furthermore, the initial significant difference between both conditions in session 1 disappeared in the scaffolded sessions (see Figure 30).

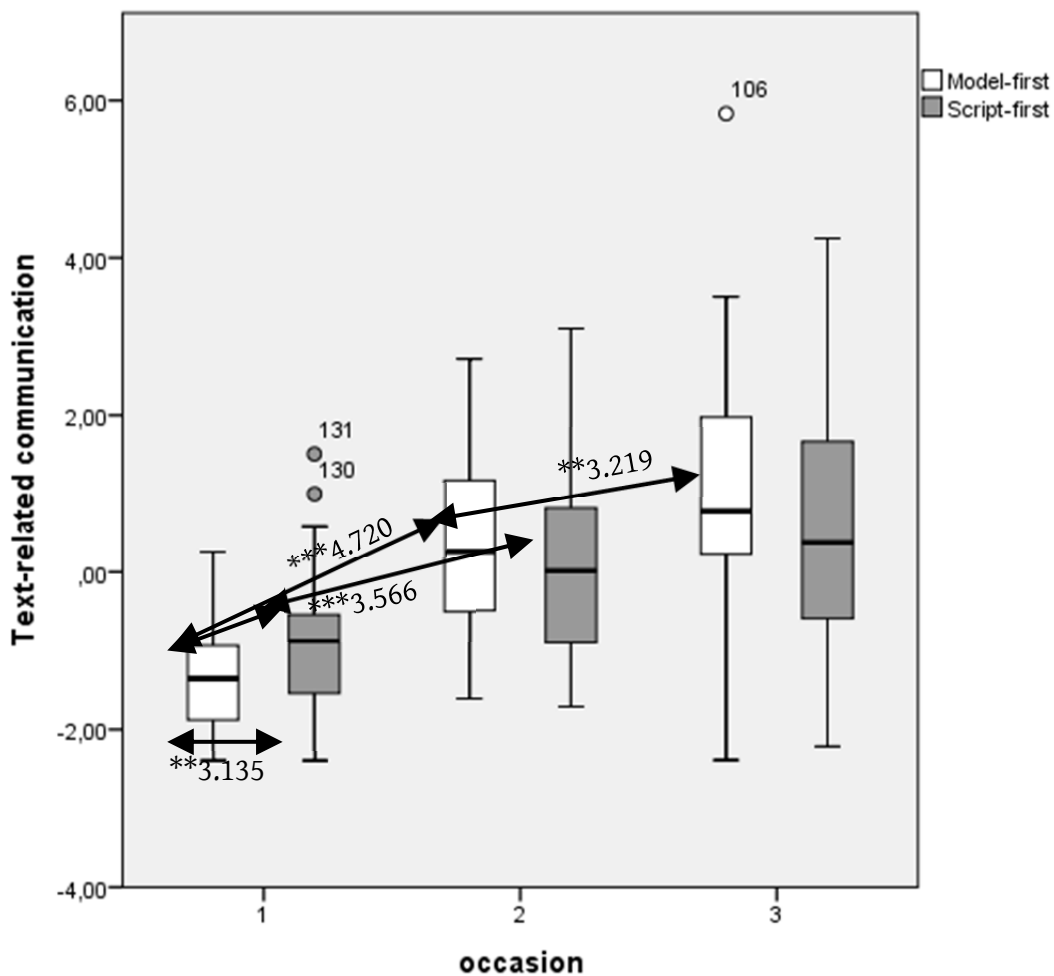


Figure 30 Text-related communication in IS3

Note: **p < .01. ***p < .001

Content-related communication intensity

In the same vein, the t-test results regarding differences in content-related communication (see Appendix 17, Table 6), i.e., combining the values for content-related chat and all HOC1-comments, confirmed the findings of the visual comparison

presented in the last subchapter (see Figures 28 and 29). They revealed a constant significant increase of content-related communication in the triads throughout all sessions in both conditions.

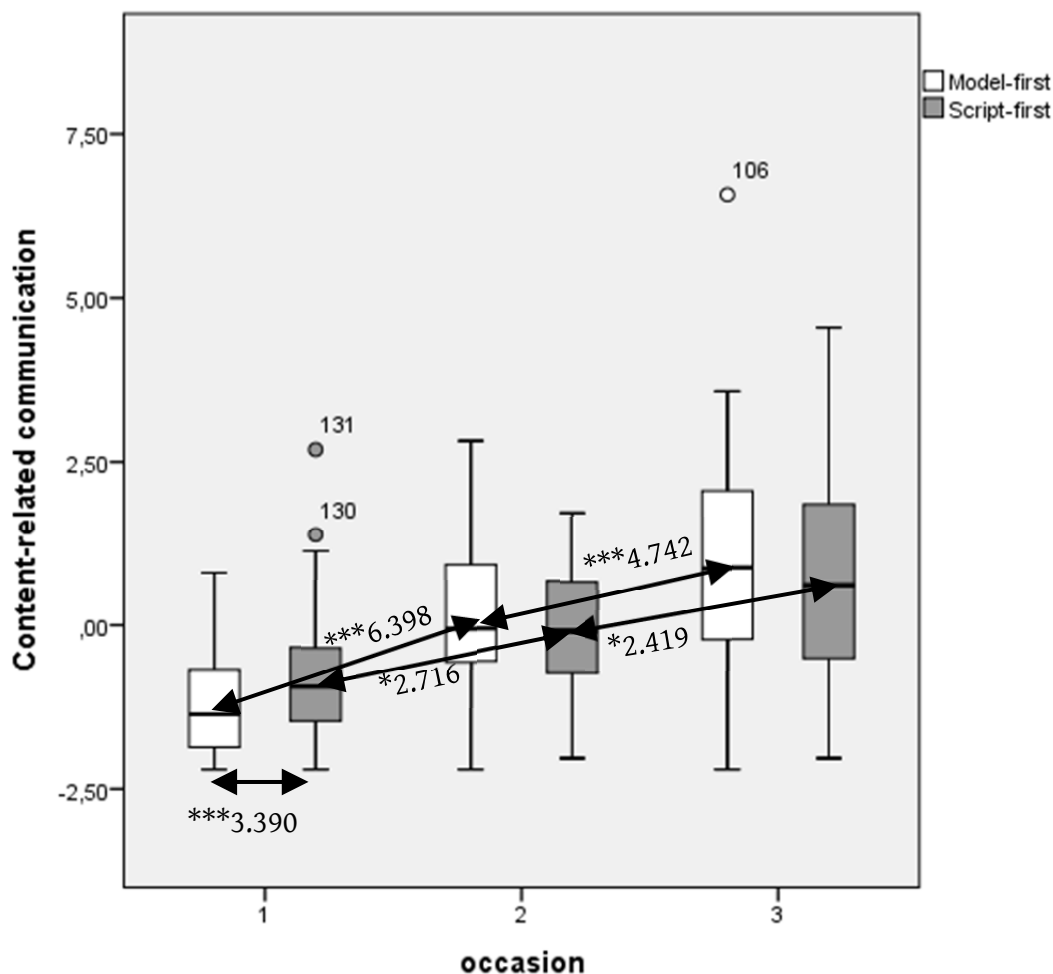


Figure 31 Content-related communication in IS3
 Note: *p < .05. **p < .01. ***p < .001

5.5.1.2 Conclusion and discussion

From the above analyses, the following conclusions can be drawn:

1. Regardless of the order of support instruments, scaffolding had a positive impact on effective work flow organization. The model stimulated the emulation of process steps, whereas the script served mainly as a fall back for timing.
2. Both overall and text-related interaction intensity increased significantly throughout the sessions, irrespective of condition and scaffolding instrument. Concerning the focus of communication, content-related discussions both in the chat and in the comments increased continuously, independent of the support

method used. This indicates that the students realised the importance of focusing on content in the targeted activity.

3. However, some interesting differences between the conditions regarding the participants' adherence to the scaffolding instruments were observed which resulted in differences regarding workflow and preferred communication channel. They will be discussed in the following.

A trade-off effect was observed between workflow *effectiveness*, i.e., including the appropriate steps for thorough content elaboration, and workflow *efficiency*, i.e., leading to the delivery of a finished product. Whereas the triads in the model-first condition displayed a more process-oriented workflow, adhering to the task steps recommended and modelled in the scaffolding instruments, the triads in the script-first condition seemed to be more product-oriented, establishing their own workflow by skipping time-consuming elaboration phases. This trend was already observable in the first, unscaffolded collaborative session. Most of the model-first triads applied the principles and strategies they had been given in a F2F session for individual synthesis production, whereas nearly all script-first triads seemed to neglect the input they had been given in the week prior to the intervention. A qualitative inspection of their workflow revealed that they basically produced three summaries of the source texts as fully written-out texts, pasted them together in the joint Google document, and adjusted the transition fragments to some extent. In this way, they succeeded in delivering a finished product, albeit not a synthesis of three source texts, but rather a combination of three summaries. Clearly, in their mental representation of the task, the finished product had priority over the process.

As the model-first triads could not finish their synthesis in the first, unscaffolded session, they might have felt a clearer need for scaffolding, leading them to cling to the instructions in the scaffold instruments. In this vein, the first learning-by-doing session provided the ground to investigate the impact of the scaffolding mechanisms on the collaboration processes, following Kirschner et al. (2004), who stated that

[o]nce a learning need becomes salient (perception), the educational affordances will not only invite but will also guide her/him to make use of a learning intervention to satisfy that need (action). The salience of the learning intervention may depend upon factors such as expectations, prior experiences, and/or focus of attention (Kirschner et al., 2004, p. 15)

Clearly, in the model-first group, this need was experienced, and therefore, the students took up the educational affordances offered by the model in the second session, and, to a less explicit extent, those offered by the script in the third session.

In the "script-first" condition, triads established an *efficient* workflow from the first, unscaffolded, session on. Therefore, they apparently did not feel the same need of being supported in the process. There were even some "refractory" triads (35% of the

participants in this condition) that did not change their work flow and collaboration intensity at all during the whole intervention. Elen (2013) coined the term "instructional disobedience" for this kind of unexpected behaviour in online learning.

Due to their higher degree of adherence to the task steps proposed or modelled in the scaffold instruments, the triads in the model-first condition had a more explicit planning phase, engaging in deep *content* elaboration of the source texts. However, this had no influence on their *linguistic* elaboration, i.e., the amount of textual borrowing. This indicates that there is no direct link between content elaboration and the linguistic elaboration (rephrasing of the content) that is necessary to avoid patch-writing. A possible explanation is that, after finishing content elaboration and after deciding on the structure of the synthesis, students in the model-first condition reverted to the original texts and copy-pasted them anyway, regardless of the rephrasing that had been going on while discussing the content in their L1. The consequence drawn from this finding is that linguistic elaboration needs to be addressed more specifically when scaffolding students in synthesis writing (see also discussion section of the article in 5.4).

The difference regarding the adoption of the modelled workflow is reflected in the attitudes towards scaffolding instruments which were expressed by the participants in questionnaires and interviews. In both conditions, the second scaffold in row that was administered in session 3 was considered less helpful than the first scaffold administered in session 2, as Figure 32 reveals.

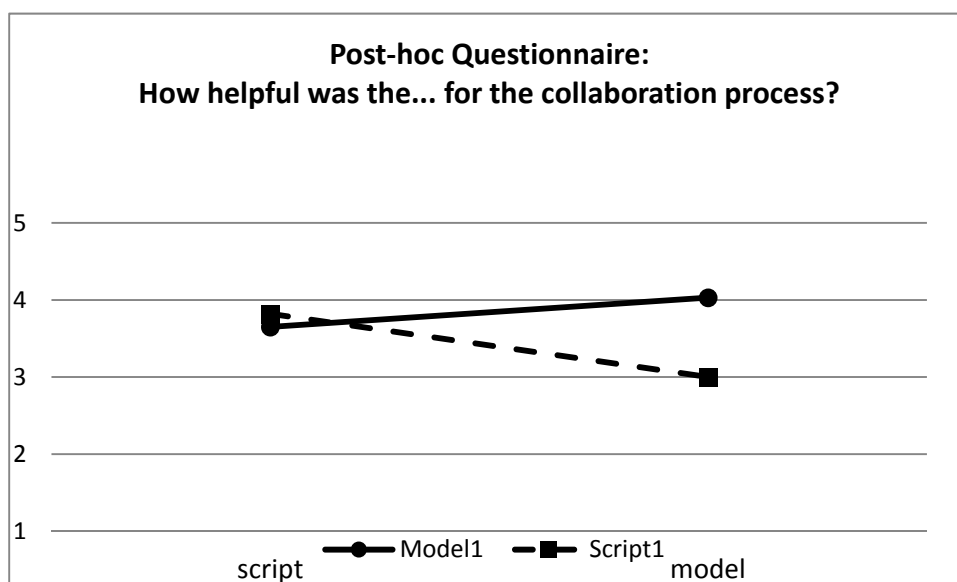


Figure 32 Attitudes towards usefulness of the scaffolding instruments

In the post-hoc focus group of the script-first session, all six participants declared to prefer the script above the model. From their statements it became clear that they found the video model far too prescriptive. The following statement by a participant of this focus group nicely summarises the prevailing opinion in this group (translated to English, followed by the Dutch original in italics):

I liked the script better because you can interpret it your own way. I also liked that it said "you should spend about 15 minutes on this". This was useful because then you realise "uh, I already spent half an hour on this, so I'd better shorten this step". And the video, well, I actually found - the advantage was that it showed examples but well as this was the second time [referring to the scaffold sequence] - you see the video and you think "oh no, but I'd do this differently", well, the examples also put you in a kind of a corset. I do it in another way and therefore I liked the script better.

Ik had het liefst het stappenplan omdat je daar meer zelf uw eigen invulling aan kunt geven. Ik vond het ook goed dat daarin stond van "kijk daar moet je ongeveer een kwartiertje over doen" dat was nuttig dat je zoiets hebt van "oei ik ben daar al een half uur aan bezig da's misschien toch niet zo slim allez dat zou normaal gezien korter moeten zijn". En met dat video dat vond ik eigenlijk - het voordeel was dat het voorbeeldjes bevatte maar omdat het dan ja het was de tweede keer die video - je ziet de video en dan denkt je van "oe nee dat zou ik toch anders hebben of" goh door die voorbeelden zit je ook een beetje meer in een korset zo - ik doe het op een andere manier en dan vond ik het stappenplan handiger.

In the post-hoc focus group of the model-first condition, the opinions about the usefulness of both scaffolding instruments were less unanimous. While the script was considered useful by all participants, some attributed an added value to the video based on its explicitness. The general opinion was that both scaffolding instruments had their value, as the following statement resumes:

I found the combination of both interesting because I remembered better the steps from the video. But I also liked to have the script as a hand-out next to me, so I could have a look again.

Ik vond de combinatie van de twee interessant omdat ik de stappen vooral heb onthouden uit de video. Maar ik vond ook wel handig dat ik dat blad naast me liggen had met de stappen dat ik daar ook nog eens naar toe kon kijken.

The observed differences in learning behaviour and attitudes towards the scaffolding instruments seem to suggest a difference in the groups of the script-first and model-first condition. Indeed, the participants of the two conditions were samples of different, though comparable populations. All participants were students of the third year of a programme with German L2 as a major subject. However, the participants in the script-first group were enrolled in a Linguistics and Literature programme at UGent, and the

participants in the model-first group were enrolled in an Applied Languages programme at the HoGent. Furthermore, the intervention in the latter group formed part of a course in advanced writing training which was taught by the researcher herself, whereas in the script-first group, the intervention was a (credited) part of a linguistics course where the researcher acted as a guest lecturer. As mentioned before, this cross-institutional design was adopted in order to enlarge the sample size, and consequently enhance statistical power of quantitative analyses. Leakey (2011), who described this common practice in CALL studies, warned that "[t]he problem with such a construct is that one is unlikely to be comparing like with like (e.g. different locations, timetabling, teachers, resources and curriculum being the main hurdles)" (Leakey, 2011, p. 244).

To control for a possible difference between the two groups involved, a pre-hoc questionnaire on learning style was administered. To this aim, the Ehrman & Leaver (2003) construct was used which is a validated instrument in language learning research. The questionnaire consists of 30 items, 3 of which target a specific dimension of learning style each, viz. field dependent vs. field independent, field sensitive vs. insensitive, random vs. sequential, global vs. particular, inductive vs. deductive, synthetic vs. analytic, analogue vs. digital, concrete vs. abstract, levelling vs. sharpening, and impulsive vs. reflective learning style preference. All subscales in this construct are based on scholarship in educational psychology and/or L2 learning and contribute to the overarching learning style characteristic posited by Ehrman and Leaver, viz. synopsis vs. ectasis. The authors summarise the two types as follows: "Synoptics 'trust their guts', and ectenics tend not to" (Ehrman & Leaver, 2003, p. 395), in other words, a synoptic learner is someone who easily picks up language from context, and an ectenic learner is someone who prefers to know the rules before engaging in language-related interaction. Each of the 30 statements consisted of two alternative options representing one of the two poles of the respective cognitive style (see Appendix 20: Part 1 of the pre-hoc questionnaire 2). Participants had to indicate their tendency towards one alternative or the other on a five-steps-scale.⁴¹ In addition to the 30 questions from the Ehrman & Leaver construct, three questions were added to investigate learner preference regarding individual vs. collaborative learning. These three questions originated from the construct used in IS2 (see Appendix 10). This addition was deemed necessary, as collaboration plays an important part in IS3, and the Ehrman & Leaver construct does not cover this aspect.

Cronbach's Alpha of the Ehrman & Leaver learning style construct (30 items, i.e., excluding the three additional questions that are not in the original construct) for the

⁴¹ In the original Ehrman & Leaver construct, a 9-steps-scale is used. It was decided to reduce the options because the 9-steps-scale seemed confusing.

population in this intervention was .721, indicating an acceptable reliability for statistical inference testing⁴². An independent samples test of means revealed that there is no significant difference between both conditions ($t=-.234$, $df=68$, $p=.816$), which means that the students in both conditions were comparable regarding their preferred learning style.

To conclude, it can be stated that the differences in behaviour and attitudes towards the scaffolding instruments that were observed in the analysis of the two conditions, cannot be explained by a difference in learning style. Beside a sequencing effect due to the consecutive use of the scaffold instruments, other, external factors, might be at play here, which will be discussed through an activity-theoretical lens in the concluding section of this chapter.

⁴² Although acceptable, this value is lower than expected, as the E&L construct is widely known and used. As I could not find any published figures about the reliability of the concept, I contacted the authors with this request who did not seem to dispose of this information. Betty Lou Leaver alerted me to the fact that the test had never been used in a writing context by the authors, but only in the context of oral performance (personal communication via e-mail). However, I do not see how this could have influenced the result, as I did not correlate the test with writing performance. The changes I made with regard to the original construct were the translation to Dutch (double-checked by a native speaker), and the rescaling of the 9-steps to a 5-steps scale. The additional 3 items for collaborative vs. individual learning style were not included in the calculation of Cronbach's Alpha.

