Using computers and digital tools as a way to support workplace tasks is ubiquitous today, and it is not really a concept that initially strikes one as a novelty. However, the digital tools available to us are not always intended for the tasks that we would actually want to get help with. In this dissertation, I have explored whether or not a digital tool could be a good fit for helping language teachers with the spike in workload experienced during periods surrounding written submissions. My area of focus has been the upper-secondary level English language courses offered within the Swedish school system.

The idea of a supporting tool for English as an Additional Language teachers was explored in the context of Second Language Acquisition models, and the design of the resulting tool was conducted based on interviews with in-service teachers. The tool was then evaluated together with in-service teachers. Specific areas of interest were the functionality needed to support the teachers in their reading, the established routines that the tool would have to fit within, and, of course, the changes in workload experienced as a result of implementation.
Automated Text Analytics in Swedish English Classrooms

The Design, Development, and Evaluation of a Supporting Tool
AUTOMATED TEXT ANALYTICS IN SWEDISH ENGLISH CLASSROOMS

The Design, Development, and Evaluation of a Supporting Tool

DANIEL OCIC IHRMARK

LINNAEUS UNIVERSITY PRESS
Automated Text Analytics in Swedish English Classrooms: The Design, Development, and Evaluation of a Supporting Tool
Doctoral Dissertation, Department of Languages, Linnaeus University, Växjö, 2023

ISBN: 978-91-8082-011-0 (print), 978-91-8082-012-7 (pdf)
© Daniel Ocic Ihrmark
DOI: https://doi.org/10.15626/LUD.486.2023
Published by: Linnaeus University Press, 351 95 Växjö
Printed by: Exakta Print AB, 2023
Abstract


The scope of this dissertation is the design, development, and evaluation of a focus-on-form diagnostics tool in the Swedish upper-secondary English subject context. The use of corrective feedback on student texts has been discussed extensively in previous research, but the implementation of an automated diagnostics tool as a support for the practice less so.

The teachers' needs are explored through a survey with the goal of gaining insight into problems learners commonly have when attaining grammar, vocabulary, and spelling as experienced by in-service teachers. The focus is on how a digital diagnostic tool could fit within the teachers’ current routines. By combining the survey results with a review of previous research, a design specification for the tool is produced. The tool is then evaluated in teaching practice as a way of exploring the possible benefits and issues of implementing automated language diagnostics as a support for teachers. The evaluation data was produced through interventions and debriefing interviews with the participating teachers.

The results of the survey on teachers' needs indicated that any automated attempt at supporting language diagnostics within the school context would have to be able to identify a series of error patterns, such as sentence structure, punctuation, verb tenses, subject-verb agreement, article usage, and spelling. In addition, the tool would have to fit within an already established practice aimed at providing learners with the necessary support for them to succeed, while also avoiding pitfalls such as creating avoidance issues or overwhelming the students. The results of the intervention study used to evaluate the tool indicated that the respondents found it likely that the implementation of the tool would result in a higher workload initially, but would be beneficial long-term. In terms of fit with current routines and ability to identify relevant language patterns, the tool was experienced as conceptually sound. However, some issues, such as the limited target variant, technical issues surrounding on-site implementation, and the overwhelming form of the output data, remain.
The use of automated text analysis for the focus-on-form reading performed by the teachers was found to provide functional support for them. Based on the results, automated diagnostics tools are indicated as being a fitting venue for future research and development.

**Keywords:** SLA, EAL, CALL, DDL, Educational Linguistics, English, Upper-secondary, Education Technology, Language Diagnostics, Tool
# Contents