5.5.2 Discussion of the study in the light of Activity Theory

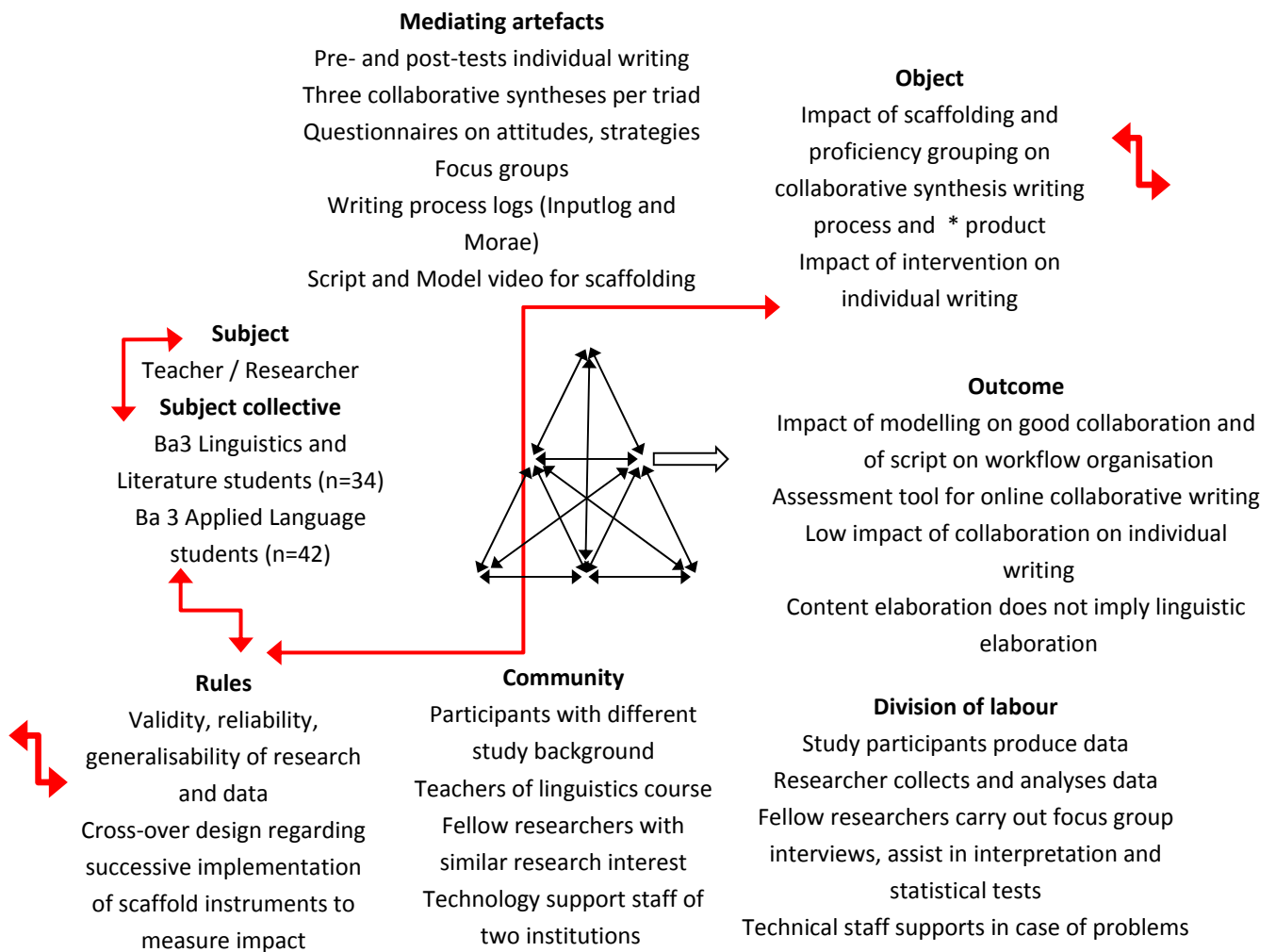


Figure 33 System of the research activity in IS3

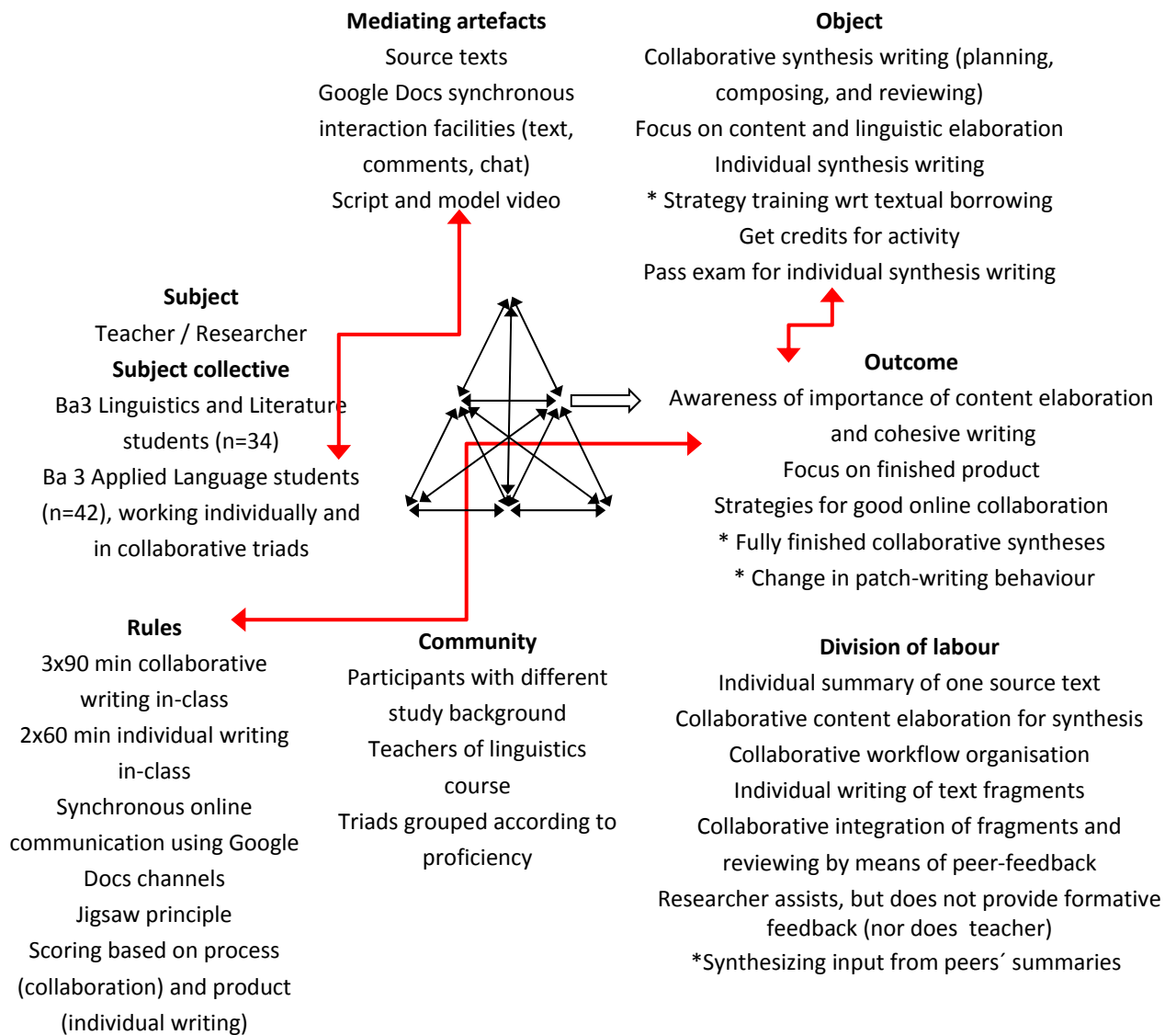


Figure 34 System of the educational activity in IS3

In total, eight major contradictions can be observed in the assessment of IS3. Each of the two activity systems reveals three internal contradictions. In addition, there are two contradictions *between* both AS'.

The Activity system of the *educational* activity of IS3 (see Figure 34) illustrates three important secondary contradictions, the first between artefacts and subject collective, the second between object and outcome of the activity, and the third between the rules and the (intended) outcome.

1. The *model video* which was used to support the participants in the collaborative workflow was not perceived as an adequate scaffold instrument by the students in the script-first condition. Although this is in part related to the scaffolding

sequence, the attitudes expressed in the questionnaires and focus group led to the assumption that this instrument was perceived as too prescriptive by this group. Video models have been reported to stimulate writing strategy development through emulation in secondary education and with freshman students in tertiary education (Couzijn, 1999; Raedts, Daems, Van Waes, & Rijlaarsdam, 2009; Van Steendam et al., 2010). However, they might lose their effect in subsequent years of tertiary education, as students by then have already established a routine in working their own way through study-related tasks.

2. The intervention mainly targeted the object of teaching students to elaborate content and integrate their own voice (i.e., restructure and rephrase source content) when synthesising source texts. To this aim, a *jigsaw* principle was applied to create the extrinsic need for content summarisation and joint elaboration. However, the design only partially led to the intended effect. In the script-first group, it was jeopardised by a different mental representation of the task in which process *efficiency* prevailed over process *effectiveness*, leading students to create their own workflow and skip time-consuming but effective task steps for content and linguistic elaboration. This means that these students did not internalise the researcher-teacher's priorities regarding the object of the activity, but they set their own priorities instead. From an AT viewpoint of learning, these students actually displayed a high level of metacognition, because they did not just follow the researcher's set path for collaboration, but engaged in their own routines. As Engeström (1987) puts it:

[A] truly high level of metacognitive awareness in learning requires (...) not just balancing the components of the learning situation but 'seeing through' the inherent contradictoriness of the learning tasks, i.e., their double nature as unities of exchange value and use value (Engeström, 1987, p. 130).

In the model-first group, triads actually followed the modelled workflow and elaborated the source content in the planning phase. However, this did not lead to the intended linguistic elaboration, because after planning, they copy-pasted fragments of the original sources anyway. This lack of linguistic elaboration which manifests itself in a high similarity between the source texts and the syntheses were observed in both the collaboratively and the individually produced texts. Actually, as posited in the concluding remarks of the article in 5.4, it might even be the case that the collaborative design in itself hampered the development of linguistic elaboration. Finding one's own voice is difficult enough for an individual writer, let alone combining three different voices into one coherent style.

3. The third contradiction concerns the rule that the whole collaboration needed to take place in class. Due to a tracking problem discovered during the analysis of the collaboration processes in IS2, it was decided to screen-capture the whole process in IS3. To this aim, the collaboration needed to be restricted in space (classroom) as well as in time (classroom session). Especially the latter restriction turned out to be counterproductive for the learning activity, as many triads were not able to finish their products in the given time. Beside causing students' dissatisfaction with delivering half-finished products, it also implied that not all triads even got to jointly reviewing their synthesis, which is an important part in collaborative writing.

The Activity system of the *research* activity of IS3 (see Figure 33) illustrates three important system-internal secondary contradictions. Two of them occurred between the subject collective and the rules or the subject. The third contradiction occurred between the rules and part of the outcome envisaged in the original research design (see 5.1.2).

4. To measure the impact of condition, i.e., the scaffold sequence, on the activity outcome, a cross-over design was implemented. The two groups for the crossed conditions were convenient samples with a comparable educational background and language proficiency. Although no difference regarding learning style was detected, the discrepancies in the attitudes and learning behaviour of the two groups led to the assumption that the different educational background actually did play an important role. Therefore, the actual intention of the cross-over design was jeopardised, as the differences observed cannot be attributed with certainty to the sequence of the scaffold instruments, and "educational background" must be taken into account as a confounding variable. In the same vein, Kirschner et al. (2004), discussing social constraints and conventions, pointed to the danger of neglecting the factor of "educational experience":

"Learners are products of their educational experience and, as such, are used to certain types of education and have been socialised to study, learn and act in specific ways. Denying or neglecting these given constraints will guarantee failure, both of the environment and the learning" (Kirschner et al., 2004, p. 18)

5. The difference between the two groups involved was intensified by the two different roles of the subject in the activity, more specifically, the relationship between the subject and the subject collective. In one condition (model-first), the

subject was researcher and teacher of the subject collective at the same time, in the other condition (script-first), she only acted as a researcher. This might have influenced the students' behaviour (and perhaps even their attitudes) during the learning activity. The subject's role, then, would be a second confounding variable that hampers a robust interpretation of the results.

6. The third contradiction concerns the rule that the whole collaboration needed to take place in class in order to ensure the availability of the process data for analysis. It turned out to contradict the object of fully finished collaborative syntheses in the educational activity (see item 3 of this list). This, in turn, entailed a major consequence for the research activity, as it implied that a comparison of products in order to detect a (hypothesised) translation of process quality to product quality (see design of IS3 in Chapter 5.1.2) was jeopardised.

Furthermore, there are two external, tertiary and quaternary, contradictions between both AS', both involving rules:

7. The first, tertiary, contradiction occurred between the rules of the research activity and the mediating objects of the learning activity. The cross-over design with a stepwise implementation of scaffold instruments that was necessary from a research validity point of view proved to be counterproductive. Students clearly experienced an effect of "over-scaffolding", paying less attention to the second scaffold they received.
8. A second external contradiction can be stated between the rules of the learning activity, more specifically, the end-of-term assessment for the course within which the educational activity was embedded in case of the model-first group, and the object of the research activity. The object of the research activity included both individual and collaborative writing, but actually the collaborative writing intervention formed the core part (also expressed in classroom-time used) of the research AS object. A tension occurred with the rules of the learning activity, as the end-of-term assessment only included individual synthesis writing. As far as the collaborative writing part is concerned, it was presented as an exercise to improve synthesis writing, and not rewarded with credits. As it is actually an *institutional* rule or rather a habit carved in stone that restricts the assessment options to individual forms, it seems adequate to classify this as a *quaternary* contradiction between the educational activity and the overarching institutional assessment activity. The case is different for the script-first group.

Here, all activities included in the intervention were jointly rewarded with credits that accounted for a small percentage of the final mark. Furthermore, the collaborative intervention was presented as a preparatory task for another course-related group task in Google Docs that was rewarded with a large percentage of the final mark. Thus, in the script-first group, this external contradiction did not occur.

There are four main consequences from the observed contradictions. They concern both educational and research-related caveats.

Firstly, from an educational point of view, scaffolding instruments with different scopes like a script and a model video should not be spread over several (collaboration) sessions. It makes more sense to choose the most appropriate form according to the educational background of the students, or at most, to combine them in one session, followed by another, unscaffolded, session for practising.

Second, there seems to be an urgent need to raise students' awareness of textual misappropriation (i.e., extensive patch-writing) and to stimulate *linguistic* elaboration of source texts. Consequently, appropriate strategies have to be trained in addition to content elaboration. These might include rephrasing strategies and skills, and the training of cohesion building through the appropriate choice of cohesive ties (see also section "Directions for further research").

Third, from a research point of view, cross-institutional designs should be avoided, even when the comparability of the groups seems high. A valuable alternative to increase sample size, if time allowing, are repeated studies in the same institution over several academic semesters or years.

Finally, when investigating complex writing tasks, time allowance plays an important role and needs careful consideration. In the case of IS3, clearly, the time needed for the thorough collaborative elaboration and reviewing of a synthesis was underestimated. Although the process had been tested in interactions with fellow researchers, and timing had been thoughtfully premeditated, this procedure failed to preview the time investment needed by the students. There are two possible solutions for interventions targeting a complex writing task. The first option is to maintain the time limit, but shorten the source text length. However, this seems counterproductive with reference to the object of the learning activity. It would imply a considerable reduction of the task complexity, given that the ability to select and summarise content of a source text constitutes an important learning sub-goal. Another option is to maintain the original complexity of the task with representative source text lengths, but expand time allowance, e.g., by spreading the collaborative planning and reviewing phases over two classroom sessions, and allowing an intermediate individual asynchronous writing phase *in the Google document* to facilitate tracking of text genesis a posteriori.

Chapter 6

Discussion of results through an activity-theoretical lens

This chapter presents a synthesis of all three intervention studies carried out in the course of this PhD through an activity theoretical lens. The first subchapter zooms in on the AS of the research activity, providing a short description of its constituents and unravelling contradictions within the network of AS' that are connected with it, i.e., the educational and the institutional AS. In the second subchapter, all constituents of both the research and the educational activities are analysed in depth to describe their historico-genetic transformation in the course of the research cycle.

6.1 Contradictions involving the research AS

Figure 35 presents an overarching research AS of all three intervention studies. In total, 162 students took part in the three studies as subject collective. With regard to the mediating artefacts, beside different platforms and programs that served both the educational and the research activity, diverse tracking and analysis tools and methods were used to investigate writing processes and products (see detailed description in subchapter 6.2.7.1). In addition, anonymous questionnaires and post-hoc focus groups helped to assess participants' attitudes towards the educational activities.

The main objective (AS: object) of the intervention studies was to explore the potential of diverse online technologies to support the learning-to-write process. These technologies were embedded in different task designs according to a specific pedagogic approach each. The outcome was the result of the analysis that each time led to the definition of a new object and, accordingly, a new mediating artefact for the next intervention. This means that the principle of *knotworking* that characterises the third-generation activity system proposed by Engeström (2001), and that is needed to achieve *expansive learning*, was adopted.

Level III of the AS, i.e., the basis of the triangle, describes the rules, the community, and the division of labour. As stated in the introduction, there are general rules to be respected in research activities, viz., the validity, the reliability, and the generalizability of the method and the results. At the same time, classroom-based research needs to take into account the principle of equality, i.e., all students have to be offered the same opportunities to learn. The community was composed by scientists who steered and encouraged the beginning researcher, such as the doctoral advisory committee that met once a year to evaluate the candidate's progress and give advice on the route to follow, and the research communities that met at conferences to exchange experiences and ideas. In addition, fellow teachers-researchers at the institution also collaborated by providing educational advice, allocating lesson time to carry out interventions, and conducting focus group interviews. Technical staff assisted to ensure a smooth running

of the interventions. Last but not least, the participants themselves also shared in the labour of the research AS by giving informed consent to being observed and by providing feedback about their experience with the tasks and the tools.