1. Introduction .................................................................................................................. 1
   1.1 The Disposition of this Dissertation ................................................................. 3
2. Methodology .................................................................................................................. 8
   2.1 A Pragmatic Approach ....................................................................................... 8
   2.2 A Question of Ecological Validity ...................................................................... 10
      2.2.1 The Diagnostic Needs of In-Service EAL Teachers .................................... 11
      2.2.2 Available Tools and Methods .................................................................... 12
      2.2.3 Teachers’ Evaluations of the Tool ............................................................. 13
   2.3 Survey Analysis Method ...................................................................................... 13
   2.4 Interview Design ................................................................................................. 15
      2.4.1 Thematizing the Interviews ........................................................................ 15
      2.4.2 Study Design ............................................................................................... 16
      2.4.3 Sampling and Respondent Recruitment ...................................................... 17
      2.4.4 The Interview Guide ................................................................................... 19
      2.4.5 Transcription and Analysis ........................................................................ 20
      2.4.6 Verifying the Data ..................................................................................... 23
   2.5 The Respondent Drawing ..................................................................................... 24
   2.6 Ethical Protocol ..................................................................................................... 26
3. Background .................................................................................................................... 30
   3.1 The English Language Subject in the Swedish Context ..................................... 30
   3.2 Educational Linguistics ....................................................................................... 34
   3.3 Second Language Acquisition ............................................................................ 38
      3.3.1 The Innatist Model of SLA ....................................................................... 39
      3.3.2 The Cognitive Model of SLA .................................................................... 42
      3.3.3 The Social Constructivist Model of SLA ................................................... 45
      3.3.4 SLA in the Current Project ........................................................................ 47
   3.4 L2 Writing ............................................................................................................. 48
   3.5 Corrective Feedback ............................................................................................. 50
   3.6 Computer-Assisted Language Teaching ............................................................ 57
   3.7 Teacher-Computer Interaction .......................................................................... 61
   3.8 Background Summary ......................................................................................... 65
4. Interviewing EAL Teachers Regarding Diagnostic Needs ........................................ 68
   4.1 Interview Analysis .............................................................................................. 68
      4.1.1 Analysis: Pilot Interviews ........................................................................... 69
6.2.1 Quincy ........................................................................................................... 161
6.2.2 Brita ............................................................................................................... 167
6.2.3 SUS and SF UES at University ................................................................. 173
6.2.4 Themes at University ............................................................................... 174
6.3 Conclusion of Intervention Studies ............................................................... 177
7. Results ................................................................................................................. 179
  7.1 Swedish EAL Teachers’ Diagnostic Needs ............................................. 179
  7.2 Tool Development ....................................................................................... 180
  7.3 Intervention Studies ..................................................................................... 182
8. Discussion ............................................................................................................. 184
  8.1 Interpretation of Results .............................................................................. 184
  8.2 Methodological Considerations ................................................................. 195
  8.3 Limitations ..................................................................................................... 196
  8.4 Lessons Learned ........................................................................................... 197
     8.4.1 Mistakes .................................................................................................. 197
     8.4.2 Strengths .................................................................................................. 198
     8.4.3 Weaknesses .............................................................................................. 198
9. Conclusion ............................................................................................................. 200
  9.1 Contributions .................................................................................................. 202
  9.2 Future Research ............................................................................................. 202
  9.3 Future Development ....................................................................................... 202
10. Sammanfattning på svenska ............................................................................ 204
  10.1 Inledning ....................................................................................................... 204
  10.2 Metod ............................................................................................................ 205
  10.3 Engelsklärares diagnostiska behov i svensk gymnasieskola .......... 206
  10.4 Verktygsdesign ............................................................................................ 207
  10.5 Interventioner ............................................................................................... 209
  10.6 Slutsatser ...................................................................................................... 210
Works Cited ............................................................................................................ 213
Appendices: ........................................................................................................... 227
Appendix A: Interview guide 1 ............................................................................. 227
Appendix B: Consent form 1 ................................................................................. 229
Appendix C: Quick start guide ............................................................................. 233
Appendix D: User testing invitation ..................................................................... 238
Appendix E: Interview guide 2 ............................................................................. 239
Appendix F: Consent form 2.................................................................................. 241
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAL</td>
<td>American Association for Applied Linguistics</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>CALL</td>
<td>Computer-Assisted Language Learning</td>
</tr>
<tr>
<td>CEFR</td>
<td>Common European Framework of Reference</td>
</tr>
<tr>
<td>CF</td>
<td>Corrective Feedback</td>
</tr>
<tr>
<td>DDL</td>
<td>Data-Driven Learning</td>
</tr>
<tr>
<td>EAL</td>
<td>English as an Additional Language</td>
</tr>
<tr>
<td>ECTS</td>
<td>European Credit Transfer and Accumulation System</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a Foreign Language</td>
</tr>
<tr>
<td>ESOL</td>
<td>English for Speakers of Other Languages</td>
</tr>
<tr>
<td>GDPR</td>
<td>General Data Protection Regulation</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-Computer Interaction</td>
</tr>
<tr>
<td>IELTS</td>
<td>International English Language Testing System</td>
</tr>
<tr>
<td>JDK</td>
<td>Java Development Kit</td>
</tr>
<tr>
<td>JLT</td>
<td>Java LanguageTool</td>
</tr>
<tr>
<td>LW</td>
<td>Learning-to-Write</td>
</tr>
<tr>
<td>L2 Writing</td>
<td>Second Language Writing</td>
</tr>
<tr>
<td>NLTK</td>
<td>Natural Language ToolKit</td>
</tr>
<tr>
<td>OECD</td>
<td>The Organization for Economic Cooperation and Development</td>
</tr>
<tr>
<td>POS</td>
<td>Part-Of-Speech</td>
</tr>
<tr>
<td>RAM</td>
<td>Random Access Memory</td>
</tr>
<tr>
<td>SAMR</td>
<td>Substitution, Augmentation, Modification, and Redefinition model</td>
</tr>
<tr>
<td>SF UES</td>
<td>Short Form User Engagement Scale</td>
</tr>
<tr>
<td>SLA</td>
<td>Second Language Acquisition</td>
</tr>
<tr>
<td>SNAE</td>
<td>Swedish National Agency of Education</td>
</tr>
<tr>
<td>SUS</td>
<td>Short-form Usability Scale</td>
</tr>
<tr>
<td>TCI</td>
<td>Teacher-Computer Interaction</td>
</tr>
<tr>
<td>TESOL</td>
<td>Teachers of English to Speakers of Other Languages</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TPCK</td>
<td>Technological pedagogical content knowledge</td>
</tr>
<tr>
<td>UCD</td>
<td>User-Centered Design</td>
</tr>
<tr>
<td>UX</td>
<td>User Experience</td>
</tr>
<tr>
<td>VFU</td>
<td>Verksamhetsförlagd Utbildning (In-service Practice Placement)</td>
</tr>
<tr>
<td>WLC</td>
<td>Writing-to-learn-Content</td>
</tr>
<tr>
<td>WLL</td>
<td>Writing-to-learn-Language</td>
</tr>
<tr>
<td>ZPD</td>
<td>Zone of Proximal Development</td>
</tr>
</tbody>
</table>