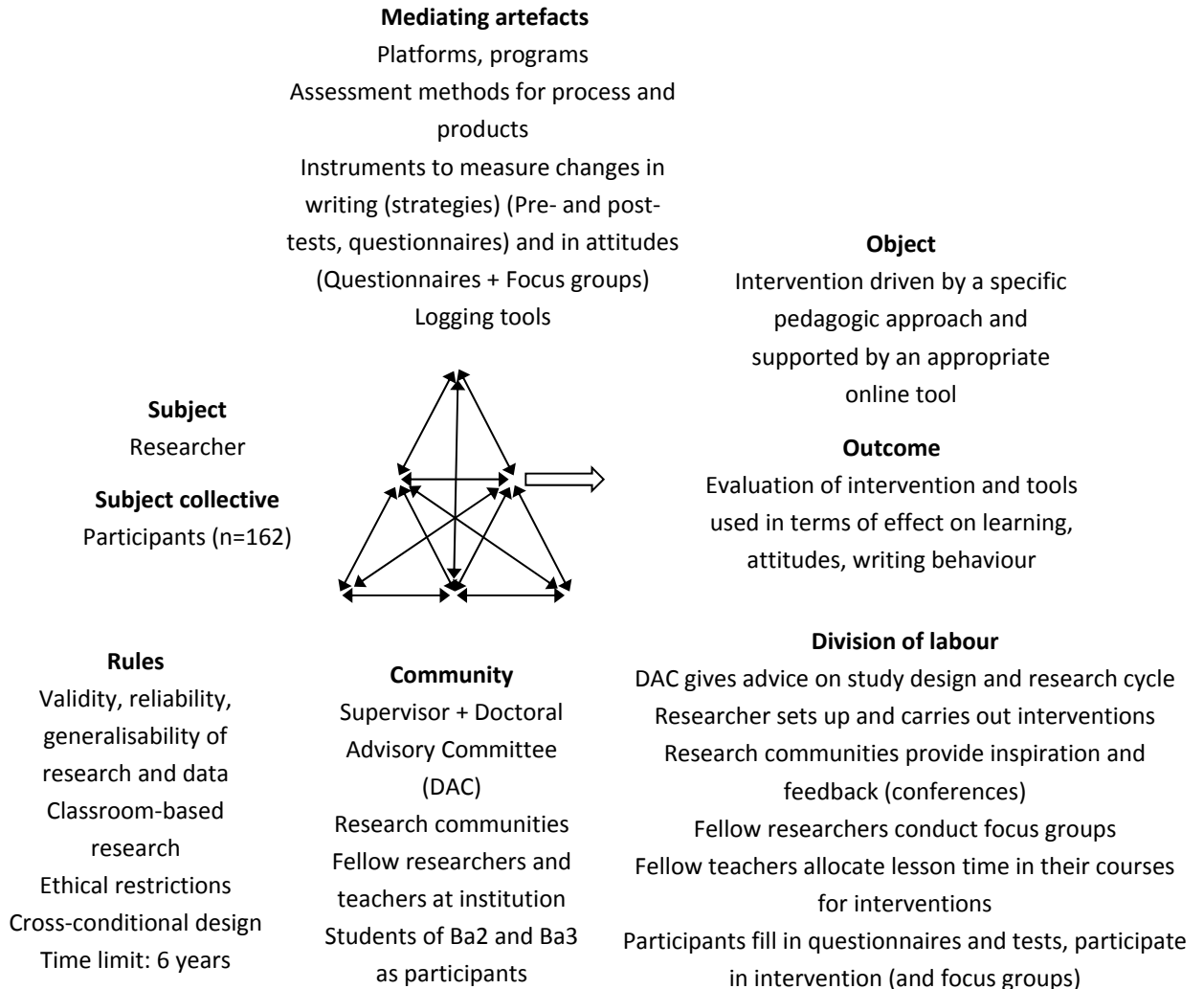


Figure 35 An overview of the general research AS of this PhD

All three intervention studies of this PhD were "quasi-experimental" in nature, as they were carried out with convenient sample groups formed by intact classes inside the classroom, and there was no control group without treatment. An advantage of classroom-based research is its high ecological validity, but it also implies some challenges regarding research design and variables. Activity theory helps to unravel the underlying tensions (contradictions) causing these challenges. They boil down to external tensions between the research AS and (a) the overarching institutional AS, and (b) the embedded educational AS. Figure 36 presents an overview of the research AS

and its relation to connected AS'. The double-sided arrows represent the three main contradictions between those AS'.

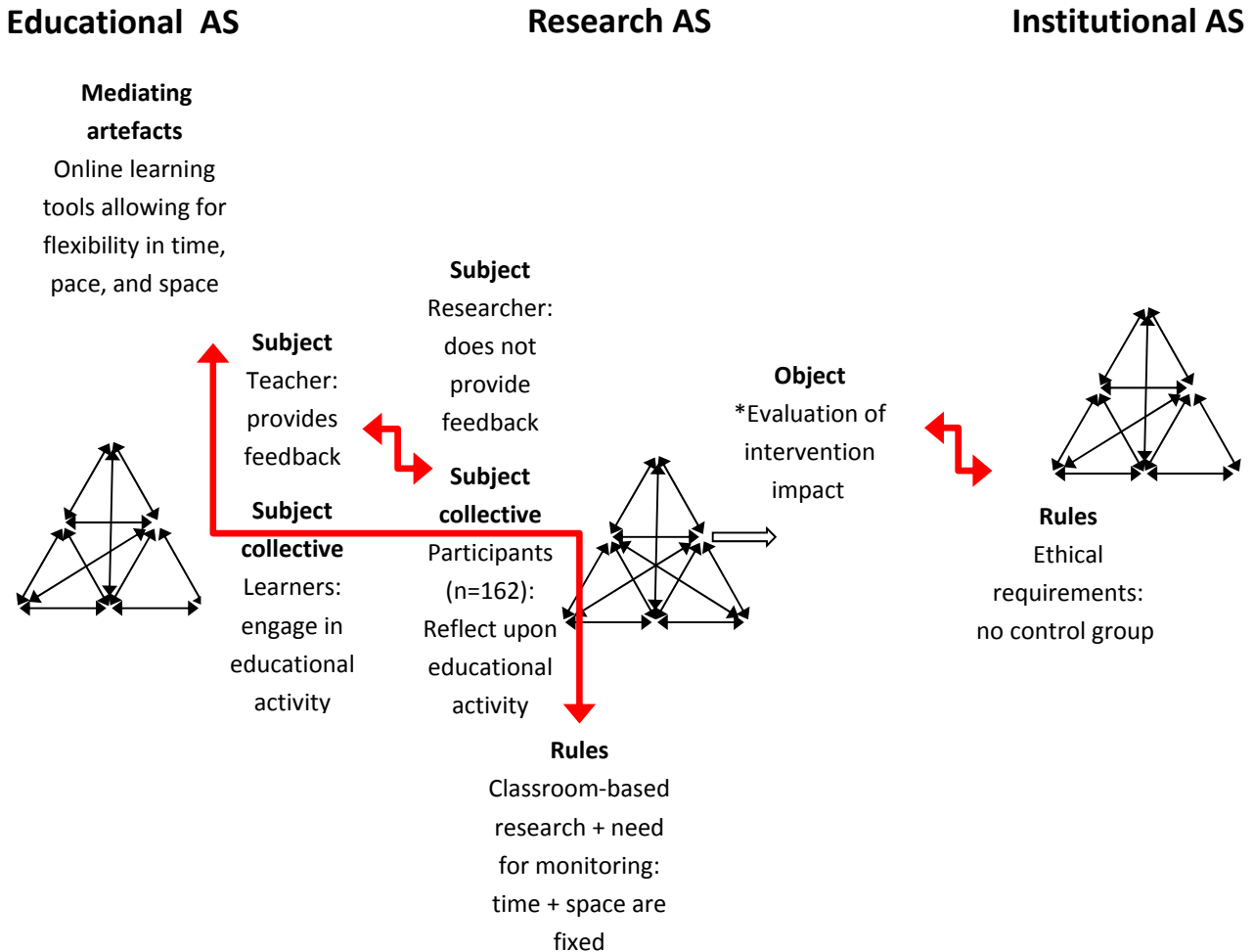


Figure 36 The research AS embedded in the network of connected AS', and the resulting external contradictions

1. There is an external tension between the object of a pedagogical research activity to assess the impact of an intervention, and an institutional rule that imposes ethical restrictions for educational activities. An experimental design with control groups can help to research to a clearer extent the specific differential impact of an intervention. However, given the ethical restrictions in most authentic instructional settings, this is often not a feasible option, leading to the situation that the lack of a control group is a typical characteristic of classroom-based research. This was also the case for the three intervention studies of this PhD. IS1 involved only one treatment group. In IS2 and IS3, conditions were manipulated adopting a cross-sectional design to look at between-conditions comparison. In addition, in IS3, group

composition by proficiency level was added to the design in order to measure within-condition differences.

2. A tension that characterises all three intervention studies is rooted in the very nature of the mediating artefacts used in the educational activity, i.e., the tools to support the online educational activities. One of the main characteristics of online tools (for learning) is the flexibility in time, space, and pace they allow. However, these mediating artefacts were at the same time an important element of the object of the research activity, i.e., investigating writing and learning processes supported by these tools. To ensure that this object of the research activity can be reached, some rules had to be respected that clash with the nature of the online tools. Their inherent flexibility was severely restricted by the fact that all interactions with and through the online tools, be it in individual or collaborative activities, needed to be carried out in class to control (and record) settings and conditions. In IS1 and IS2, this rule mainly served to safeguard that all processes of the educational activity would be running smoothly and that no unforeseen technical problems would jeopardise the research activity object. In IS2, students could in part experience the flexibility of online tools for collaboration, as one part of the writing task was carried out outside the classroom. It was actually the tracking problem discovered during the analysis of these collaboration processes that triggered the decision to screen-capture the whole process in IS3. To this aim, the collaboration needed to be restricted in space (classroom) as well as in time (classroom session). This conundrum provoked sometimes bewildered reactions by the participants. For example, although group members of collaborative activities in IS2 and IS3 sat together in one classroom (sometimes even next to each other), they were instructed to communicate via the computer only. Students understood this was related to the research activity, but they still found it counterintuitive, which is an understandable reaction. In IS1, the online learning modules were also accessible from outside the classroom. Nevertheless, students were asked to attend the classroom sessions so that their online behaviour could be tracked. The following fragment of a discussion in one post-hoc focus group of IS1 illustrates students' bewildered reaction to this rule (English translation followed by the Dutch original in italics):

S1: You could actually do it all the same from home. I think she should explain a bit more why you have to come to the classroom session for this. S2: I agree. You can actually do it from home. Thus, the following week, there were a bunch of students missing because they thought, well I can do it from home, too.

S1: Je kon het eigenlijk even goed thuis doen. Ik vind dat ze een beetje meer moet aantonen waarom je naar de les moet komen daarvoor. S2: Ik ben akkoord met wat je zegt, je kan het

eigenlijk van thuis uit doen, dus de week erachter waren er al heel wat minder studenten omdat ze dachten van ja, dat kan ik ook van thuis uit doen.

3. Closely connected to the above mentioned tension is another one arising from the dual role of the subject as a (guest) teacher in the educational activity and as a researcher in the research activity (see also Blin, 2005, p. 78). To investigate the impact of the online pedagogies and technologies on students' learning activity, the researcher could not interfere by providing additional human feedback, thus, acting as a "typical classroom teacher". However, she was physically present during the intervention to ensure a smooth running of the intervention. This contradiction was mentioned several times by the participants in questionnaires and focus groups (see chapter 3.3.4). At the same time, this tension also posed an ethical dilemma for the teacher-researcher, as the participation in the research activity should be disassociated from the marks related to the educational activity of the course in which it is embedded. However, the teacher in her might be influenced by the observations of the researcher in her, subconsciously rewarding and punishing, as the case may be, students' degree of engagement during the intervention. In addition, the students are used to be marked for all educational activities. If the activities carried out during the intervention were not compensated by an extrinsic stimulus in the form of a mark, they might feel less motivated to participate. Therefore, ideally, a disassociation of both roles (and accordingly, the rules) should be aimed at, which, however, is difficult to realise in action research.
4. Finally, the same dichotomy applies to the role of the subject collective, being students and participants at the same time. As students, the main objective of their participation in the educational activity is to pass the end-of-term exam or to gain credits. However, as participants of the research activity, they are constantly being monitored through intrusive tracking technologies or asked to reflect on the educational activity in questionnaires and focus groups which could cause a certain - understandable - fatigue attitude. Murray and Barnes (1998) rightly pointed out that this tension even might jeopardise the validity of the students'-participants' actions and responses, as those might be shaped according to the alleged expectation of the teacher who will, at the end of the day, mark the results. Indeed, in particular situations during the intervention studies, this tension has become overt. A very nice example for this is the "secret conversation" of a collaborative triad, preparing and commenting in the (supposedly invisible) chat their (visible) discussion in the document that was discussed in chapter 4.3.1.2 (see also Appendix 15). Another episode where this tension became manifest occurred during a focus group interview after IS1. Precisely to avoid a potential inhibition on the students' part to speak out freely, the post-hoc interviews in IS1 and IS3 were conducted by fellow researchers who were unknown to the learners. The researchers explained that the interview would be recorded and transcribed for research purposes only, and names

would be anonymised. Still, at one specific point of a focus group conversation in IS1, a participant asked for reassurance that the teacher-researcher would not listen to the tape, alluding to the possibility of flattery (see Appendix 18). These two examples illustrate a subliminal contradiction which is constantly present in and typical for classroom-based action research. However, there is also counter-evidence to this statement. Students were very open in their questionnaire responses regarding attitudes towards task and tool, conscious of their role as participants in a research study. They seemed eager to contribute to the enhancement of task and tools, indicating major design flaws from their viewpoint as well as minor bugs. On several occasions, their comments disclosed internal as well as external contradictions in the activity-theoretical system, as the examples of statements that were included in the discussion of the individual studies revealed (see subchapters 3.3.4, 4.3.2, and 0).

6.2 Summarising overview of the three intervention studies according to AS constituent

The researchers face the fact that all their skillful efforts to make the participants acquire and apply the culturally more advanced models according to a plan have been partially futile. A genuine expansive cycle inevitably produces not only civilization but also an ingredient of wilderness. To get a theoretical grasp of this wilderness, to find and understand something unexpected as a piece of the history of the future is the reward. (Engeström, 1987, p. 335)

[T]he consideration of sociocultural theory calls into question the scope of language learning itself. Once broader contextual factors are brought into the equation, it is difficult to know when or where to draw the borders of inquiry (Warschauer, 2005, p. 48)

According to the fourth principle of *historicity* of an AS, specific constituents in the overarching research AS were deliberately changed in the course of the intervention design cycle. These *transformations* were motivated by the *primary contradictions* encountered within a constituent, the *secondary contradictions* between constituents of one AS, viz. of the educational or of the research activity, or the *tertiary contradictions* between the AS of the educational activity and the research activity, as explained in

subsections 3.3.4, 4.3.2, and 0). In the following, all transformations are shortly described and discussed per *constituent* of an AS.

6.2.1 Subject collective

Table 10 Transformations of the research AS regarding the subject collective

Study	Year	Institution / Course	n	CEF	Characteristics
IS1	Ba2	HoGent / Summary writing	38	(B1-) B2	<ul style="list-style-type: none"> • Major concern with accuracy • Highly motivated to improve activity-related skills
IS2	Ba2	UGent / Linguistics practicum	48	(B1-) B2	<ul style="list-style-type: none"> • Motivated to acquire general language proficiency
IS3	Ba3	M1 HoGent / Advanced writing skills	42	B2- C1	<ul style="list-style-type: none"> • Motivated to improve activity-related skills • Used to highly structured educational activities
		S1 UGent / Linguistics course	34	B2- C1	<ul style="list-style-type: none"> • Motivated to improve activity-related skills • Not used to highly structured educational activities

Table 10 provides an overview of the student populations involved in the three intervention studies. First of all, it is important to note that the subject collective was a different group of learners each time, thus the overall number of participants involved in the intervention studies amounts to 162. On a positive note, this means that their initial status of acquaintance with online technologies for language learning in class was comparable, i.e., inexistent (with a few exceptions who had prior experiences with individual online pattern-drill in secondary education). On the other hand, this also entails the negative consequence that the students could not compare the different technologies and pedagogies that were used throughout the interventions which might have provided interesting insights from the user perspective.

The interventions were carried out at two different institutions, a University College (HoGent) and a University (UGent), respectively. This had a twofold reason. On the one hand, it allowed the researcher to adapt the rules regarding external reward for the participation in the activities (see below), and on the other hand, it served to increase the sample size in IS3. An important difference regarding prior experience with highly

structured activities during higher education between both groups turned out to have important, though unexpected, repercussions on the results of IS3, as was discussed in section 5.5.1.2.

The change from Ba2 to Ba3 was motivated by a tension between the subject collective and the overarching object of the intervention studies, viz. academic writing. The focus in academic writing should be primarily on HOC issues. However, in the second year, students expect (and need) more guidance in LOC issues due to their lower general proficiency level of the targeted L2. This tension was solved by selecting two student groups of the third year for IS3.

6.2.2 Subject

Table 11 Transformations of the research AS regarding the subject

Study	Role	Other persons involved	Characteristics
IS1	Teacher	Teacher of F2F lessons in first half of course Fellow researcher for focus group interviews	Acts as tutor-monitor during all sessions, ensuring technically smooth process, but provides no additional feedback
IS2	Guest lecturer	Responsible teacher of the course Fellow teacher of the course	Acts as teacher in introductory lesson, and as tutor-monitor during intervention, occasionally assisted by the teacher, no additional feedback
IS3	Teacher	Fellow researcher for focus group interviews	Acts as teacher in introductory lesson, and as tutor-monitor during intervention, providing no additional feedback
	Guest lecturer	Responsible teacher of the course Fellow teacher of the course Fellow researcher for focus group interviews	Acts as teacher in introductory lesson, and as tutor-monitor during intervention, providing no additional feedback

The role of the researcher-teacher has an important impact on both the learning AS and the research AS. It defines to a certain degree the decision power regarding the rules to be implemented (e.g., for assessment of the *outcome* of the learning activity), the object

of the activity, the mediating artefacts that can and will be used, and the distribution of labour. On two occasions (IS1 and in one group of IS3), the researcher was at the same time the teacher of the participants. So, she could independently decide the rules for assessment and timing. In IS2 and the second group of IS3, she had the role of a guest lecturer, and the rules for assessment, timing, and even, to a certain degree, the object of the activity were adapted in agreement with the responsible teacher and other fellow teachers of the course to suit the overarching activity framework of the host course. However, as the focus of the research activity in all intervention studies was on the online tools in combination with a specific pedagogical approach, the researcher enacted exactly the same role in both settings. She acted as a teacher in introductory lessons that were related to, but not a part of, the intervention in IS2 and IS3. During the sessions of the actual intervention, she acted as a tutor-monitor, ensuring a smooth technical process, taking note of eventual bugs and/or content-related problems, assisting in case of technical problems, but not providing additional feedback. Thus, she acted as a facilitator and monitor, but did not take up the moderator role that is typical in classroom-based education (Ahern, 1998, as cited in Müller-Hartmann & Schocker-v.Ditfurth, 2010, p. 29). The adopted teacher role implies consequences regarding task control. From an activity-theoretical viewpoint, this aspect is included in the distribution of labour, and will be discussed in the respective section of this chapter. The double role of teacher-researcher in IS1 and one group of IS3 caused a *secondary contradiction* that sometimes led to bewildered reactions of the students regarding the missing moderator role of their teacher, as described above.

6.2.3 Object

The object (...) is also a nexus of power and resistance in language educational contexts. For example, who decides what the object of activity is? How will the outcome be evaluated, and by whom, and with what effects? How tightly are the actions and operations monitored? What is the level of agency as enacted/enactable by the participants? These are questions that benefit from a systemic examination of relevant communities, their governing rules, and divisions of labor. (Lantolf & Thorne, 2006, p. 223)

As Lantolf and Thorne rightly pointed out in the above quoted fragment, several competing objects can be involved in a language educational activity. For example, there might be (conflicting) objects of the participants involved in the activity (the teacher's vs. the learners' objects), e.g., a different perspective regarding the importance of specific skills and knowledge to be acquired. Even among the subject collective, different objects might prevail while engaging a language learning activity.