Demo output site using Mimer sample:  
[https://eightypercentdemo.herokuapp.com/](https://eightypercentdemo.herokuapp.com/)

GitHub repository and access to tool:  
[https://github.com/Daniellhrmark/dissertationsoftware](https://github.com/Daniellhrmark/dissertationsoftware)
Acknowledgements

I have been a student at the Department of Languages at LNU since the autumn of 2012, and it feels somewhat fitting to close out my decade here with the publication of this thesis. This place has given me the opportunity to not only grow as a researcher during my PhD studies, but also as a language teacher. I have learned so much by being allowed to teach together with you all over the last couple of years, and I have had a very good time doing so. This also extends to Kora and Ahmad, who have let me talk about some of my favorite topics with their Digital Humanities students.

The research that went into this dissertation has been a mainly pleasant experience, although the pandemic made for a bit of a rough patch. I have been very lucky in terms of my supervisors, and Jukka and Marie have been wonderful discussion partners both on the topic of the dissertation and in general. Thank you both for all of the time and effort you have spent on this! Ylva has also spent an incredible amount of time helping me figure out the admin side of the whole PhD thing, and a myriad of other things. Thank you for keeping us on track!

I also want to thank the EdLing group at LNU for the discussions, feedback and all of the fika. By my estimate we have had somewhere around 100 fikas or so together at this point, which is a respectable number. I would also like to acknowledge the support I have had from my fellow doctoral students in both the Department of Languages and the Department of Swedish. These types of projects can feel very lonely and disorientating at times, and life is a lot easier when you can go talk to other people going through the same thing. It makes me very happy that we have been steadily growing in number.

I am running out of space on the page, but I still have far too many people that I would like to thank for their involvement and support throughout this entire process. I want to especially thank Jessica, Lene, and Shirley for providing thorough and constructive feedback on my work at the 50% seminar, 80% seminar, and examining review respectively. I hope this final version is to your liking (if not, mail me and I will revise it again).

It is so weird how these acknowledgment sections often mention family and friends at the end, even though they are probably the most important people in our support networks. I never would have attempted this if my parents had not supported me trying weird things throughout my life, so thank you, mom and dad, for doing so. Maria, you are the best, and I am rushing through this so that I can come home to you a bit quicker. Johan, Maria falls asleep at 21:00, I will see you at the pub around 21:30.

Växjö, March 2023
Daniel Ihrmark
1. Introduction

This dissertation explores the design, application, and evaluation of a digital tool as a supporting aid for teachers of English in Swedish upper-secondary classrooms. This is accomplished through the creation of a tool that uses corpus linguistic resources to provide focus-on-form analysis of student texts in accordance with the diagnostic needs of in-service teachers as depicted through an interview study and previous research. Research on the topic also suggests that the focus-on-form reading takes up a considerable amount of time and effort, while also being difficult to address in planning and practice due to the workload experienced by many teachers (Ferris, 1999).

The use of corrective feedback on student texts has been discussed extensively in previous research over the last decades, for instance by Semke (1984), Sheppard (1992), Truscott (1996), Sheen (2007), and Ferris (2010). While the research community has not arrived at a conclusion regarding its use and efficiency, Ferris has proposed criteria for successful implementations. However, these criteria are based on experiences and observations gained from teaching university students and often come from classrooms in inner-circle English-speaking nations.

The main contribution of this dissertation is creating an understanding of how international research on corrective feedback on student texts fits within the Swedish context and creating a starting point for automated support tools for use in upper-secondary English as an Additional Language (EAL) classrooms within the Swedish school system. The digitalization of the school environment is a matter of fact at this point (Nationell digitaliseringsstrategi för skolväsendet, 2017), but it has mainly been engaged with developing digital aides for the students and general demands for “digital competencies” amongst teachers (Eurydice, 2019). While this is important, the continued support of teachers in their craft through innovations should also be considered a central venue for improvement, per the Education Act’s (SFS., 2010) requirements for research informing practice.

The questions answered in the current project are:
1. What does previous research on corrective feedback in language teaching indicate as important to the design of a digital tool for EAL teachers in Sweden?
2. What actions must the tool be able to carry out in order to provide aid to Swedish EAL teachers?
3. Which practices and routines must the tool fit within in order to be implemented?
4. How is a tool designed based on these questions experienced by in-service EAL teachers in Sweden?

These are addressed through the design and development of a digital tool based on the results of questions 1, 2, and 3. Question 4 is then aimed at evaluating the resulting tool in practice, with a focus on the key areas highlighted by the previous questions. The evaluation is intended to highlight the possible use of these types of tools as an answer to the issues to do with the time and effort demanded by focus-on-form reading, and the resulting high workload imposed on teachers.

The scope of this dissertation is focus-on-form, or grammatical, diagnostics of written EAL production, but the introduction of digital, automated supporting tools for language teachers is new ground, in many ways. As such, the insights gained from the project do not only concern the topic of the EAL scope, but the wider concept of computer-aided language teaching in the classroom environment. As the number of qualified teachers in Sweden continues to be problematically low and the workload increases (Lärarnas Riksförbund, 2019; OECD, 2019), these types of solutions that can enable teachers to remain detail-oriented and engaged with individual student production in an efficient way are likely to become increasingly important.

Automating focus-on-form language analysis for EAL classroom purposes appears, conceptually, as a straightforward way of reducing the teachers’ workloads. Such a tool could also provide data that would benefit lesson planning, individual development planning, and identifying the appropriate levels of study materials and assignments. Furthermore, formative feedback based on automatically generated data could serve as a way of emphasizing progress as a motivator for students tied to concrete aspects of language development. This type of diagnostic data can also be used as a way of pinpointing small, concrete areas for struggling students to engage with. While there are previous automated solutions to assessment, such as Versant, these tend to focus on prompt-specific scoring and grading rather than providing information that could be used for formative feedback (Streeter et al., 2011). This change in focus of the automated support is both the main functional gap and the main research gap addressed in this dissertation. Previous efforts have explored similar approaches in other languages outside of the Swedish context,
for instance Galician (Gamallo Otero et al., 2015), but without in-practice teacher evaluations.

While students can benefit indirectly from the tool, the main target userbase for the current project is the teachers who will be providing the actual feedback to the students. To keep this at the forefront, the design process draws on User-Centered Design (UCD) (Interaction Design Foundation, 2019) and has a user experience focus (Earnshaw, Tawfik and Schmidt, 2018). This is accomplished by involving in-service EAL teachers throughout the project through participatory design (Hayes, 2011), starting with an interview study to create focal points for the tool development. The tool is built using the Python coding language and makes use of pre-existing software libraries and resources for language analysis, such as the Natural Language Toolkit (NLTK) and the Java LanguageTool (JLT).

In order to avoid going blindly into these new applications of digital resources in language classrooms, this dissertation roots its perspective in Educational Linguistics as an overarching field. Educational Linguistics could be briefly defined as applied linguistics in the broad sphere of education, dealing with language education, language in education, and language as practice within educational institutions and systems (Hult, 2008; Spolsky, 2008). An important aspect of Educational Linguistics for this project, and for the field in general, is the emphasis on the responsibilities surrounding the implementations of new concepts in school environments, such as the impact on student well-being and social effects. This provides a background against which possible detrimental side effects of new practices can be discussed. This dissertation takes this into consideration through the contextualization of the tool within second language acquisition theory, as well as through considering what the possible effects of larger-scale implementation could be based on the evaluations of the tool.

1.1 The Disposition of this Dissertation

Beyond this introductory chapter, the second chapter discusses the epistemology at play within the research project. The approach is based on Pragmatism as a mixed-methods framework (Feilzer, 2009). The chapter also introduces Ecological Validity as a central aspect in the research design (Schmuckler, 2001), as it pertains to both applying findings from the international research community in a Swedish context and the development of new tools for the educational system. The main framework afforded by Pragmatism in the current dissertation is the pragmatic inquiry as a guide for the research design (see 2.1). The design process for the actual tool is guided by the framework for User-Centered Design (UCD), as suggested by the Interaction Design Foundation (2019, see 3.7).

These frameworks interact with one another to provide cohesion to the project as a whole, but also as a way of making sense of how the different parts
of the project go together. The pragmatic inquiry informs how the study should be conducted from a research perspective, while the UCD framework dictates how the development of the tool should proceed in order to provide a beneficial product for its stakeholders, i.e., the teachers. The frameworks have also impacted the formulation of the research questions, as these build on each other in a similar way to the stages of the two processes. In practice, this means that the first research question informs the second research question, and so on. The last research question also serves to validate the results of the prior questions from both the research perspective, afforded by the pragmatic inquiry, and the user-experienced viability of the resulting tool, through the UCD perspective.