Whereas some might have a mastery-oriented mindset, with a genuine interest in acquiring the targeted skill, others can be more performance-oriented, with the main object to pass the end-of-term exam. When the educational activity is embedded in a research activity, additional objects of the latter come into play. Finding a pedagogically sound balance between the objects of both overlapping activities is one of the most intriguing and difficult tasks when designing educational intervention research.

The overarching targeted *educational object* of the intervention studies was to support L2 academic writing development, addressing both the process and the product levels. The main object of *research activity* was to investigate in what form and to what extent specific online technologies used to translate different pedagogical approaches (i.e., the *mediating artefacts* of the educational activity) can support the educational object. In the course of the intervention studies, both objects of the educational activity and of research activity underwent changes. These changes were caused by tensions discovered between the educational and the research AS, or within both, which drove the innovation cycle. In Table 12, the most important changes are summarised.

Table 12 Transformations of the research (and educational) AS regarding the object

Study	Object of educational activity	Object of research activity	Rational for innovation / object changes
IS1	Summary writing from spoken source	<ul style="list-style-type: none"> Investigate affordances of individual online module (curios) in combination with cognitive constructivist and instructivist pedagogy 	<ul style="list-style-type: none"> Perceived shortcomings of F2F setting to teach targeted task
IS2	Summary writing from spoken source + Synthesis writing from written sources	<ul style="list-style-type: none"> Investigate affordances of individual online module (curios) in combination with cognitive constructivist and instructivist pedagogy Investigate affordances of online tools (learning path + forum + Google Docs) in combination with social constructivist pedagogy Compare individual and collaborative settings for both tasks 	<ul style="list-style-type: none"> Lack of human feedback perceived as major shortcoming of individual online setting -> compare with collaborative online setting Problems of tracking cognitive processes while listening / note-taking -> change of educational activity to facilitate research activity

IS3	Synthesis writing from written sources	<ul style="list-style-type: none"> • Investigate affordances of online tool (Google Docs) in combination with social constructivist pedagogy, supported by scaffolding instruments 	<ul style="list-style-type: none"> • Synthesis writing from written sources is a more natural task in view of research skills to be developed • Collaborative settings can lead to deep content elaboration which is main concern wrt strategy development for academic writing • But: collaborative writing requires more instructional support
-----	--	---	---

Summary writing from oral input, which was the targeted activity in IS1 and IS2, involves both the spoken and the written language. The transition from spoken input to written output complicated the tracking of the cognitive processes. By changing the object of the activity to writing from *written* input in IS2 and IS3, the research focus was restricted to observing *writing*. This was convenient for tracking the text elaboration process which, in turn, allows to a certain degree to draw conclusions about underlying cognitive processes. However, summary writing from one input source text was not deemed challenging enough in terms of cognitive and linguistic effort. Therefore, the object of the research and educational activity was adapted to *synthesis* writing from *written* input.

6.2.4 Outcome

The outcome of an activity is closely related to, but not identical with, the object of the same activity. As Engeström (1987) posited, the outcomes of an activity can be *intended* or *unintended*. Moreover, the outcomes also might differ among the participants involved in the activity.

The educational activities carried out in the course of this PhD research aimed to foster strategies that are essential in academic writing (see chapter 1.1.4) as an intended outcome. At the process level, writing and summarising strategies need to be built up in order to cope with the cognitive overload induced by the complexity of the task. This was done by scaffolding participants in the process, encouraging them to elaborate the source texts in terms of content and language, to outline the text to be produced before

writing, and to adopt recursive rather than linear writing patterns. At the product level, the main outcome was to raise students' awareness of an audience, which included broadening their focus of attention in writing and reviewing to both local (LOCs) as well as global (HOCs) aspects of text quality, striving for coherence and cohesion of their products.

The outcomes of the intervention studies, especially those that were not fully achieved, but only hinted at, each time triggered the design of the following study, thus complying with design-based research principles. These are the outcomes that, following the conventions established by Lantolf and Thorne (2006), were marked with an asterisk in the AS schemes presented for each intervention study (in the subchapters 3.3.4, 4.3.2, and 0). An overview of the main outcomes of the three intervention studies is presented in Tables 13 and 14.

Table 13 Outcomes of the educational AS' of all interventions

Study	Outcomes of educational activity	
	Realised	* Hinted at
IS1	<ul style="list-style-type: none"> • Raised self-confidence regarding aspects of the task • Enlarged set of listening and writing strategies • Awareness of summary quality features (HOCs) 	<ul style="list-style-type: none"> • Perceived learning gain wrt task • Raised self-confidence regarding self-assessment
IS2	<ul style="list-style-type: none"> • Improved writing strategies and skills for academic writing • Acquaintance with collaborative writing and online facilities • Perceived learning gain 	<ul style="list-style-type: none"> • Strategies for good collaborative writing
IS3	<ul style="list-style-type: none"> • Awareness of importance of content elaboration and cohesive writing • Efficient workflow for synthesis writing • Strategies for good online collaboration 	<ul style="list-style-type: none"> • Fully finished collaborative syntheses • Change in patch-writing behaviour

Table 14 Outcomes of the research AS' of all interventions

Study	Outcomes of research activity	
	Realised	* Hinted at
IS1	<ul style="list-style-type: none"> • Evidence of "noticing" of task-relevant features both through pre-post-test comparison and analysis of self-evaluation • Models are beneficial to stimulate noticing of HOCs • Students prefer pre-programmed corrective feedback over models for self-evaluation 	<ul style="list-style-type: none"> • Evidence for changes in writing strategies

IS2	<ul style="list-style-type: none"> • Definition of best task for research aim • Benefits of collaborative writing condition for content elaboration, but low impact on overall product quality • Effective online instruction of individual content elaboration and planning • <i>Google Docs</i> as effective tool for collaboration 	<ul style="list-style-type: none"> • Effective instruction for collaborative planning and writing • Insight into writing processes including collaborative communication • Impact of condition on product quality
IS3	<ul style="list-style-type: none"> • Benefits of modelling for good collaboration and script for workflow organisation • Assessment tool for online collaborative writing • Low impact of collaborative intervention on individual writing • Content elaboration does not imply linguistic elaboration 	<ul style="list-style-type: none"> • Measure impact of process on product • Support strategies to avoid textual misappropriation

6.2.5 Division of labour

The division of labour in an educational activity primordially occurs between the teacher (subject) and the learners (subject collective). There is a mutual relationship between the division of labour and the educational roles both actors (are willing to) engage in during the activity. These roles are defined by the following two aspects:

1. Teacher presence (virtual or real)
2. Learner responsibility (individual or shared with peers)

They develop in diametrically opposed directions: While direct or indirect teacher presence decreases, i.e., the teacher ceases to adopt the role of moderator, the learners' responsibility for their own learning process increases. Comparing both parameters in the three intervention studies, the following picture emerges (see Table 15):

Table 15 Teacher presence and student responsibility in the three intervention studies

Virtual teacher presence						
		IS1	IS2	IS3		
Level		Individual scaffolding		Collaborative scaffolding		
Moderation provided by (virtual) teacher	Process	Learning path		Fine-grained Jigsaw Script	Model video	Jigsaw Script
	Content	Pre-programmed corrective feedback	Model solutions	--	--	--
Learner's role	Process	follow learning path		follow script	follow model	follow script
	Content	read feedback	self-assessment	share information with peers and discuss peer-assessment		

Learner responsibility

As the division of labour in the case of online learning is not separable from the instructional scaffold mechanisms employed (i.e., the mediating artefacts of the educational activity), they were included in Table 15 and in the following description. They will be discussed in more detail in the respective section below (see 6.2.7).

The pedagogical online instruments (learning path and integrated testing tool) used to support individual learning-to-write aimed to scaffold the (listening, reading, planning, composing, reviewing) processes in two different ways. On the one hand, a fine-grained individual learning path served to decompose the complex task into manageable units. On the other hand, feedback was provided through the integrated testing tool in the form of pre-programmed corrective feedback according to instructivist principles, or of models for self-assessment according to cognitive constructivist principles.

In IS1, students were strongly guided by the (virtual) moderating teacher, as their labour in the activity consisted in following the learning path and reading the pre-programmed feedback. Only in the part of the module where feedback was provided in the form of models, students were stimulated to take over responsibility for their learning by using the model as a basis for self-assessment. In the individual learning part of IS2, virtual teacher presence was still high in the individual learning path. However, in this case, the open answer questions with model solutions outweighed the

closed answer questions with pre-programmed feedback. In this way, a higher responsibility for their learning was transferred to the students (subject collective).

In contrast with individual scaffolding, the support instruments used for scaffolding collaborative learning-to-write only targeted the process level. The responsibility to provide feedback on a content and language level was entirely on the collaborative group, meaning that the students had to rely on peer feedback, and no feedback was provided by the teacher. The instruments used to scaffold the collaborative writing process changed from IS2 to IS3. The collaborative learning path in IS2 contained a fine-grained workflow script based on the *Jigsaw principle* (Aronson, Stephan, Lides, Blaney, & Snapp, 1978), which means that each group member had exclusive access to a piece of information (in this case, one of the source texts to be synthesised by the group) and was held responsible for sharing and discussing this information with the peers. The script used in IS3 also followed the *Jigsaw principle*, but was less fine-grained and also less intrusive than the one in IS2, as it was presented on a sheet of paper to be consulted instead of an online learning path to be followed. The video used as second scaffold instrument in IS3 was watched by the students before engaging in the collaborative process. It modelled process steps according to the script and additionally contained examples of a mocked-up online collaboration. Comparing the two scaffold instruments used in IS3, the model video could be interpreted as being more prescriptive in that it contained real-life examples of interaction, and therefore, learners might feel compelled to emulate this behaviour. The script left more room for interpretation, and therefore, also put a greater responsibility for the process on the learner.

An important point of consideration is that the above discussed inherent characteristics of the scaffold instruments used only can *afford* certain actions or behaviour to be taken up by the learners. Whether the latter indeed take over the responsibility for their learning, as suggested through the instruments and the pedagogy used, still depends to a large extent on their willingness and motivation.

6.2.6 Rules

Throughout the interventions, several secondary and tertiary contradictions involving rules were discovered. Different kinds of rules were established that depended on the requirements of the educational or the research AS, or even the overarching course or institution AS. They comprise rules for assessment as well as rules for timing of the activity. The researcher had a different role in the two institutions involved, being a

researcher at the University, and a teacher / researcher at the HoGent.⁴³ This impacted her relationship with the subject collective, i.e., the participants and her role within the community involved in the research AS, i.e., as a responsible teacher in her own course or as a guest lecturer in the courses of fellow teachers, as described above. These roles also had repercussions on her ability to define rules for assessment and for timing in the educational and the research activities.

An important rule of the overarching institutional AS is that attending classroom sessions is not compulsory in either of the institutions involved. Actually, this rule fully complies with the flexibility in time, space and pace that characterises online learning. Nevertheless, from the viewpoint of the research activity, the participants' presence in the classroom during the intervention was necessary to facilitate process tracking. As described in detail above (see chapter 6.1), this tension between the institutional rules / the characteristics of online learning on the one hand and the research activity rules on the other led to bewildered reactions, especially in IS1. Some of the university students participating in IS3, conscious of the institutional rules and not feeling obliged to the researcher-guest lecturer, did not attend one or more of the five classroom sessions in which the intervention (including pre- and post-tests) took place, which led to data loss.

Table 16 Formal rules in the three intervention studies

	Timing	Place	Assessment based on	%¹
IS1	6 x 110 min. ²	<ul style="list-style-type: none"> Guided tasks in online module and self-assessment in class Summary writing at home 	(no formal assessment)	0
IS2	2 x 110 min.	<ul style="list-style-type: none"> Guided tasks for elaboration and planning in class Summary/synthesis writing at home 	<ul style="list-style-type: none"> final products (1 individual, 1 collaborative) 	5
IS3	3 x 90 min. + 2 x 60 min. ²	<ul style="list-style-type: none"> All writing phases (planning, translating, reviewing) in class 	<ul style="list-style-type: none"> individual pre- and post-tests (products) collaborative products 	10 (S1)

⁴³ To avoid confusions, it should be mentioned that in the meantime (after the intervention studies), both institutions merged, and the former University College department now forms part of the University as a faculty department.

	(50% individual credit, 50% joint credit)	
	• individual pre- and post- tests (products)	7.5 (M1)

¹ of all course credits ² incl. pre- and post-tests

To stimulate classroom-based participation in the educational activity, the latter had to be integrated in the overarching course assessment. After all, the participants needed an extrinsic motivation and were entitled to an extrinsic reward in the form of credits for their dedicated participation in the intervention. As Table 16 illustrates, participation in the activity was awarded with a mark that counted for 5 to 10 % of the whole course credit. This means that the assessment of the participation did not have a decisive weight in the final mark, but still served as an incentive. In IS1, no credits could be attributed to participation, as the course rules established prior to the intervention did not allow any form of continuous assessment. In IS2 and IS3 S1 (script-first condition, i.e., the UGent group), the assessment of the collaborative products each time consisted of a combination of an individual mark for individual contributions and a joint mark for the group product. In IS3 M1 (model-first condition, i.e., the HoGent group), the original intention to assess the collaborative product had to be abandoned during the intervention for two reasons. First, too many groups did not succeed in finishing their joint syntheses in the foreseen time, which hindered a consistent assessment of the products, and second, individual group members were concerned about jeopardising their final course mark through what they regarded as "failed collaboration". In fact, this was not the case at all, as their collaboration was *effective*, albeit not *efficient*, as discussed in chapter 5.5.1.2. To safeguard their motivation for the collaborative intervention, it was resolved not to formally assess the collaborative product. The fact that students in M1 did not succeed in finishing their syntheses was also related to another rule-based tension, as explained in chapter 5.5.2. The whole collaboration process was recorded for tracking during the classroom sessions. Therefore, time allowance could not be extended beyond the 90-minutes limit of the classroom session.

6.2.7 Mediating artefacts and their affordances

6.2.7.1 Mediating artefacts of the educational activity

Different kinds of mediating artefacts were involved in the *educational* activities of this PhD:

1. Online platforms in which the tools used were embedded,
2. Programs and tools used for scaffolding and supporting students while engaged in the activities,

3. Source texts to be summarised or synthesised,
4. Language(s) (L1 and L2) used in the interaction (student-computer / student-student).

Table 17 Mediating artefacts used in the educational activities

	Platform + program(s)	Other scaffolding instruments	Source texts	L1/L2
IS1	<i>Zephyr curios</i>	--	<ul style="list-style-type: none"> • 4 spoken radio features on current affairs topics of socio-cultural interest from German-speaking societies 	L2
IS2	<i>Minerva curios</i>	--	<ul style="list-style-type: none"> • 2 spoken radio features, see IS1 	L2 + L1
	Forum Learning Path <i>Google Docs</i>		<ul style="list-style-type: none"> • 2 sets of 3 written sources each: factual information in bulleted form, persuasive text, quotes with different opinions on the matter 	
IS3	<i>Dokeos</i>	Video Script	<ul style="list-style-type: none"> • 3 sets of 3 written sources each: popular scientific texts about contemporary varieties of German and related socio-political discussions representing different viewpoints (collaborative writing) 	L2 + L1
	<i>Minerva</i> <i>Google Docs</i> <i>MS Word</i>		<ul style="list-style-type: none"> • 2 sets of 2 written sources each: idem (individual writing) 	

In the following, first, the languages used in the activities and the choice of source texts, platforms, and programs will be reflected on briefly. Then, the technological tools will be discussed in more detail, elucidating their assumed and perceived affordances, the actual uptake of the latter during the activities, and the constraints that hindered the uptake of the affordances in the educational activities.

Language is an important mediating artefact, especially in activities related to language learning, where the improvement of the L2 is always included in the object(ive). It is important to clearly delineate the use of the L1 and the L2, according to the prevailing object. As the objects changed over time, so did language use. In the individual online modules in IS1 and IS2, only German (L2) was used. In the collaborative activities, the scaffolding and support instruments were formulated in German, whereas students were allowed and actually encouraged to use their mother tongue while collaboratively elaborating content and discussing workflow. This was

deemed necessary to ensure that joint content elaboration would not be hampered by a lack of linguistic resources. However, there were some triads that tried to use the L2 as much as possible in their interactions, which led to some interesting showcases of playful code switching between the L2 and a "chat variety" of the L1 which contains a lot of dialectal Flemish characteristics, a few English expressions, and of course non-verbal symbols like smileys (see examples in the appendices of the articles included in 5.2 and 5.3).

Another important mediating artefact were the source texts. They were subject to changes in the course of the interventions, not only with regard of the mode (spoken vs. written), which was adapted to the object of the activity, but also with regard to their content focus. For IS1 and IS2, news features regarding current affairs topics instigating public discussions within German-speaking societies were chosen, as all students involved majored in German, and therefore were expected to have an interest in current socio-cultural developments of these societies. As the activity in IS3 was integrated in a content-focused linguistics course at the University, taking up a considerable amount of total course time, the scope of the source texts had to be adapted in order to fit the course subject. In accordance with the responsible and the assisting teachers of the course, the overarching (linguistic) subject "Varieties of contemporary German" was chosen. However, as the second group of participants studied Applied Languages, they did not have the required background nor the interest to process scientific texts with a linguistic scope. Furthermore, at HoGent, the activity formed part of a writing skills course. Therefore, a reasonable trade-off was found in using popular scientific texts on the matter. The texts were downloaded from websites of quality press (e.g., *Der Spiegel* and *DIE ZEIT*), of the *Goethe-Institut*, or other informative websites. The topics included *Kiezdeutsch* (a much discussed variety spoken among youth in urban areas where different languages and cultures form a linguistic melting pot), *Denglisch*, i.e., English words and phrases in everyday German, the recognition of Austrian German as an official language in the EU, and the Swiss discussion about diglossia in (pre-)primary education (see Appendix 22).

The affordances of the two different scaffolding instruments, viz. the video model and the script, to support the collaborative work flow and performance of the educational activity have been discussed in detail in chapter 5.5.1.2. They actually represent a non-virtual alternative for the scaffolding through an online learning path that was used in IS2. As they were presented in a different medium than the actual work space (script on a sheet of paper) or *before* the actual collaboration (video model), they were perceived as less intrusive than the fine-grained workplan in the online learning path. These three cognitive tools functioned as *secondary artefacts* in the educational activities in that they were "representations both of primary artefacts and of modes of action using them" (Cole, 1999, p. 91). In the case of IS2 and IS3, they were used to

ensure the intended uptake of the affordances of the *primary artefacts*, i.e., the technologies used to support the educational activity (see discussion below).