In Figure 1 on the next page, these frameworks have been visualized in the context of the research questions, study design, and corresponding chapters of the dissertation. The figure shows the progression of the dissertation project and disposition from left to right, and the outcomes of the frameworks from top to bottom. The intention is to showcase how the steps of the research design build on and inform one another throughout the progression of the project, while the realization of the epistemological paradigm takes place through the steps of the pragmatic inquiry being matched to the steps of the UCD framework, resulting in research questions based on the steps of both frameworks. It is important to note the overlapping of the steps in the different frameworks, as they do not match 1-to-1. This is also seen in the relationship between the research questions and the frameworks, as questions 2 and 3 cover several aspects of both the pragmatic inquiry and UCD framework. The research questions are then operationalized through focal areas and methodological choices, labeled in Figure 1 as study design. These are then tied to their corresponding chapters in the dissertation in order to emphasize the disposition as mirroring the progression of the study.
Figure 1: Overview of project progression and dissertation disposition
Chapters 1 and 2 have been omitted from Figure 1, as they serve as an introduction to the dissertation and the paradigm in which it takes place. This should not be seen as downplaying the importance of chapter 2, which provides the necessary context for the current thesis, but is rather a result of the overarching nature of its contents placing it outside of the more practice-oriented design of the study. In terms of its influence on the study design, the second chapter could be seen as being summarized by the pragmatic inquiry framework presented as a starting point at the top of the figure.

The third chapter provides an overview of the EAL subject in the Swedish school system as a context for the dissertation and comparisons to international frameworks of proficiency in order for the appropriate levels to be made clear. The chapter then moves on to a review of previous research on the main background topics: Educational Linguistics, Second Language Acquisition (SLA), Corrective Feedback in SLA, Computer Assisted Language Learning (CALL), and Human-Computer Interaction (HCI). As previously mentioned, Educational Linguistics provides the umbrella under which this project takes place, but the placement should more specifically be considered as the intersection between the other fields. The applications proposed for the tool designed in the project will stem from the sociocultural model of language acquisition with an emphasis on feedback and scaffolding, although other models are also discussed to provide an understanding of other acquisition models that might be at play within the educational context.

Chapter 4 presents an interview study on in-service upper-secondary English as an Additional Language (EAL) teachers’ diagnostic needs, feedback routines, and student text interactions. The interview study makes use of semi-structured interviews with the addition of a respondent drawing. The conclusions drawn from the interview study and relevant insights from the previous research are then carried into chapter 5, which details the design and development of the tool. This chapter also includes an evaluation of the tool in a controlled environment consisting of an EAL learner corpus built for the project. The material for the corpus was sourced from an online repository of student texts.

The dissertation then moves on to present results from classroom use of the tool in chapter 6. The interventions consist of usage reports from two teachers in upper-secondary EAL classrooms and reflective usage reports from two university level EAL instructors. The debriefs are constructed from the same interview instrument used for the initial survey of upper-secondary teachers’ routines and diagnostic needs. The upper-secondary participants in the intervention study had access to the tool for one semester, and university instructors used the tool during a short session and then discussed their experiences of the tool based on their teaching for the semester.
Chapter 7 presents a summary of results and chapter 8 a discussion of findings in context with previous research and the interpretations of the results as they related to the research questions. The discussion chapter also discusses the methodological considerations and limitations of the project. The conclusion in chapter 9 outlines the main findings and contributions of the dissertation and suggests future directions for research based on the results, as well as recommendations regarding future development based on the interventions. The 10th chapter provides a summary in Swedish.
2. Methodology

This chapter presents the foundations of the project by accounting for the epistemology on which it is based. Pragmatism is introduced as a guiding principle for the research design and application. Pragmatism is also explored as a framework for the mixed-methods approach within the current project.

The chapter also accounts for ecological validity as an important consideration throughout the project. After providing the definitions of ecological validity within the scope of the dissertation the research plan is presented step-by-step. By reflecting on the epistemological approach and the main driving concept of the research questions, a concrete overview regarding the scope, intent, and design of the project is presented. This is followed by an overview of the approach to analysis used for the interview data, and the formulation of the interview instrument.

2.1 A Pragmatic Approach

As this project uses a mixed-methods approach by combining qualitative and quantitative data, it has been placed within the pragmatic paradigm where the choice of method is ultimately decided by the practical needs involved in solving the problem (Feilzer, 2009, p. 2). Morgan describes the Pragmatic paradigm as the “paradigm of choices” due to the need for constant conscious choices involved in mixed-method research designs (Morgan, 2014, p. 8). This statement originally comes from Michael Quinn Patton (2015), who explored the application of mixed methods within the social sciences, moving away from the original, much broader, sense of the Pragmatic inquiry introduced by John Dewey (1997) in How We Think (Morgan, 2014, p. 46).

The project could be further classified as an inquiry within the Pragmatic paradigm, as an “explicit attempt to produce new knowledge by taking actions and experiencing their results”, fulfilling the criterion presented by Morgan (2014, p. 8). Inquiry within the Pragmatic paradigm is further described as “confronting situations that fall outside your existing knowledge…” and adding new knowledge through the results experienced in order to navigate the new situation in the future (Morgan 2014, p. 7). In essence, the Pragmatic paradigm builds on pragmatism as a social philosophy, where actions lead to experiences that accumulate into beliefs that guide future actions that lead to new experiences and so forth (Morgan, 2014, p. 43). As the outcome of this project is the creation of a new interaction through the implementation of a digital support tool in the practice of EAL-teachers, and the goals are centered around an evaluation of the results of the implementation within the teaching practice, the situation described by Morgan seems well fitting. One central component
that comes into play through the texts of Dewey, Morgan, and Feilzer is the concept of the “warranted belief”.

A warranted belief is based on the researcher’s repertoire of knowledge going into the research and dictates the choices made in the research design. As it plays into the formulation of research questions, method choices, and focal points, it is tightly connected to the beliefs of the field at the time of research formulation. While describing the history of pragmatism, Richard Ormerod succinctly summarizes the relationships as: “beliefs are guides to actions and should be judged against the outcomes rather than abstract principles” (2006, p. 1).

In his exemplification of the problem-solving process according to the pragmatic perspective, Dewey provides five steps where active choices decide the final product of the process.

![Figure 2: Dewey’s five steps of inquiry (adapted from Morgan, 2014, p.8)](image)

These five steps would then fit into creating and formulating the beliefs that guide the initial research design, while the research is then carried out as an action based on those beliefs. Following Ormerod’s description, this action must also be designed to make the correctness of those beliefs visible by its outcome.

In application, however, the Pragmatic paradigm has developed vastly since Dewey’s initial definition. In Feilzer’s 2009 article “Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm”, pragmatism is explored as mixed-methods research while put in the context of (post)positivism and constructivism/interpretivism. These two perspectival contexts make up the sides in what Feilzer dubs the “Paradigm Wars”, on which the often-strict divide between quantitative and qualitative approaches is based (2009, p. 1). This is a deceitfully simple description, or as Feilzer puts it: “in a (admittedly rather simplistic) nutshell”, but it is a necessary starting point for a discussion on modern pragmatism.

While positivism and constructivism enforce rather strict views on the nature of truth, pragmatism has been described as sidestepping “the contentious issues of truth and reality” in order to engage more directly with the problem at the core of the research being conducted (Feilzer 2010 p. 8). As such, the
epistemological implications of the pragmatic paradigm are profound, yet difficult to define. As a natural effect of mixed-method research, the constraints of paradigms related heavily to either the qualitative or the quantitative become troublesome. While several attempts have been made to formulate a framework more fitting with the practices of employing both qualitative and quantitative methods simultaneously (see: Tashakkori and Teddlie, 1998; Greene, Benjamin and Goodyear, 2001; Creswell and Plano Clark, 2007) pragmatism stands out as a functional solution by allowing the focus to reside not on the truth of interpretation or the nature of reality, but rather on the specific problem being researched (Feilzer, 2009, p. 2). In the context of this study, this would mean allowing the focus to reside on how the results can be of practical use for the teachers.

2.2 A Question of Ecological Validity

Reflecting on the choice of the research problem for this dissertation, the value of a tool like the one proposed is dependent on the question of ecological validity. Ecological validity is used to describe and discuss the generalizability of the results of an experiment in real-world scenarios (Schmuckler 2001). Schmuckler (2001) provides a description of the different designs evaluations of ecological validity have taken through the years and arrives at the conclusion that, at least within psychology, the exact meaning of the term continues to be debated.

Kasprowicz & Marsden (2017) use the term ecological validity to discuss the relationship between experiments on the benefits of input-based practice for L2 German learners, and their operationalization becomes centered on the replication of real-world output scenarios in regards to comprehension tests. Kinginger (2013) uses ecological validity to discuss the issue of representative samples regarding identity in studies on the efficiency of Study Abroad-type language projects. Rogers & Cheung (2020) produced a meta-study based on finding learning gains across several experiments conducted in a real-world setting, and use ecological validity to discuss previous studies conducted in a laboratory setting. In these projects, ecological validity can be taken to mean different things, i.e., either the validity of a study in the actual population it is researching or whether or not an experiment is actually representing the real-world scenario it is trying to apply results to.

Returning to the conceptualization of the term, Schmuckler describes the first dimension of ecological validity as “a concern with the setting or environment in which the research takes place” in regard to representability outside of the specific environment in which the experiments took place (2001, p. 421). The second dimension Schmuckler ties to stimuli in psychology, which for the experiments described in his summary can be equated to the application of the tool in the current study within the experiment contra in a real situation. Finally,
the third dimension of ecological validity is presented as the "nature of the task, behavior or response", meaning whether or not the action performed during the experiment is actually representative of the action as undertaken in a real situation and accounts for all variables therein (Schmuckler, 2001, p. 423).