To fit the requirements of the different pedagogical approaches adopted in the three intervention studies, different tools had to be chosen. They ranged from tools provided within or linked via the institutions' VLEs to open access online software. The platforms were chosen depending on the institution where the intervention was carried out, whereas the programs were selected in function of their fit with the envisaged pedagogical approach. HoGent did not dispose over an assessment tool that was flexible enough to incorporate all answering options and all types of feedback needed to accommodate the different pedagogic approaches. Actually, the institution had a cooperation agreement with an external partner who provided access to their assessment tool *Edumatic* via HoGent's VLE *dokeos*. Although this tool is user-friendly in terms of input facilities and interface characteristics, *Edumatic* comes as a black box with no possibilities to adjust feedback and exercise types to the (educational) designer's needs. In this specific case, an "open answer" exercise type with the option to provide model answers for self-evaluation was missing which made it impossible to use the tool for the design of a module that integrated learning sequences based on a cognitive constructivist approach. Therefore, the assessment tool developed at and provided by UGent (*curios*) was chosen. This assessment tool can only be accessed through a platform of UGent, whereas IS1 was carried out at HoGent. Therefore, the platform *Zephyr* was chosen, an open VLE that is intended to support courses outside the official study programme of the university, and therefore is accessible for users with a guest account. A guest course was created on this platform to provide access to the *curios* assessment modules in IS1. In IS2, the individual online modules were also implemented in *curios*. As the target group were students at UGent, the university's VLE *Minerva* was used as a platform to provide access to the modules. In addition, two built-in tools of this VLE were used for the collaborative part of IS2, viz. a forum (for content discussion) and a learning path (used by ways of a fine-grained script to steer the collaborative workflow). For both the individual and the collaborative writing, Google Docs (GD) documents were created. This was also the tool for collaborative writing implemented in IS3. For the individual pre- and post-test production in this study, MS Word was chosen in order to be able to record the writing process with the keystroke logging tool *Inputlog* (see discussion of the artefacts for the research activity below). Furthermore, a video model (provided via the VLE) and a script (hand-out on paper) were two artefacts used to scaffold students in the process of establishing an effective workflow for the collaborative task.

In the following, the different online technologies used will be closely examined, taking into account affordances, uptake, and constraints. All mediating artefacts were carefully chosen based on their affordances for the envisaged educational activities. However, not all affordances were taken up as intended. AT helps to understand the

underlying tensions that can hinder the uptake of affordances. Kirschner et al. (2004) labelled these tensions *physical, logical, or cultural constraints*. In the following, all technologies used in the interventions will be scrutinised, including a brief review of the constraints / tensions that were encountered during process analysis and/or reported by the learners. Table 18 provides an overview of those technologies and their affordances according to the scheme established in chapter 1.2.3.2. The degree to which the affordances fit the educational setting (with an underlying pedagogic approach) is indicated by a score ranging from "very high" to "low".

Table 18 Affordances of the technologies used in the intervention studies

Tool	Approach	Setting	Utility		Usability	
			Educational f.	Linguistic aff.s	Social f.	Technological f.
			Educational aff.s	Linguistic aff.s	Social aff.s	Technological aff.s
Curios	Instructivist		+	+/-	n.a.	+/-
	Cognitive constructiv.	Ind. learning	+	-	n.a.	+/-
Forum		Coll. synchronous planning	+/-	+/-	+/-	+/-
Online learning path	Socio-constructivist	Scripting coll. workflow	+/-	n.a.	-	+/-
Google Docs		Coll. writing	+	-	+	+
		Coll. planning and writing	++	+/-	+	+
	(n.a.)	Ind. writing	+/-	-	n.a.	+

Note: ++ "very high", + "high", +/- "moderate", - "low", n.a. "not applicable"

Abbreviations: Coll. = Collaborative, Ind. = Individual, aff.s = affordances, f.= functionality

Curios

This online assessment tool of the UGent provides an array of templates for different exercise types, ranging from multiple-choice over fill-the-gap to open answer forms. The flexible architecture allows for a broad range of options regarding feedback provision, score attribution, embedding of other resources like weblinks, videos, etc., time restriction, etc.. It therefore has the affordances to support different pedagogic

approaches online. In the case of IS1 and IS2, a "moderate constructivist" approach was applied, combining closed exercise types for knowledge acquisition and preparatory skills training, isolating specific problems, and open answer exercise types for complex skills training. Therefore, the affordances of the tool for both instructional approaches are presented separately.

- a) For instructivist knowledge building and well-delineated skills training
 - Educational affordances: High
 - Flexible architecture allows for different forms of exercises and possibilities to disclose automated pre-programmed feedback and student's results
 - Missing: Option for stepwise feedback disclosure, 1) highlighting problem zones 2)(+3) disclosure of correct response (and/or (elaborated) feedback) (like *Hot Potatoes*⁴⁴)
 - Cultural constraint: Students sometimes skip or very superficially browse feedback which leads to a loss of the desired "pattern-drill effect" through repetitive structure in exercise suites
 - Linguistic affordances: Moderate
 - Predictable errors on specific linguistic levels can essentially be remediated by pre-programmed feedback, except for errors that require more than pattern-matching technology
 - Possible logical constraint: The uptake of this affordance depends highly on the designer-teacher's experience regarding students' possible errors
 - Technological affordances: Moderate
 - The user interface for the students is intuitive but not appealing
 - Time-consuming input due to complex input interface
 - Possible logical and cultural constraint: Users might experience the input interface as daunting
 - Logical constraint: Browser-sensitive (can lead to malfunction)
- b) For strategy and skills training according to cognitive-constructivist principles, providing model answers
 - Educational affordances: High
 - Template for open answer questions + possibility to provide informative feedback without score attribution, instigating the learners to expand their zone of proximal development through questions for self-reflection

⁴⁴ <https://hotpot.uvic.ca/>

- Possible cultural constraint: Students might feel uncomfortable when asked to auto-evaluate their own input (see discussion above)
- Linguistic affordances: Low
 - Feedback on unpredictable input can only be given through the provision of models for auto-correction, as no NLP-based technologies for automated writing evaluation are embedded.
- Technological affordances: Moderate
 - Same as in a) (though input of models with questions for reflection is not time-consuming)

To conclude, it can be stated that *curios* proved to have sufficient affordances to support the instructivist part of the online module targeting knowledge building and well-delineated skills training. Although it also has the basic affordances to support the cognitive constructivist part, the tension between students' expectations with respect to the required feedback (LOC-focused, performance-oriented, provided by a human assessor) and the feedback actually provided (HOC-focused, mastery-oriented, provided online) placed a serious constraint on their uptake in this context. It remains to be investigated whether automated feedback provided by an assessment tool using NLP technologies to process unpredictable input could add sufficiently to the affordances in order to create an *added value* of this technology to individual learning regarding academic writing (see sections "The added value of online technologies in the learning process" and "Directions for further research").

Forum

The forum tool provided through UGent's VLE has the amenities of current forum tools, including an option for nested or linear visualisation. In IS2, it was used as a tool for synchronous content elaboration and planning of the collaborative synthesis. The AT analysis revealed that it didn't fit the requirements for this educational activity in a satisfactory manner due to lacking affordances, especially related to spontaneous interaction.

- Educational affordances: Moderate
 - Group interaction facilities rather suit asynchronous communication
 - Tracking facility for both student and teacher of interaction history
- Linguistic affordances: Moderate
 - Due to interaction facility, the forum in principle affords uptake of linguistic forms while engaged in meaning-focused conversations according to Swain's (2006) principle of *languaging*
 - Possible cultural or logical constraints:
 - Depends on the amount of L2 used in the interaction

- Depends on students' willingness to constructively comment on their peers' input
- Social affordances: Moderate
 - Interaction facilities at hand, but direct feedback related to specific parts of a message can be given only by (cumbersome) quoting
 - Constraints due to individually differing pace between collaborating peers and time delay in reaction to peers' posts
- Technological affordances: Moderate
 - Posts were sometimes delayed due to server problems -> inappropriate for synchronous use and spontaneous interaction

Although a forum basically affords the necessary interaction for collaborative content elaboration and planning, students clearly found the interaction facilities cumbersome and felt the lack of a tool that afforded quick spontaneous interaction. This is why the majority declared to prefer F2F planning above planning via an online forum. This constraint was solved in IS3 by transferring the planning to the shared social space in the GD chat.

Constraints to the affordances of the tool for the activity are related to its (necessary) combination with other tools (i.e., the online learning path, an individual MS Word document) for parallel use which led to window hopping and time delay of synchronous activities. To conclude, the *added value* of this tool actually has to be considered in combination with those other tools. IS2 revealed that this combination of tools did rather hamper than foster the joint activity (see also conclusion).

Online learning path

The online learning path is a tool provided within *Minerva*. It consists mainly of the visualisation of a stepwise procedure to accomplish a complex task or a series of connected tasks. Beside a description of the step, additional elements like links to a *curios* exercise or to external resources or other kinds of educational material can be added to each step. As soon as the user has clicked on a substep of the learning path to visualise its content, a green checkmark appears automatically at the left hand side of the step.

- Educational affordances: Moderate
 - Stepwise guidance through a complex task is possible
 - Additional information can be provided through links and embedded items
 - Checkmarks visualise which steps have already been processed (clicked on)
 - Logical constraints:

- The checkmark appears automatically after clicking on the step, i.e., it has no informative value about the actual completion or even due processing of the required task step
 - Depends highly on individual student's willingness to carefully read and respect workflow instructions
- Social affordances: Low
 - Workflow is highly predefined -> no social interaction with regard to workflow is stimulated
- Technological affordances: Moderate
 - Missing: option for user-defined status indication for each step (see above)

Google Docs

One of the main affordances of GD (Google Docs) that informed the decision to base the educational design of the collaborative writing activity on this tool is that it allows to separate different communication streams through offering multiple channels for interaction. The emerging text itself is the first communication channel, as students cut and paste individually produced text pieces and edit their own and their peers' texts. The second communication channel is the comments function which can be used to discuss text-related issues without immediately interfering with the original text: (a) Ask for clarification regarding peers' text chunks, (b) suggest editing according to higher-order concerns (i.e., changing the order, adding or deleting propositions), or lower-order concerns (suggestions for grammatical or lexical corrections or improvement). Last but not least, the chat allows for a more general communication related to workflow, socialising, etc.

GD was used in different contexts and with different purposes in IS2 and IS3. In IS2, it was used for collaborative text elaboration and reviewing (synchronous or asynchronous) and suggested for individual writing. In IS3, GD was used for the entire collaborative text production including planning, writing, and reviewing. In the following, the affordances of the tool for each of these purposes will be discussed separately.

- a) For collaborative text elaboration and reviewing
 - Educational affordances: High
 - Comment function allows for both synchronous and asynchronous discussion of joint text
 - Version tracking system allows for text editing which can be tracked in the "document history" view

- Missing affordance: It is impossible to view changes in documents (only visible in document history) together with the comments related to those changes (only visible in actual text status) (<-> e.g. MS Word)
 - Possible logical constraint: comment function has to be activated by clicking on the "insert comment" button that is identified by an icon.
 - Linguistic affordances: Low
 - No automated writing correction tool embedded⁴⁵ (<-> MS Word)
 - Social affordances: High
 - Students can decide as a group whether to meet for synchronous collaboration or to use the tool for asynchronous collaboration
 - Technological affordances: High
 - Smooth and relative bug-free functioning
 - Intuitive use
 - No accidental loss of work due to automated saving every few seconds, and due to version tracking system
 - Logical constraint: Browser-dependent (some malfunctions reported by students using Windows explorer)
- b) For collaborative synthesis production including content elaboration and planning
- Educational affordances: Very high
 - Three different channels for interaction (emerging text document, comments, chat) that can be assigned to different work phases with specific purposes: Collaborative elaboration of a text with regard to content and language aspects, work flow discussion, reviewing
 - All channels could basically also be used for interaction with a tutor (e.g., the teacher), to steer the process and provide additional formative feedback. However, this affordance was not taken up in IS3, as this was not conform with the research objective
 - Possible logical constraint: Inconsistent or unintended use of channels. As GD was not conceived as an educational tool in the first place, the channels have not specifically been designed for collaborative planning (which is the case for educational tools developed for collaborative text planning, e.g., TC3 (see description in chapter 5.1.2))
 - Linguistic affordances: Moderate (- potentially High)

⁴⁵ NB: This has changed since IS2 and IS3 were carried out. In the actual version of GD, there is a spellchecker embedded that, however, has a much lower reliability than the one in MS Word.

- Due to its interaction facilities, GD in principle affords *linguaging* in the sense of Swain (2006)
- Possible cultural or logical constraints:
 - Depends on the amount of L2 used in the interaction
 - The uptake depends heavily on students' willingness to constructively comment on their peers' input and critically review the joint product
 - There is no corrective if the joint solution to a language-related problem is incorrect, i.e., a revision was unsuccessful
- Social affordances: High
 - Chat allows for spontaneous interaction needed for the establishment of a good work flow and of social coherence in the group
 - (Nested) comments function fosters text-related interaction and a common sense of ownership and responsibility for the joint product
 - Possible physical constraint (when used in combination with another tool for individual elaboration, i.e., window-hopping is required): reactions on peers' request can remain unnoticed
- Technological affordances: High
 - (see above)
 - Logical constraint in synchronous writing: when automatically saving, GD encounters version conflicts which can lead to instable visualisation of the joint text ("jumping text" on screen)
 - Missing (see "social affordances"): Alert function that can be activated by collaborating peers to request response, e.g., "pushing" pop-up window or flashing symbol on screen (like e.g. a MS Windows alert to request user action)
- c) For individual writing and reviewing
 - Educational affordances: Moderate
 - Both student and teacher can reconstruct text genesis through version tracking facility
 - Linguistic affordances: Low
 - No automated control of spelling and grammar (like in MS Word)
 - Technological affordances: High (see a))

Both the usability and the usefulness of GD were rated high by the students. The only problem that was frequently mentioned in the questionnaires was time-delay in the workflow. This constraint was caused by the synchronous use of GD for the collaborative text production with MS Word for the production of individual text fragments. Students were recommended to write their individual text pieces (i.e., their summaries of the source texts and the individual parts of the collaborative synthesis) in

a Word document, in order to separate the individual from the collaborative work space, and to avoid confusion or version conflicts. Therefore, they needed to change windows during the individual work phase in order to follow-up what was going on in the collaborative work space. When individual members of the group occasionally focused for a longer period exclusively on their individual work space without checking the collaborative work space, their peers had to wait for reactions. There were several solutions to avoid this possible constraint: Some students decided to split their screen in two parts, one for Google Docs, and the other one for their individual Word document. Like this, they could fully profit from the affordances of the two work spaces. Some groups agreed in the chat on the time when they would meet again in the collaborative work space before "splitting up" into their individual work spaces. Of course, this required discipline from the individual group members, and the chat sessions revealed many occasions where students posted several reminders to their peer(s) in the chat, desperately waiting for their contributions to be able to continue the collaborative work. This physical constraint to the affordance of the chat to promote a quick exchange of ideas was explicitly mentioned by one student during the collaboration in the chat itself. After he had waited in vain for response to a question asked in the chat - his peers were focused on writing individual fragments in Word and therefore did not check the joint work space in GD - he posted: "They should add a buzzer in Google Docs" (Dutch original: "Ze zouden een buzzer moeten toevoegen aan Google Docs").

In this context, it is interesting to note that some students in the scripted session made use of the time indications in the script as a "backing" for their request to speed up the work flow. In this vein, the script that was a secondary artefact actually contributed to the perceptibility of the affordances of the primary artefact for the collaborative task. Model and script afforded a shift of GD from being both an (unintended) object and a mediating artefact in the first, unscaffolded session, to a mere artefact in the subsequent sessions. In the first session, students were not yet acquainted enough with GD, and therefore, the potential of its facilities was not exploited to the maximum. In sessions 2 and 3, the video model and the script alleviated the process, modelling or describing the potential of the Google Docs facilities for effective interaction. This example illustrates Engeström's second principle that the role of constituents in an activity system are variable.

From the perspective of the targeted skills in the educational activity, GD indeed afforded the envisaged interactions. However, as the clear differences in usage of the channels for interaction between the first, unscaffolded, session, and the subsequent sessions reveal (see chapter 5.5.1), the scaffolding in the form of a model video and a script was needed to take up all of these affordances in the intended form. Due to physical constraints (students forgot the - short - explanation of the comments function in the introduction, and/or did not find the button to add a comment), half of the triads did not use the comment function at all in the first, unscaffolded, session. The

perception, and consequently, the uptake of the chat affordance was not hindered by this physical constraint, as the chat window opens automatically, as soon as more than one user is logged into the document⁴⁶. On the other hand, the chat has an important drawback for maintaining the communication flow concerning content and text elaboration in that the text "disappears" as soon as the discussion has filled the length of the screen and there is no scrollbar or other option to revert to the discussion that has disappeared. This means that the chat affordances are high for a short exchange about workflow issues, and for social cohesion of the group, but low for a discussion that is likely to extend over a longer period during the collaboration.

Beside the above described physical constraints, also cultural constraints to the perception of an affordance were observed. The difference in uptake of the affordances of script and model by the two groups of participants in the two conditions (script-first vs. model-first) discussed in chapter 5.5.1.2. can be interpreted in this light. Kirschner et al. (2004) defined cultural constraints as "learned conventions shared by a group" (Kirschner et al., 2004, p. 13). In the case of IS3, these conventions were conditioned by the learning experiences in higher education of both student groups. Whereas students at HoGent (i.e., the participants of the model-first group) were used to being scaffolded while engaging in study-related tasks, students at UGent (i.e., the participants of the script-first group) were not. Thus, their institutional learning culture worked as a cultural constraint to take up the affordance of the scaffolding instruments used as *secondary artefacts*. This is also reflected in the fact that they did not rate the usefulness of scaffolding instruments equally high as the HoGent students. Kirschner et al. (2004) adverted to a potential problem arising from these kind of constraints: "Learners are products of their educational experience and, as such, are used to certain types of education and have been socialised to study, learn and act in specific ways. Denying or neglecting these given constraints will guarantee failure, both of the environment and the learning" (Kirschner et al., 2004, p. 18).

Indeed, the process inspection revealed that about a third of the UGent triads in IS3 did not change their collaboration workflow they established in the first, unscaffolded session which had led to a satisfactory outcome in their view, i.e., a finished product. From an instructional point of view, however, their activity was not satisfactory, because they did not engage in deep collaborative elaboration and thorough collaborative revision, but just pasted their individual fragments into one joint text (see chapter 5.5.1.2). Thus, their workflow resembled rather that of a *cooperative* activity. Due

⁴⁶ This functionality also has changed since IS2 and IS3 were carried out. In the actual version of GD, the chat window only opens by clicking on the symbol(s) for the peer(s) that are synchronously logged into the document.

to the cultural restriction, they did not perceive the affordance of the scaffolding instrument to transform their workflow into a truly collaborative one.

It is interesting to note that this cultural restriction did not hamper the uptake of the affordances of the mediating artefacts provided in IS2 which was also carried out at UGent. There are two possible explanations: First, this might be related to a difference in the nature of the supporting artefacts. However, this seems highly unlikely, as the tools used in IS2 (an individual online module following cognitive constructivist principles and an online learning path to script the collaborative workflow) are even more prescriptive than the secondary artefacts used in IS3 in that they subdivide the task into a highly fine-grained stepwise process. A second, tentative, explanation is the fact that the participants of IS2 were only in their second year of undergraduate studies, and therefore, were still more susceptible to being scaffolded in educational tasks, as their secondary schooling experience was relatively fresh in their memory.