The ecological validity of the tool is the main research result addressed in this study, i.e., will such a tool function in a real-life scenario, and as such a definition of ecological validity is needed. Reaching a large enough, and diverse enough, sample of field observations for a generalizing statement regarding the efficiency of the tool in the entirety of the Swedish school system is beyond the resources of the current project, although some generalizing insights might be possible through emerging patterns (Larsson, 2009). The focus will thus follow the concept of ecological validity as applied by Kasprowicz & Marsden (2017) and, to some extent, Rogers & Cheung (2020) by attempting to design the dissertation project in a way that properly emulates the classroom scenario, before venturing into actual classrooms.

However, due to the design of the current study, ecological validity has to be accounted for in regard to each of the research questions individually. The next section outlines the questions in order of address and presents the specific definition of ecological validity in each case. These are discussed in an overlapping fashion, categorized thematically under the headlines “The Diagnostic needs of In-Service EAL Teachers”, which deals with the ecological validity involved in research question 1, “Available Tools and Methods”, which discusses ecological validity in research questions 2 and 3, and finally “Teachers’ Evaluation of the Tool”, which is concerned with research question 4.

2.2.1 The Diagnostic Needs of In-Service EAL Teachers

The literature that comprises the background section on corrective feedback in SLA and EAL stems from research conducted at different levels of education, both university and upper-secondary, and in different contexts. Some of the studies referenced depict second language learners of English in native speaking countries, while others describe English language education in different countries where English has different functions and is of differing status in society. Relating to a common, albeit often criticized, visualization of the English-speaking world, most of the studies brought up in the previous research section have been conducted in Kachru’s inner and outer circles, while the current project is, arguably, situated in an expanding circle context (1985). However, as discussed by Peterson (2019), how English functions and how younger outer circle learners acquire proficiency in the language today is quite drastically different from the situation just one or two generations back.

This is where the first dimension of ecological validity becomes the most apparent, as the context of the current project is different from the research environment that informs the criteria for successful corrective feedback in EAL.
In order to gain an understanding of how applicable the previous research cited in this project is in the Swedish context, an interview study was conducted and the focal points and experiences of in-service teachers in the Swedish context were compared to the previous research to create an understanding of what is relevant and what new needs the Swedish context might bring.

2.2.2 Available Tools and Methods

Once an understanding of what is relevant functionality for the teachers and what is theoretically sound information to include has been built from the interview study and literature review, the design of the actual tool begins. This also moves the dissertation into the second dimension of ecological validity, as designing the tool in practice means designing stimuli for a desired effect in the classroom, i.e., something that supports the teacher in gaining an overview of the students’ written production and adapting their lesson plans accordingly.

Returning to pragmatism as the guiding notion of the dissertation, the design of the tool builds on the warranted beliefs arrived at up to this point but it must also consider the intended use within the school system. The tool must become a useful option within the real ecology of the Swedish upper-secondary EAL classroom. However, the primary testing and development were carried out using a learner corpus built for the purposes of this project; a controlled environment that, much like Schmuckler (2001) warned against, lacks many of the variables at play in the real scenario. To maintain ecological validity, the results from the testing in the controlled environment were later reviewed in light of the field tests. The final field studies were carried out in classrooms, but as development was in its final stages at that point in time it must be mentioned as a weakness of the design of the current project.

In order to prepare for implementation in a classroom the variables missing from the controlled testing environment, such as teaching preferences, preferred modes of student feedback, and the teachers’ working situation, had to be anticipated from previous research, curricula, and syllabi. The interview study was used as a complement to these documents and served as a part of the background materials that more directly tied into the teachers’ lived experiences.

The third dimension of ecological validity brought up by Schmuckler (2001) is the “...nature of the task, behavior or response”, which in the case of this project means performing the task in the classroom and school environment. Using the tool in this environment creates new needs that could not be tested with the learner corpus used for the development of the tool itself. In many ways, it was not until this late step that the use of the tool could be said to start resembling in earnest the task as performed in the intended setting.
2.2.3 Teachers’ Evaluations of the Tool

Finally, the project arrives at an explicit evaluation of the tool and method proposed. Through exploring the use of the tool in the field, issues with functionality and implementation that were not apparent in the controlled environment could be discovered and corrected. Furthermore, in-service teachers involved in the field tests could more directly influence the tool and fit it to their needs.

This is also where anything conclusive can be said regarding the ecological validity of the project as a whole. While previous sections have attempted to recreate certain parts of the real-world application, the field tests fulfill the criteria suggested by Schmuckler’s (2001) overview in a more direct manner. This dissertation will be unable to provide a generalizable conclusion that could be considered valid for the entirety of the Swedish school system. However, the field tests should provide enough insight to indicate whether or not digital diagnostic tools for EAL education aimed towards use by the teachers in the manner proposed is an interesting route for further development.