Finally, a general primary tension should be mentioned that arose between the electronic mode of the mediating artefacts used for learning in IS1 (and, to a less explicit extent, in IS2) and the pen-and-paper mode of mediating artefacts that students declared to prefer for learning in view of exam preparation, as the following statement of a student in an intermediate questionnaire of IS1 illustrates (translated to English, followed by the italicised Dutch original):

It would be a lot more useful if those grammatical structures were handed out as printed copies. Then we could resort to it while summarising and while studying for the exam. Of course, the exercises will always be readily available via the pc, but during exam preparation, my pc remains switched off.

Het zou veel nuttiger zijn om deze belangrijke grammaticale structuren op papier uit te delen. Zo kunnen we er steeds makkelijk naar teruggrijpen tijdens het samenvatten en het studeren. De oefeningen zijn natuurlijk steeds beschikbaar via de pc, maar tijdens het studeren blijft mijn computer uitgeschakeld.

Furthermore, individual students reported to generally disfavour the use of online technologies for educational purposes because they generally perceive these technologies as belonging to their free-time activities, and consequently, as inappropriate for educational activities. The following statement of a participant of focus group 1 after IS1 is representative for this attitude:

To me, a computer is related to relax time and it doesn't really fit in the classroom.

Ik vind toch, een computer, dat dat nog altijd heeft iets van ontspanning en dat dat niet echt past bij een les.

6.2.7.2 Mediating artefacts of the research activity

In view of the overarching research activity, the following additional mediating artefacts were used to monitor and track the process, evaluate the product, and learn about participants' attitudes, self-reported writing strategies, and their learning style:

1. Screen-capturing or keystroke logging programs (*Morae* and *Inputlog*),
2. Answers and scores of students in online modules (*curios*) and final texts produced,
3. Tools and methods for process and product assessment (analysis based on *Inputlog*-outcome, and partly on automated extraction of linguistic text features),
4. Questionnaires and focus group interviews.

In the following, a brief account of the tools used and their fit for the research activity is presented, including problems that caused data loss. Special attention is given to the tools that served as mediating artefacts for both the educational *and* the research activity.

Curios

The advantages of the flexible architecture of *curios* in terms of feedback provision and answer options from an educational designer's perspective were mentioned above. This flexibility also makes *curios* an interesting tool from a researcher's perspective, as it has the necessary affordances to create questionnaires, both anonymous ones (interesting to track attitudes), and identifiable ones (interesting to track individual pre- to post-hoc changes, e.g., in writing strategies). The ICT Educational Technology group at UGent⁴⁷ deserves a special mention here, as they have been open to suggestions regarding new exercise types for specific task / questionnaire item requirements. Furthermore, data loss when working with *curios* is excluded, as all interactions and results are stored on the university's server, and can be retrieved any time from immediately after until years after data collection. They can also easily be exported to an xml-format for further statistical research. On the down side, it has to be noted that input is relatively time-consuming (in comparison with other programs like *Hot Potatoes* and *Edumatic*, or free online survey programs) and that the decision tree structure of the interface can get confusing, especially in exercises where a high amount of possible answers need to be pre-programmed.

⁴⁷ <http://icto.ugent.be/>

Google Docs

GD has functionalities that make the tool interesting from a research point of view. Via the version tracking facility, a text's genesis can be restored in detail. All changes that were made in the document can be attributed to the person who originated it and the exact time. In IS2, this version tracking system was used for process analysis of the collaboratively written texts. However, the tool version available then only allowed to track the comments that were still in the document at the time of consultation. This means that all comments that had been erased together with the text fragment they referred to, and all comments that had been marked as "solved" by one of the participants, were not available for analysis. In the GD version that was available when IS3 was carried out, all comments that were made during text genesis could be retrieved afterwards. However, the chat history was not saved. Therefore, to gain a full picture of the group interactions during text genesis, it was decided to use the screen-capturing software *Morae* for process analysis.

GD is a freely available cloud-computing tool. However, there are several caveats to be considered when setting up educational projects using GD (and similar tools). To prevent institutions from using this tool without requesting an institution access liable to costs, *Google* restricts the number of users that can open an account within a certain period from PC's that can be identified by their IP addresses as belonging to the same institution. Identifying this problem and trying to solve it caused a considerable loss of time during the first classroom session of IS2. In IS3, the problem was avoided by asking all participants to open an account prior to the first intervention session from home or at different times from a university's PC. Furthermore, *Google* keeps changing its policy regarding the personal data that is allegedly requested to open an account. It is part of the responsibility of the researcher-teacher to alert students to the fact that all personal data which is revealed can and will be used for commercial and other purposes, and that they should not give away phone numbers and other personal details, even if *Google* keeps sending the request to do so (which, up to date, is not a condition to use the tool). Another potential problem with using freely available online tools is that the versions can change at any given time, causing a different behaviour and/or different conditions for its use, which ultimately can jeopardise the comparability of data collected at different points in time. Last but not least, to safeguard that no data will be lost accidentally, it is recommended that the researcher-teacher creates the documents, i.e., holds the status as document owner, and invites the student groups as participants.

Morae

Morae is a screen-capturing software that includes components for process recording (*Morae Recorder*), process analysis (*Morae Manager*), and process observation in real-time (*Morae Observer*). Due to its sophistication (and the high cost), it is mostly used in

commercial software usability studies. As the department holds a license of this tool, it was used both in the preliminary usability studies, and in IS3. Both the recorder and the manager components are straightforward to use, and the flexibility of the analysis instruments allows for a broad range of applications, from a micro-analysis at the keystroke level (see 2.2.2.2) to a macro-analysis, defining process phases to analyse time on task (see 5.2). A potential ethical constraint for the use of screen recording and keystroke logging tools is that participants must declare their informed consent to being observed. None of the participants of PS1, PS2, and IS3, in which tracking tools were used, declined to give their consent. However, to a certain degree, these tools *intrude* on the natural classroom processes, as they might instigate participants to behave differently than under "normal", i.e., unobserved circumstances. This is a conundrum that cannot be solved when carrying out research in naturalistic classroom environments.

A technical problem that is worth mentioning is related to data storage. The files produced by *Morae Recorder* during a 90-minutes-process require a high amount of data storage room (approximately 600 MB). When these have to be stored by a whole class of participants at the same time on a server after finishing a session, a bottleneck is created that can cause immense time delays, depending on the server capacities. This caused data loss in IS3 due to participants who interrupted the storage process in order to leave the classroom for their next lesson.

Inputlog

Inputlog (Leijten & Van Waes, 2013) was specifically developed by a research team at Antwerp University to support and further writing process research. It is freely available for researchers who are registered as users. This Windows-based keystroke logging tool enables researchers to record data of a writing session in MS Word. The recorded session can be played back for visual process analysis. Furthermore, *Inputlog* creates several output files with readily available analysis results, ranging from process-product ratio over pausing behaviour to linguistic annotation (see 5.4 for a detailed description of the features used in IS3). As *Inputlog* exclusively works in combination with MS Word, it could not be used for the investigation of online collaborative processes. However, it was of great use for the analysis of individual pre- and post-tests that were produced in MS Word.

Two physical constraints should be mentioned here, as they had an impact on students' attitudes or even caused data loss. The first one concerns the fact that two facilities of MS Word are automatically deactivated when opening a writing session through *Inputlog*, viz. the automated spelling and grammar checker and the drag-and-drop facility. As these are functionalities on which students like to rely (especially when writing in a foreign language) and of which they make frequent use, they reacted

bewildered to their loss in the individual pre- and post-test sessions. These functionalities also have to be reactivated in MS Word after the use of Inputlog, which is especially important in a PC-classroom setting.

The second physical constraint arises from the fact that Inputlog creates a file folder as output. Normally, students don't have the required access to store a file folder on a university server or the VLE. Therefore, other storage hardware needs to be used in order to retrieve the data, which always carries the danger of data loss. In IS3, an unidentified problem during transfer of Inputlog files to an external hard disk led to corrupted files that could not be processed for play back analysis.

It goes without saying that *mediating artefacts* play a pivotal role in CALL research. This is the case not only for the educational activity, but also for the overarching research activity. The tools for *both* activities need to be chosen carefully in order to avoid unintended tensions and the risk of data loss. Platforms and programs offered within the institutional context have the advantage of being well-supported by the technology team of the institution; however, they might not have all affordances to meet the requirements of the envisaged learning activity. As there is an ever expanding (virtual) world of good tools freely available out there, it is worth exploring their affordances in classroom-based activity research. At the same time, it is a vibrant and fast changing world, therefore, a CALL designer-researcher needs to stay alert to new upcoming tools and their potential for language learning.

To conclude, it can be stated from this chapter that taking an activity-theoretical perspective, more specifically, investigating the constituents of the activity systems of the interventions allowed the researcher to thoroughly investigate both the research-related aspects and the educational aspects. It helped to pinpoint contradictions that were the result of design flaws or inherent problems of the network of neighbouring AS' and that, in some cases, hindered the uptake of certain affordances of mediating artefacts.

In the following, a general conclusion will be drawn from the results, first answering the main research questions, then broadening the view towards the *added value* of the technologies used, and finally, pointing out promising areas for future investigation.

Conclusion

An increase in one or more of the standards [effectiveness, efficiency, and satisfaction of those learning] without a concomitant decrease in any of the others means success. This is the proof of the pudding. (Kirschner et al., 2004, p. 18)

Insights into online support in L2 academic writing instruction

The studies carried out in the course of this PhD and their analysis through an activity theoretical lens contribute to the research areas of CALL and L2 Writing instruction by providing insight into the potential of online technologies to foster L2 academic writing. In Chapter 1, the theoretical-methodological foundations for the research were laid, framing it within a sociocultural approach on language learning and learning-to-write. Adopting Kirschner et al.'s (2004) design-based research cycle for CSCL studies to describe the research cycle formed by all studies carried out in the course of my PhD, six overarching research questions were posed at the end of Chapter 1 (see Table 19). The main answers obtained by the studies of this PhD are briefly summarised in the right column next to the questions. In the following, I will expand on these answers.

Table 19 Six overarching research questions of this PhD with summarised answers

RQ	Answers
1	<p>What are <i>stumbling stones</i> for advanced language learners engaging in online tasks generally, and in summarising tasks specifically?</p> <ul style="list-style-type: none">• Lack of cohesion between exercises focusing on language and task focusing on meaning• Listening for gist• Planning before writing
2	<p>What are the <i>affordances of different online pedagogies</i> and according <i>technologies</i> for academic writing activities?</p> <ul style="list-style-type: none">• Educational, linguistic, social and technological affordances need to be evaluated separately• Curious has medium-high affordances to support individual learning, Google Docs has high

		affordances to support collaborative learning, when used in combination with instructional scaffolding
3	What <i>contradictions</i> arise within the educational activity, and between the educational activity and the overarching research activity?	<ul style="list-style-type: none"> • Goal-oriented mind-set of students (due to institutional rules) <-> mastery-oriented cognitive and socio-constructivist CALL design • Obtrusive monitoring for research purpose <-> flexibility in time, place, and space of online technologies
4	What are the learners' <i>attitudes</i> towards the support using online technologies?	<ul style="list-style-type: none"> • Positive towards pre-programmed feedback in closed tasks, mixed towards model solutions and self-evaluation • Positive towards individual learning path, mixed towards collaborative learning suite incorporating different tools • Cautiously positive towards collaborative writing via GD, supported by model or script
5	Which affordances of the pedagogies and technologies are <i>taken up</i> by learners? What are the <i>constraints</i> that limit the uptake?	<ul style="list-style-type: none"> • Depends on learners' characteristics and on sound pedagogical design: Combination of tools, secondary artefacts for support • Tool specific constraints range from technical compatibility problems to cultural constraints caused by learner histories
6	What is the <i>learning gain</i> from the different intervention studies? What <i>outcomes</i> did students (not) achieve? Consequently, where should innovation focus on?	<ul style="list-style-type: none"> • Increased awareness of text quality features • Better strategy use (more planning, more recursive writing, more reviewing, especially of HOCs) • Better content elaboration in collaborative settings • NOT: better textual appropriation • NOT: better overall text quality • (unintended side-effect) Adoption of GD for collaboration in other contexts

In order to answer RQ1, preliminary usability research was carried out in view of the preparation of my own intervention studies. The two preliminary studies shed light on advanced language students' online learning behaviour by observing them while engaging in a broad array of tasks in the language learning environment DUO. They were described and discussed in Chapter 2, and the answer to RQ1 was formulated in the concluding subchapter (see 2.3). The main conclusions in view of preparing my own

online learning material can be summarised in two points, one related to content and the other to aspects of *didactic usability*. Regarding content, P2 revealed that listening for gist and planning before writing were two important strategies to focus on. With respect to *didactic usability*, a term that was coined in the article about P1 (see 2.1.2), critical factors for learning success in online language learning modules that were singled out are salience of task instructions and the need to dovetail preparatory language exercises with the corresponding tasks targeting skills development.

Chapters 3-5 were devoted to the description of the intervention studies and the main results obtained. RQ2-RQ6 were addressed in a cyclic manner in all interventions, albeit to differing degrees. IS1 was analysed with a specific focus on learners' attitudes (see Chapter 3), i.e., to answer RQ4. A stark contrast was found between students' positive attitudes towards pre-programmed feedback in closed tasks inspired by instructivist learning theory and their rather negative attitudes towards feedback in the form of model solutions and reflective questions for self-evaluation. Another subchapter focusing on attitudes is 4.3.1.2, which compared students' positive response to technologies used to support *individual* writing in IS2 with a critical attitude towards the technologies to support *collaborative* writing in the same intervention. The educational design adopted in IS3 for collaborative writing using Google Docs for synchronous CMC and synthesis production had a far higher acceptance than the combination of technologies used in IS2. Although this is beyond the scope of the study, it is interesting to mention at this point that *Google Docs* has been permanently and broadly adopted by the HoGent students involved in IS3 for study-related purposes *after* the intervention, to the extent that they started to use it spontaneously for other tasks and in other classes where collaboration was involved, e.g., in collaborative translation exercises, or to prepare a joint *MS PowerPoint* presentation. This could be interpreted as an "unintended" positive outcome of the activity. It should be noted in this context that within their "probabilistic" view on learning, Kirschner et al. (2004) gauge the success of instructional design for CSCL not only by its causal effect on the individual learning gain ("effectiveness"), but also - even alternatively - by a satisfactory learning experience of the participants. In this regard, IS3 surely passed "the proof of the pudding" (Kirschner et al., 2004, p. 18).

RQ6 was specifically addressed in the subchapters 4.2, 5.5.1, and 5.4. The article about IS2 sheds light on the impact of pre-programmed feedback to support content and linguistic elaboration and planning in comparison to peer feedback or joint content elaboration and reviewing through CMC on the product. The results indicated no difference with respect to linguistic quality of the outcome, but an interesting difference in favour of the collaborative setting with respect to content selection.

The results of the process and product analysis of collaborative writing in IS3 were in line with the expectations resulting from IS2, but they also revealed an unexpected effect. Although the collaborating triads overall displayed an expected learning gain

regarding workflow efficiency and interaction focus on HOCs which translated to a better content selection, this joint elaboration process did not have the intended effect on linguistic elaboration, i.e., rephrasing and cohesion building. Instead, the similarity rate between the produced syntheses and the source texts increased throughout the intervention, indicating that students copy-pasted more fragments from the source texts. This result was corroborated by the individual pre-post-tests comparison, which also revealed a higher textual borrowing rate in the post-tests, indicating that the higher writing process efficiency observed did not translate to a better product. Consequences drawn from these conclusions will be discussed below (see "Directions for further research").

All intervention studies were discussed through an activity-theoretical lens at the end of the respective chapters (see 3.3.4, 4.3.2, 5.5.1.2), indicating main contradictions encountered within the educational activity (thus, answering RQ 3), and between the educational and the overarching research activity that caused changes in the educational design of the subsequent intervention study. In chapter 6, the successive development of the intervention studies was described, singling out all constituents that together shape the educational activity, thus providing an exhaustive answer to RQ 3 and RQ 5, which are closely related from an AT point of view. The concluding critical discussion of affordances of the mediating artefacts used in the intervention studies (see 6.2.7) on the one hand revealed their potential for effective support of learning activities in individual and collaborative contexts (RQ 3). On the other hand, it also provided evidence that the *uptake* of these affordances highly depends on the combination with other tools used within the educational, as well as in the overarching research environment, and on other external factors that can impose physical, cultural, or logical constraints (RQ 5).

The added value of online technologies in the learning process

To conclude, the results and insights from the AT evaluation should be combined to broaden the view towards the *added value* of the technologies used, taking into account the envisaged educational objective. To this aim, the formula developed by Colpaert (see 1.2.3.2) should be recalled:

$$\begin{aligned} & \text{Added_value(technologyX)} \\ & = \text{FIT (requirements (optimally_designed_learning_environment), affordances} \\ & \quad \text{(technology))} \end{aligned}$$

From this formula, two conditions can be derived that have to be met in order to attribute an added value to technology used for learning:

1. The technologies used in the interventions have to fit the requirements of the learning environment designed to foster a specific objective;
2. The learning environments has to be optimally designed for the targeted objective.

Actually, the second condition is a prerequisite for the first one to make sense. The best chosen or engineered tools cannot prove their added value when implemented in a poorly designed learning environment. It is therefore important at this point to recall the main objective of the learning environments created. Academic writing is a complex activity that requires the mastery of several underlying cognitive skills and abilities, viz. elaborate the content of source texts, compare different sources, select macropropositions, establish a coherent outline, and compose and review a text, juggling with a threefold mental representation of the same (Kellogg, 2008). These skills are highly dependent on language proficiency (Plakans, 2009), which, in turn, is reflected in relevant linguistic knowledge on the one hand, viz. lexical resources to quote, rephrase, and summarise, and to establish a cohesion on textual level that guides the reader in understanding the underlying coherence, and in the necessary reading, listening, and writing strategies, on the other hand. It stands to reason that these different cognitive and language-related skills and knowledge areas cannot be targeted altogether in one learning environment, let alone in one intervention study. Therefore, specific objectives were emphasised in each intervention study in function of the participants' proficiency level and of insights from the previous study and scholarship on L2 writing and academic writing (see section 0).

In IS1, listening strategies, linguistic means to summarise and build coherence, and awareness of summary quality features with a specific focus on HOCs were the main outcomes envisaged. The learning environment that was created, an individual online learning path, combined different task types according to instructivist learning theories on the one hand targeting acquisition of linguistic means and listening skills, and cognitive-constructivist ones on the other hand, targeting awareness of relevant HOC features through self-evaluation based on model solutions. The assessment tool *curios* that was used to create the learning environment proved to provide the necessary affordances to support those outcomes. Therefore, it is fair to attribute an added value to the technology used concerning the instructivist part of the module, which is also corroborated by the attitudes of the study participants. However, important cultural constraints hampered the uptake of the affordances of the model solutions. The participants reported a lack of self-confidence, combined with a certain resistance to change their performance-oriented mindset into a mastery-oriented one, causing them to underestimate the added value of this technology-supported approach for their learning. In essence, they missed corrective feedback. It would therefore be interesting to investigate whether expanding the linguistic affordances of the technology by adding

automated corrective feedback could help to improve the (experience of) added value (see "Directions for further research").

To counter the "missing feedback conundrum" that resulted from IS1, a different pedagogical approach was added to the design in IS2, broadening the perspective from individual learning to collaborative learning according to socio-constructivist principles, and comparing both settings in terms of their effectiveness to support the targeted outcomes related to academic writing. The analysis focus was twofold. On the one hand, the individually and collaboratively produced texts were compared with regard to their *linguistic* elaboration. No significant differences were detected in that respect. On the other hand, the *content* analysis of the texts revealed a better elaboration in the collaborative setting. Qualitative process analyses provided evidence that this difference originated in content-related discussions in the planning phase. Moreover, an analysis of the text genesis revealed that collaborative writing processes show the characteristics of recursive writing, which is a strategy that has often been highlighted as beneficial for academic writing. Therefore, from a pedagogical point of view, the collaborative setting seemed to be more promising to foster those writing strategies that are especially important in academic writing. However, this was not corroborated by the participants' perceived learning gain.

The AT analysis of the activity revealed that the technologies used played an important role therein. The individual online learning module was constructed in a very similar way as in IS1, and delivered via the same tool (*curios*). However, this time, it was only implemented in the individual content elaboration and planning phase, and not used for self-evaluation of the *final product* through model solutions. Participants of IS2 rated the added value of the individual learning path high, including the model solutions stimulating reflection. Therefore, an added value can be attributed to model solutions used to stimulate reflection during the writing process, but not for self-assessment of a finished written product, at least, if one considers students' attitudes.

With regard to collaborative writing in IS2, the learning environment was actually a combination of three tools, viz., a forum for synchronous planning, a learning path for workflow support, and Google Docs for actual writing and reviewing outside the classroom (mostly asynchronous). Although all three technologies had some affordances to support the learning and writing process, it was actually the combination of different tools that proved to be counter-productive. At the same time, the analysis revealed (a) that the forum was too unwieldy to afford effective synchronous interaction for planning, and (b) that the affordances of Google Docs had not been exploited to the maximum in the asynchronous setting for writing and reviewing.

It can be concluded that the (combination of) technologies used in the collaborative part of IS2 proved to have no added value for the targeted outcome, i.e., learning to collaboratively produce a synthesis from sources. Kirschner et al. (2004) stated that "[w]ith respect to collaboration, the question is whether the elements of the

environment afford the emergence of that type of social interaction that is supportive of the acquisition of the targeted skill" (Kirschner et al., 2004, p. 25). Based on the conclusions from IS2, I would actually broaden the question, formulating it in the following way: "...whether the *combination of elements* of the environment afford...".

The insights from IS2 led to a new design of the collaborative task as a synchronous activity based on Google Docs as the only mediating technology for interaction. To support participants in the task organisation and exemplify the necessary types of interaction, two secondary artefacts were used in the form of a (non-intrusive) script and a video model. They revealed to have a positive effect on the collaborative workflow and interaction mode, albeit to a different extent for the two groups of participants. There are two important results of this intervention with regard to the added value of technologies for learning. First, GD indeed provides the necessary affordances to support a collaborative academic writing task, and second, these affordances can only be taken up to a satisfyingly degree, and therefore, lead to an added value of the technology, in combination with secondary artefacts. This result corroborates Kirschner et al.'s following affirmation: "In the 'physical' world, affordances abound for casual and inadvertent interactions. In the 'virtual' world, social affordances must be planned" (Kirschner et al., 2004, p. 15). Script and model in IS3 were used to plan the social, and, to some extent also the linguistic, affordances of GD.

Limitations of this thesis

A model of evaluation that seeks to measure effectiveness of a pedagogy or CALL object must acknowledge that a multitude of dynamics are at play in the real-life CALL classroom and in the individual learner, and that it is virtually impossible to control for all these potentially confounding variables. Every researcher must, therefore, couch his/her inferences in cautious terms and with reference to the caveats that exist (Leakey, 2011, p. 51).

As Leakey rightly pointed out in the above cited quote, inferences of action research in CALL carried out in real-life classrooms need to be couched in cautious terms. In this vein, some potentially confounding variables and limitations of the research carried out in the course of this PhD should be highlighted. These refer to a) the lack of participants' acquaintance with the technologies used, b) the intrusive character of research instruments used for data collection, c) the lack of control groups in the design of the studies, and d) the "black spot" in the educational and research design regarding *reading* abilities, which play an important role in writing-from-sources tasks.

Most of the students involved in the studies as participants were absolute or relative novices to CALL, and to the technologies used. This factor might have contributed to the

fact that the intended effect of the pedagogical interventions using online technologies was not always clearly measurable. In this vein, Leakey, elaborating on Hubbard and Siskin (2004) stated: "[I]mproved outcomes are more likely with improved research designs and students that are more familiar with CALL. 'Familiarity' suggests long-term exposure and longitudinal studies to monitor this" (Leakey, 2011, p. 117). The studies carried out in the course of this PhD research cannot be characterised as "longitudinal". At the same time, they are far from being one-shot intervention studies, lasting between two and six weeks, thus affording the participants the time and possibility to get acquainted with the technologies and pedagogies used.

(Writing) development research demands scholarly rigour in terms of data collection. Quantitative and qualitative data need to be combined to triangulate results and to control for confounding individual variables. Therefore, the outcome as well as the activity processes were monitored by a series of tools and methods for data collection like pre- and post-tests, questionnaires, screen-capturing, and keystroke logging. The potential pitfall of collecting too much data during the interventions is that students-participants might perceive this as "over-intrusive monitoring". This can lead to a possible disturbance of the natural classroom-based processes of learning - which are precisely the scope of the observation - and ultimately skew the qualitative data (Leakey, 2011, p. 183). However, the AT based approach to the critical analysis of the interventions allowed us to pinpoint the tensions that arose between the educational AS and the overarching research AS. Therefore, at least, they did not remind "blind spots" in the presented research.

Further, it is important to mention that no control groups were involved in the design of all three intervention studies due to ethical requirements of the institutions involved. Therefore I opted to use cross-over designs, manipulating the conditions to observe between-subjects and between-groups effects. This design allowed me to highlight certain tendencies. However, the outcomes could not be attributed with the necessary methodological rigour to the intervention in question. While statistical methods may have failed to reveal causal effects (see e.g. 5.5.1.1), I would like to qualify this statement by adding that the qualitative inspection of processes, especially in collaborative writing, was overwhelmingly convincing with regard to the great potential for "linguaging" and for joint content elaboration of this educational setting (see 5.2 and 5.3).

It stands to reason that the quality of academic writing-from-sources depends not only on the writing skills, but also to a large extent on the *reading* skills of the writer (Plakans, 2009; Wolfersberger, 2007). When writing from spoken instead of written sources (as in IS1), the *listening* skills as well as note-taking techniques play an important role. In the present PhD research, listening strategies have been tackled to a limited extent (in P2 and IS1). Reading strategies, though, have not been included in the instructional design of the studies, nor did I control for reading abilities in my research

design. This aspect would surely have enriched the research. Moreover, scholarship provides evidence that reading strategies can effectively be supported using online technologies (Hsu & Wang, 2010; Lenhard, Baier, Endlich, Schneider, & Hoffmann, 2013; Murphy, 2010; Prichard, 2010). Given that new technologies like eye-tracking can help to investigate online reading strategies in a non-intrusive way, the combination of reading and writing strategies instruction in "wreading" tasks through online technologies is a promising area for future research. However, this additional aspect was simply deemed to go beyond the scope of this PhD.

Directions for further research

Two promising areas for further research, building on the findings of the present PhD, were mentioned in the conclusions. The first one is to combine online reading *and* writing strategy instruction. The second one is to provide automated corrective feedback to stimulate self-evaluation and reviewing in academic writing. The latter approach could help solve the two challenges in L2 academic writing that remained unresolved in the interventions carried out in the course of this PhD. These challenges are a) the so-called *patch-writing* behaviour and, closely related to it, b) students' problems with establishing cohesion in their academic texts (see 5.4). The individual and collaborative learning environments were designed to stimulate content elaboration and textual appropriation through cognitive involvement and interaction with feedback (delivered by a computer or a peer), and these affordances were indeed taken up by the students. However, after elaborating and restructuring source content, they reverted to the original texts, especially when those were at their disposal in written form, and copy-pasted fragments anyway.

Future research in the field needs to specifically focus on strategies to avoid (unintended) plagiarism, and to foster reader-orientation. Several promising routes for strategy instruction to this respect in (L2) academic writing have been mapped out, targeting, among others, intertextual practices (Ivanic, 2004) and self-regulation strategies (Graham & Perin, 2007b). The question that calls for investigation is what role technologies can play to foster these two specific aspects in L2 academic writing. In this vein, writing tutors based on AWE (automated writing evaluation) have been gaining momentum in the last decennium, mainly in English as a Second Language instruction in the US. The fast evolution of NLP (natural language processing) technologies allows for an ever increasing reliability of these tools regarding the recognition of non-target-like input. AWE tools evaluate learners' texts and provide diagnostic feedback on different levels. On the global textual level targeting higher-order concerns, they can, among others, detect over- and underuse of linguistic cohesion features. Based on n-gram comparison, they can also detect (excessive) textual borrowing from a source.

AWE tools, therefore, provide the necessary affordances for writing strategy instruction. Two interesting examples of online writing tutors based on AWE are the *Writing Pal* for English (L1 and L2) (McNamara, Crossley, & Roscoe, 2013), and *escribo* for German L1 (Proske, Narciss, & McNamara, 2012). Both are built according to principles of Computer-Based Scaffolding and aim to support students by guiding them through the whole writing process. To stimulate reviewing, they provide automated feedback with a specific focus on cohesion and paraphrasing.

While computers to date certainly cannot substitute a human being in terms of NLP and interpretation, intelligent computer-generated feedback has other strengths that favour its use in an instructional context, such as time-effectiveness, immediacy of feedback, and reduced anxiety regarding loss of face (Dodigovic, 2005). However, bearing AT principles in mind, their implementation in classroom contexts should be carefully monitored, taking into consideration possible constraints that might hinder the uptake of their affordances. An interesting challenge for future research...

“Todo fim é um novo começo”

Bibliography

- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the Zone of Proximal Development. *The Modern Language Journal*, 78(4), 465-483. doi: 10.2307/328585.
- Aronson, E., Stephan, C., Lides, J., Blaney, N., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.
- Barab, S., & Squire, K. (2004). Design-based research: putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1-14.
- Basharina, O. K. (2007). An activity theory perspective on student-reported contradictions in international telecollaboration. *Language Learning & Technology*, 11(2), 82-103.
- Bax, S. (2003). CALL - past, present and future. *System*, 31(1), 13-28. doi: 10.1016/S0346-251X(02)00071-4.
- Beauvois, M. H. (1998). Conversations in Slow Motion: Computer-mediated Communication in the Foreign Language Classroom. *Canadian Modern Language Review*, 54(2), 198.
- Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing*, 16(2), 97-110. doi: 10.1016/j.asw.2011.02.001.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing*, 21(4), 348-363. doi: 10.1016/j.jslw.2012.09.006.
- Blin, F. (2004). CALL and the development of learner autonomy: Towards an activity-theoretical perspective. *ReCALL*, 16(2), 377-395. doi: 10.1017/S0958344004000928.
- Blin, F. (2005). *CALL and the Development of Learner Autonomy - an Activity Theoretical Study*. Unpublished PhD, The Open University, UK. Retrieved from <http://webpages.dcu.ie/~blinf/BlinThesis.pdf>.
- Blin, F., & Appel, C. (2011). Computer Supported Collaborative Writing in Practice: An Activity Theoretical Study. *CALICO Journal*, 28(2), 473-497.
- Blin, F., Caws, C., Hamel, M.-J., Heift, T., Schulze, M., & Smith, B. (2013). *Data and elicitation methods in interaction-based research*. Paper presented at the Global perspectives on Computer-Assisted Language Learning, Glasgow.
- Blin, F., & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance to change through the lens of activity theory. *Computers & Education*, 50(2), 475-490. doi: 10.1016/j.compedu.2007.09.017.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141-178.

- Bush. (2008). Computer-assisted language learning: from vision to reality? *CALICO Journal*, 25(3), 443-470. doi: 10.11139/cj.25.3.443-470.
- Castelló, M., Bañales, G., & Vega, N. A. (2010). Research approaches to the regulation of academic writing: the state of the question. *Electronic Journal of Research in Educational Psychology*, 8(3), 1253-1282.
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: foundations for teaching, testing and research*. Cambridge: Cambridge University Press.
- Chapelle, C. A. (2009). The Relationship Between Second Language Acquisition Theory and Computer-Assisted Language Learning. *The Modern Language Journal*, 93, 741-753. doi: 10.1111/j.1540-4781.2009.00970.x.
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, 18(1), 80-98. doi: 10.1177/0741088301018001004.
- Clark, R. E. (1989). When Teaching Kills Learning: Research on Mathemathantics. In H. N. Mandl, N. Bennett, E. de Corte & H. F. Friedrich (Eds.), *Learning and Instruction. European Research in an International Context* (Vol. 2, pp. 1-22). London: Pergamon Press Ltd.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cole, M. (1999). Cultural psychology: Some general principles and a concrete example. In Y. Engeström (Ed.), *Perspectives on activity theory* (pp. 87-106). Cambridge: Cambridge University Press.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). New York: Springer.
- Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal*, 23(3), 21.
- Couzijn, M. (1999). Learning to write by observation of writing and reading processes: effects on learning and transfer. *Learning and Instruction*, 9(2), 109-142. doi: 10.1016/S0959-4752(98)00040-1.
- Dalgarno, B. (2001). Interpretations of constructivism and consequences for Computer Assisted Learning. *British Journal of Educational Technology*, 32(2), 183-194. doi: 10.1111/1467-8535.00189.
- Davin, K. J. (2013). Integration of dynamic assessment and instructional conversations to promote development and improve assessment in the language classroom. *Language Teaching Research*, 17(3), 303-322. doi: 10.1177/1362168813482934.
- De Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. Psychology Press.
- De la Colina, A. A., & García Mayo, M. d. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. d. P. García Mayo (Ed.), *Investigating Tasks in Formal Language Learning* (pp. 91-116). Clevedon, NY, Ontario: Multilingual Matters Ltd.
- De La Paz, S. (2007). Managing cognitive demands for writing: Comparing the effects of instructional components in strategy instruction. *Reading & Writing Quarterly*, 23(3), 249-266. doi: 10.1080/10573560701277609.
- de Smet, M. J. R., Broekkamp, H., Brand-Gruwel, S., & Kirschner, P. A. (2011). Effects of electronic outlining on students' argumentative writing performance. *Journal of Computer Assisted Learning*, 27(6), 557-574. doi: 10.1111/j.1365-2729.2011.00418.x.
- Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In F. Fischer, I. Kollar, H. Mandl & J. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 275-301): Springer US.

- Dodigovic, M. (2002). Developing writing skills with a Cyber-Coach. *Computer Assisted Language Learning*, 15(1), 9-25.
- Dodigovic, M. (2005). *Artificial intelligence in second language learning: Raising error awareness*. Clevedon: Multilingual Matters.
- Donato, R. (2000). Sociocultural contributions to understanding the foreign and second language classroom. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 27-50). Oxford: Oxford University Press.
- Donato, R., & MacCormick, D. (1994). A sociocultural perspective on language learning strategies: The role of mediation. *The Modern Language Journal*, 78(4), 453-464. doi: 10.2307/328584.
- Doolittle, P. E., Hicks, D., & Triplett, C. F. (2006). Reciprocal teaching for reading comprehension in Higher Education: a strategy for fostering the deeper understanding of texts. *International Journal of Teaching and Learning in Higher Education*, 17(2), 106-118.
- Driscoll, M. P. (2000). *Psychology of learning for instruction*. Boston: Allyn and Bacon.
- Ede, L. S., & Lunsford, A. A. (1990). *Singular texts/plural authors: Perspectives on collaborative writing*. Carbondale: Southern Illinois University Press.
- Egbert, J. L. (2005). Conducting research on CALL. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 3-8). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ehrman, M., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System*, 31(3), 393-415. doi: 10.1016/S0346-251X(03)00050-2.
- Elen, J. (2013). "Instructional disobedience": Challenging instructional design research. Paper presented at the Earli 15th biannual conference, München, Germany.
- Ellis, R. (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly*, 29(1), 87-105.
- Ellis, R. (2008). *The study of second language acquisition* (Second edition). Oxford: Oxford University Press.
- Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System*, 46(0), 80-95. doi: 10.1016/j.system.2014.07.011.
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.
- Engeström, Y. (2001). Expansive learning at work: toward an activity theoretical reconceptualization. *Journal of Education and Work*, 14(1), 133-156. doi: 10.1080/13639080020028747.
- Engeström, Y., Miettinen, R., & Punamäki, R.-L. (Eds.) (1999). *Perspectives on activity theory*. Cambridge: Cambridge University Press.
- Erkens, G., Jaspers, J., Prangmsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative writing. *Computers in Human Behavior*, 21(3), 463-486. doi: 10.1016/j.chb.2004.10.038.
- Espasa, A., Guasch, T., & Alvarez, I. (2013). Analysis of feedback processes in online group interaction: a methodological model. *Digital Education Review*, (23), 59-73.
- Espasa, A., & Meneses, J. (2010). Analysing feedback processes in an online teaching and learning environment: an exploratory study. *Higher Education*, 59(3), 277-292. doi: 10.1007/s10734-009-9247-4.
- Evans, N. W., James Hartshorn, K., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System*, 39(2), 229-239. doi: 10.1016/j.system.2011.04.012.
- Faigley, L., & Witte, S. P. (1981). Analyzing revision. *College Composition and Communication*, 32, 400-414.
- Felix, U. (2002). The web as a vehicle for constructivist approaches in language teaching. *ReCALL*, 14(1), 2-15.

- Felix, U. (2005). What do meta-analyses tell us about CALL effectiveness? *ReCALL*, 17(02), 269-288. doi:10.1017/S0958344005000923.
- Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, 21(1), 40-58. doi: 10.1016/j.jslw.2011.12.002.
- Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32 (Special Issue 02), 181-201. doi:10.1017/S0272263109990490.
- Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Hillsdale, NJ: Erlbaum.
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365-387.
- Foster, P., & Ohta, A. S. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, 26(3), 402-430.
- Galbraith, D. (2009). Cognitive models of writing. *German as a foreign language*(2-3), 7-22.
- Galbraith, D., & Rijlaarsdam, G. (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction*, 9(2), 93-108. doi: 10.1016/S0959-4752(98)00039-5.
- Gallien, T., & Oomen-Early, J. (2008). Personalized versus collective instructor feedback in the online courseroom: Does type of feedback affect student satisfaction, academic performance and perceived connectedness with the instructor? *International Journal on E-Learning*, 7(3), 463-476.
- Garrett, N. (1987). A psycholinguistic perspective on grammar and CALL. In W. F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation* (pp. 169-196). Lincolnwood, IL: National Textbook.
- Gaver, W. W. (1996). Situating action II: Affordances for interaction: The social is material for design. *Ecological Psychology*, 8(2), 111-129.
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology*. Hillsdale, NJ: Erlbaum.
- Glendinning, E., & Howard, R. (2003). Lotus ScreenCam as an aid to investigating student writing. *Computer Assisted Language Learning*, 16(1), 31-46.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. New York: Longman.
- Grabowski, J., Becker-Mrotzek, M., Knopp, M., Jost, J., & Weinzierl, C. (2014). Comparing and combining different approaches to the assessment of text quality. In C. Heine, D. Knorr & J. Engberg (Eds.), *Methods in writing process research*. Frankfurt / Main: Peter Lang Verlag.
- Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187-207). New York / London: Guilford Press.
- Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476.
- Graham, S., & Perin, D. (2007b). What we know, what we still need to know: Teaching adolescents to write. *Scientific Studies of Reading*, 11(4), 313-335.
- Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(2), 165-198. doi: 10.1017/S0958344013000013.
- Hämäläinen, R., & Arvaja, M. (2009). Scripted collaboration and group-based variations in a Higher Education CSCL context. *Scandinavian Journal of Educational Research*, 53(1), 1-16. doi: 10.1080/00313830802628281.

- Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research*, 11(4), 459-479. doi: 10.1177/1362168807080963.
- Hanjani, A. M., & Li, L. (2014). Exploring L2 writers' collaborative revision interactions and their writing performance. *System*, 44(0), 101-114. doi: 10.1016/j.system.2014.03.004.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi: 10.3102/003465430298487.
- Hayes, J. R. (2006). New directions in writing theory. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 28-40). New York / London: Guilford Press.
- Hayes, J. R., & Flower, L. (1980). Identifying the organization of the writing process. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.
- Hémard, D. (2003). Language learning online: Designing towards user acceptability. In U. Felix (Ed.), *Language learning online: Towards best practice* (pp. 21-43). Lisse: Swets & Zeitlinger.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, 56(4), 473-493.
- Hoadley, C. P. (2002). *Creating context: Design-based research in creating and understanding CSCL*. Paper presented at the Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473. doi: 10.1093/applin/amp048.
- Howard, R. M., Serviss, T., & Rodrigue, T. K. (2010). Writing from sources, writing from sentences. *Writing and Pedagogy*, 2(2), 177-192.
- Hsu, H.-Y., & Wang, S. (2010). The impact of using blogs on college students' reading comprehension and learning motivation. *Literacy Research and Instruction*, 50(1), 68-88.
- Hubbard, P. (1988). An integrated framework for CALL courseware evaluation. *CALICO Journal*, 6(2), 51-72.
- Hubbard, P., & Siskin, C. B. (2004). Another look at tutorial CALL. *ReCALL*, 16(02), 448-461. doi: 10.1017/S0958344004001326.
- Hung, H. T. (2011). Design-based research: designing a multimedia environment to support language learning. *Innovations in Education and Teaching International*, 48(2), 159-169.
- Ivanic, R. (2004). Intertextual practices in the construction of multimodal texts in inquiry-based learning. *Uses of intertextuality in classroom and educational research*, 279-314.
- Jalkanen, J. (2013). *Dynamics of sustainable pedagogical development: Insights from higher education language teaching*. Paper presented at the Global perspectives on Computer-Assisted Language Learning, Glasgow.
- Jamieson, J., & Chapelle, C. A. (2010). Evaluating CALL use across multiple contexts. *System*, 38(3), 357-369. doi: 10.1016/j.system.2010.06.014.
- Johnson, K. E. (2006). The sociocultural turn and its challenges for second language teacher education. *TESOL Quarterly*, 40(1), 235-257.
- Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, 45(1), 65-94.
- Karagiorgi, Y., & Symeou, L. (2005). Translating constructivism into instructional design: Potential and limitations. *Educational Technology & Society*, 8(1), 17-27.

- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25(0), 4-22. doi: 10.1016/j.jslw.2014.05.005.
- Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 355-365.
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1), 1-26.
- Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology*, 44(1), 78-94.
- Kinzie, M. B., Sullivan, H. J., & Berdel, R. L. (1988). Learner control and achievement in science computer-assisted instruction. *Journal of Educational Psychology*, 80(3), 299.
- Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25(1), 105-121.
- Kirschner, P. A. (2002). Can we support CSCL? Educational, social and technological affordances for learning. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 7-47). Heerlen: Open University of the Netherlands.
- Kirschner, P. A., Martens, R. L., & Strijbos, J.-W. (2004). CSCL in higher education? A framework for designing multiple collaborative environments. In J.-W. Strijbos, P. Kirschner & R. L. Martens (Eds.), *What We Know About CSCL: And Implementing It In Higher Education* (pp. 3-30). Dordrecht: Kluwer Academic Publishers Group.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284. doi: 10.1037/0033-2909.119.2.254.
- Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58(1), 79-97. doi: 10.3102/00346543058001079.
- Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly*, 35(4), 537-560.
- Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. New York: Routledge.
- Kuteeva, M. (2011). Wikis and academic writing: Changing the writer-reader relationship. *English for Specific Purposes*, 30(1), 44-57. doi: 10.1016/j.esp.2010.04.007.
- Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*. doi: 10.1093/applin/amu047.
- Lantolf, J. P. (Ed.). (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.
- Lantolf, J. P., & Pavlenko, A. (2001). (S)econd (L)anguage (A)ctivity theory: Understanding second language learners as people. In M. Breen (Ed.), *Learner contributions to language learning: New directions in research* (pp. 141-158). Essex: Pearson Education Limited.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- Leakey, J. (2011). *Evaluating computer-assisted language learning: an integrated approach to effectiveness research in CALL*. Bern: Peter Lang.
- Lei, X. (2009). *Understanding writing strategy use from a sociocultural perspective: a multiple-case study of Chinese EFL learners of different writing abilities*. Unpublished doctoral dissertation, University of Hong Kong, Hong Kong. Retrieved from <http://hdl.handle.net/10722/56573>.
- 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.

- Lenhard, W., Baier, H., Endlich, D., Schneider, W., & Hoffmann, J. (2013). Rethinking strategy instruction: direct reading strategy instruction versus computer-based guided practice. *Journal of Research in Reading, 36*(2), 223-240. doi: 10.1111/j.1467-9817.2011.01505.x.
- Leontiev, A. N. (1978). *Activity, consciousness and personality*. Englewood Cliffs, NJ: Prentice Hall.
- Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford: Oxford University Press.
- Levy, M. (2013). Design-based research and the quest for normalization in CALL. In J. C. Rodriguez & C. Pardo-Ballester (Eds.), *Design-based research in CALL* (11 ed., pp. 31-40).
- Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. New York / London: Routledge.
- Lodewyk, K. R., Winne, P. H., & Jamieson-Noel, D. L. (2009). Implications of task structure on self-regulated learning and achievement. *Educational Psychology, 29*(1), 1-25. doi: 10.1080/01443410802447023.
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126-141.
- Lund, A. (2008). Wikis: a collective approach to language production. *ReCALL, 20*(01), 35-54. doi: 10.1017/S0958344008000414.
- Lyon, C., Barrett, R., & Malcolm, J. (2004). *A theoretical basis to the automated detection of copying between texts, and its practical implementation in the Ferret plagiarism and collusion detector*. Paper presented at the Plagiarism: Prevention, Practice and Policies 2004 conference.
- Manchón, R. M. (2012). Multiple perspectives in the study of L2 writing development: An introduction. In R. Manchón (Ed.), *L2 writing development: Multiple perspectives*. (pp. 1-16). Boston / Berlin: Walter de Gruyter.
- Manchón, R. M. (2011). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3-16). Amsterdam: John Benjamins.
- Martínez Esteban, N., & Roca de Larios, J. (2010). The use of models as a form of written feedback to secondary school pupils of English. *International Journal of English Studies, 10*(2), 143-170.
- Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. (ED 427 087). Aurora, CO: Eric Document Reproduction Service.
- McAndrew, P., Taylor, J., & Clow, D. (2006). Methods for evaluating learning, collaboration and technology use in distributed virtual environments and mobile environments: Supporting first-aider training. The Open University-Knowledge Network. Retrieved from <http://kn.open.ac.uk/public/getfile.cfm?documentfileid=10015>.
- McNamara, D., Crossley, S., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods, 45*(2), 499-515. doi: 10.3758/s13428-012-0258-1.
- Mehanna, W. N. (2004). e-Pedagogy: the pedagogies of e-learning. *Research in Learning Technology, 12*(3), 279-293. doi: 10.1080/0968776042000259582.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 2*(1), 63-86. doi: 10.1007/s11412-006-9005-x.
- Müller-Hartmann, A., & Schocker-v.Ditfurth, M. (2010). Research on the use of technology in task-based language teaching. In M. Thomas & H. Reinders (Eds.), *Task-based language learning and teaching with technology* (pp. 17-40). London / New York: Continuum International.

- Murphy, E. (2009). Online synchronous communication in the second-language classroom. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, 35(3).
- Murphy, P. (2010). Web-based collaborative reading exercises for learners in remote locations: the effects of computer-mediated feedback and interaction via computer-mediated communication. *ReCALL*, 22(02), 112-134. doi: 10.1017/S0958344010000030.
- Murray, L., & Barnes, A. (1998). Beyond the "wow" factor - Evaluating multimedia language learning software from a pedagogical viewpoint. *System*, 26(2), 249-259. doi: 10.1016/s0346-251x(98)00008-6.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In A. J. M. Spector, M. D. Merrill, J. Van Merriënboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology (Aect)* (pp. 125-144). New Jersey: Lawrence Erlbaum.
- Nicolás-Conesa, F., Roca de Larios, J., & Coyle, Y. (2014). Development of EFL students' mental models of writing and their effects on performance. *Journal of Second Language Writing*, 24(0), 1-19. doi: 10.1016/j.jslw.2014.02.004.
- Nielsen, J. (1993). *Usability engineering*. London: Academic Press.
- Norman, D. A. (1990). *The design of everyday things*. New York: Doubleday.
- Norman, D. A. (1999). Affordance, conventions, and design. *interactions*, 6(3), 38-43.
- Norris, J. M., & Manchón, R. (2012). Investigating L2 writing development from multiple perspectives: Issues in theory and research. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives*. (pp. 221-244). Boston / Berlin: Walter de Gruyter.
- Nystrand, M. (2006). The social and historical context for writing research. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 11-27). New York / London: Guilford Press.
- Opdenacker, L., Stassen, I., Vaes, S., Van Waes, L., & Jacobs, G. (2009). *Manual for the quality assessment of digital educational material*. Retrieved from http://www.scribani.org/quadem/material_Q/handbook/Handbook.pdf.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. (Vol. 1). New York: Newbury House.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601. doi: 10.1093/applin/amp045.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3), 265-289. doi: 10.1016/s1060-3743(99)80117-9.
- Pecorari, D. (2013). *Teaching to avoid plagiarism. How to promote good source use*. Berkshire: Open University Press.
- Pederson, K. (1987). Research on CALL. In W. F. Smith (Ed.), *Modern media in foreign language education: Theory and implementation. The ACTFL Foreign Language Education Series* (pp. 99-131). Lincolnwood, IL: National Textbook Company.
- Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition*, 6(02), 186-214.
- Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561-587. doi: 10.1177/0265532209340192.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer Publishing.
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment (C-DA). *Language Teaching Research*, 17(3), 323-342.
- Prensky, M. (2001). Digital natives, digital immigrants. A new way to look at ourselves and our kids. *On the Horizon*, 9(5), 1-6.

- Prichard, C. (2010). Using social bookmark sites for independent reading projects. *The JALT CALL Journal*, 6(2), 115-128.
- Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing. *Journal of Research in Reading*, 35(2), 136-152. doi: 10.1111/j.1467-9817.2010.01450.x.
- Raedts, M., Daems, F., Van Waes, L., & Rijlaarsdam, G. (2009). Observerend leren van peer models bij een complexe schrijftaak. *Tijdschrift voor Taalbeheersing*, 31(2), 142-165.
- Raedts, M., Rijlaarsdam, G., van Waes, L., & Daems, F. (2007). Observational learning through video-based models: impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In S. Hidi & P. Boscolo (Eds.), *Writing and motivation* (pp. 219-238). New York: Springer.
- Center for Activity Theory and Developmental Work Research (2003-2004). *Activity System*. Retrieved from <http://www.edu.helsinki.fi/activity/pages/chatanddwr/activitysystem/>.
- Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System*, 40(3), 421-430. doi: 10.1016/j.system.2012.07.005.
- Rodríguez, J. C., & Pardo-Ballester, C. (2013). *Design-based research in CALL*. San Marcos, TX: Computer Assisted Language Instruction Consortium.
- Roebuck, R. (2000). Subjects speak out: How learners position themselves in a psycholinguistic task. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 79-95). Oxford: Oxford University Press.
- Rosselle, M., Sercu, L., & Vandepitte, S. (2009). Learning outcomes and learner perceptions in relation to computer-based feedback. *Indian Journal of Applied Linguistics*, 35(1), 45-61.
- Rumelhart. (1977). Understanding and summarizing brief stories. In D. LaBerge & S. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 265-303). Hillsdale, NJ: Erlbaum.
- Rummel, N., & Spada, H. (2007). Can people learn computer-mediated collaboration by following a script? . In F. Fischer, I. Kollar, H. Mandl & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning* (pp. 39-55). Boston, MA: Springer Science+Business Media.
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, 4(1), 69-92. doi: 10.1007/s11412-008-9054-4.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9(3), 259-291. doi: 10.1016/S1060-3743(00)00028-x.
- Sasaki, M. (2009). Changes in English as a foreign language students' writing over 3.5 years: A sociocognitive account. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 49-76). Bristol: Multilingual Matters.
- Scardamalia, M., Bereiter, C., & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science*, 8(2), 173-190. doi: 10.1016/S0364-0213(84)80016-6.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.
- Schulze, M., Verspoor, M., Wood, P., & Pokorny, B. (unpublished manuscript). *Towards automatic proficiency scoring in L2 writing: Balanced complexity*.

- Seedhouse, P. (2005). Conversation analysis and language learning. *Language Teaching*, 38(04), 165-187.
- Segev-Miller, R. (2004). Writing from sources: The effect of explicit instruction on college students' processes and products. *L1 - Educational Studies in Language and Literature*, 4(1), 5-33. doi: 10.1023/B:ESLL.0000033847.00732.af.
- Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532. doi: 10.1093/applin/amp047.
- Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating information: An analysis of the processes involved and the products generated in a written synthesis task. *Written Communication*, 30(1), 63-90. doi: 10.1177/0741088312466532.
- Spivey, N. N. (1990). Transforming texts. Constructive processes in reading and writing. *Written Communication*, 7(2), 256-287.
- Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, 14(3), 153-173. doi: 10.1016/j.jslw.2005.05.002
- Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10(2), 29-46.
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Bristol / Buffalo / Toronto: Multilingual Matters.
- Strijbos, J. W., Martens, R. L., & Jochems, W. M. G. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. *Computers & Education*, 42(4), 403-424. doi: 10.1016/j.compedu.2003.10.004.
- Swain, M. (1997). Collaborative dialogue: Its contribution to second language learning. *Revista Canaria de Estudios Ingleses*, 34(115-32).
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. *Sociocultural theory and second language learning*, 97, 114.
- Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). London / New York: Continuum.
- Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer-peer dialogue as a means of second language learning. *Annual Review of Applied Linguistics*, 171-185.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Tapscott, D. (2008). *Grown up digital: How the net generation is changing your world*: McGraw-Hill.
- Thorne, S. L. (1999). *An activity theoretical analysis of foreign language electronic discourse*. Unpublished PhD, University of California, Berkeley.
- Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology*, 7(2), 38-67.
- Thorne, S. L., & Reinhardt, J. (2008). "Bridging activities", new media literacies, and advanced foreign language proficiency. *CALICO Journal*, 25(3), 558-572.
- Thorson, H. (2000). Using the computer to compare foreign and native language writing processes: A statistical and case study approach. *The Modern Language Journal*, 84(2), 155-169.
- Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67-80). New York / London: Guilford Press.

- Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education*, 27(3), 379-392. doi: 10.1007/BF03179901.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327-369.
- Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing*, 5(3), 271-294. doi: 10.1016/S1060-3743(96)90005-3.
- Van Beuningen, C. (2010). Corrective feedback in L2 writing. Theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, 10(2), 1-28.
- van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272. doi: 10.1016/j.compedu.2011.07.020.
- van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245-259). Oxford: Oxford University Press.
- van Lier, L. (2004). *The ecology and semiotics of language learning. A sociocultural perspective*. Boston / Dordrecht / New York / London: Kluwer Academic
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20(4), 316-327. doi: 10.1016/j.learninstruc.2009.08.009.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Mass: Harvard University Press.
- Wang, C. X., Song, H., Stone, D. E., & Yan, Q. (2009). Integrating Second Life into an EFL program in China: Research collaboration across the continents. *TechTrends*, 53(6), 14-19.
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology, Research and Development*, 53(4), 5-23.
- Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia language teaching* (pp. 3-20). Tokyo: Logos International.
- Warschauer, M. (2005). Sociocultural perspectives on CALL. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 41-52). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26(3), 445-466. doi: 10.1177/0265532209104670.
- Wigglesworth, G., & Storch, N. (2012). Feedback and writing development through collaboration: A socio-cultural approach. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives*. (pp. 69-97). Boston / Berlin: Walter de Gruyter.
- Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72(2), 131-175.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: measures of fluency, accuracy, & complexity*. Hawaii: Univ. of Hawaii Press.
- Wolfersberger, M. (2007). *Second language writing from sources: An ethnographic study of an argument essay task*. Unpublished PhD, University of Auckland. Retrieved from https://researchspace.auckland.ac.nz/bitstream/handle/2292/2364/02whole.pdf?sequence=19&origin=publication_detail.

- Wolsey, T. D. (2008). Efficacy of instructor feedback on written work in an online program. *International Journal on ELearning*, 7(2), 311-329.
- Wong, L. H., & Looi, C. K. (2010). Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: two case studies. *Journal of Computer Assisted Learning*, 26(5), 421-433.
- Yasuda, S. (2005). Different activities in the same task: An activity theory approach to ESL students' writing process. *JALT JOURNAL*, 27(2), 139-168.
- Yutdhana, S. (2005). Design-based research in CALL. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 169-178). Mahwah, NJ: Lawrence Erlbaum Associates.
- Zhao, Y. (2003). Recent developments in technology and language learning: a literature review and meta-analysis. *CALICO Journal*, 21(1), 7-27.
- Zhou, A. A. (2009). What adult ESL learners say about improving grammar and vocabulary in their writing for academic purposes. *Language Awareness*, 18(1), 31-46. doi: 10.1080/09658410802307923.
- Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22(1), 73-101. doi: 10.1006/ceps.1997.0919.