2.3 Survey Analysis Method

This survey uses a mixed-method approach with the goal of gaining insight into problems learners commonly have when attaining grammar, vocabulary, and spelling as experienced by in-service teachers in the upper-secondary schools of Sweden. It is also used to see how a digital diagnostic tool could fit within the teachers’ current routines. This was done by performing a survey consisting of semi-structured interviews in combination with a respondent drawing. The survey was followed by a literature review intended to put the results in context with the current literature on the acquisition of English as an additional language (EAL) in the Swedish school system and corrective feedback on student texts.

The decision to begin the project with the survey rather than the literature review is guided by Glaser and Strauss’ Grounded Theory, introduced in The Discovery of Grounded Theory (1967) and elaborated on by Glaser in Theoretical Sensitivity (1978). Grounded Theory builds on the notion that a theory should be formulated from the data, rather than having the theory guide the data selection (Starrin, 1997, p. 30; Denscombe, 2016, p. 157). In practice, this is quite close to the ideas behind inductive data-driven methods applied in linguistics and elsewhere. The formulation of a theory should be an ongoing process during the entire study while remaining clearly connected to the emerging data, meaning that a project based on Grounded Theory should have fieldwork and analysis running in parallel throughout (Denscombe, 2016, p. 158).

After the initial theory description, however, Glaser and Strauss came to disagree about the application of Grounded Theory. Glaser supported a creative approach to the interpretation of data with an open perspective on possible
outcomes, while Strauss took a more prescriptive stance with regard to the categorization of data (Cho and Lee, 2014, p. 3). The current project will follow Glaser’s line of thought as it fits better with the need for a summarized list of items from the survey for tool development. As the development will deal with defined categories of syntactical, morphological, and lexical issues, detailed categorization based on respondent wording is superfluous.

Denscombe (2016, p. 157) recommends Grounded Theory for studies using explorative data in specific environments, which fits well with this study as it takes place in the relatively well-defined context of the Swedish school system. The reason for the literature study being scheduled after the survey data collection is in its final stages is an attempt to avoid any researcher bias due to preconceived notions about which findings are relevant in the survey (Glaser and Strauss, 1967, p. 33; Denscombe, 2016, p. 161; Starrin, 1997, p. 45). Although Grounded Theory guides the entire study it will be most noticeable in the interview analysis, which relies entirely on the approach of creating categories and groups, “conceptual codes”, to see patterns and understand the data on a holistic level (Starrin, 1997, p. 35).

According to Cho & Lee (2014), Grounded Theory is seen by some critics as a framework and the method in practice as content analysis. Content analysis as a method of data analysis can be applied from both quantitative and qualitative perspectives, answering questions regarding frequency by summarizing material quantitatively or providing qualitative interpretations through the use of themes respectively (Bengtsson, 2016, p. 10). The approach of applying Grounded Theory through content analysis is used in this dissertation as a way of retaining a connection to the theoretical foundation throughout the analysis.

The content analysis process described by Bengtsson (2016) is divided into four steps: Planning, Data Collection, Data Analysis, and Creating a Report and Presentation of the Results (p. 9). Planning includes formulating the research aim, the specifics of the sample and units of analysis, as well as the method of analysis and intended practical implication of the research. This makes Bengtsson’s method description a good fit for the pragmatic approach chosen for the current dissertation, as including the intended practical implication aligns with the practice-oriented focus. Data collection will be discussed in the later sections, but it is interesting to note that Bengtsson has included initial data transformation in this step through the transfer of collected data into written text, i.e., transcription, while also emphasizing the interpretation that goes into such work (2016, pp. 10-11). Transcription in the current project is further discussed in section 2.4.5.

---

1 An issue to take note of here is that I am entering into the survey with a priori knowledge of the field and topic, as I have previously engaged with language acquisition through both work and study. I would also not have been able to apply as a candidate without some level of previous understanding of the topic. The design choice here should rather be seen as intended to minimize biasing on the specific topic of the dissertation.
As for the data analysis, Bengtsson makes a distinction between two types of content analysis. *Manifest Content Analysis* stays close to the material and provides information based on the exact wording of the text or respondent, while in *Latent Content Analysis* “the researcher seeks to find the underlying meaning of the text: what the text is talking about” (Bengtsson, 2016, p. 10). The results of the interview study are presented in a partially quantitative manner, but the categories are defined using latent content analysis. However, care will be taken to include the nuances of respondents’ answers as much as possible, with a more compiled set of categories being offered after the initial categories made explicit in the interviews have been presented.

### 2.4 Interview Design

The interviews were carried out in accordance with the methodology described by Kvale in *Den Kvalitativa Forskningsintervjun*² (2014) and *Doing Interviews* (2013) for semi-structured research interviews, with the addition of a respondent drawing as an elaboration and validation of the interview interpretation. The first six steps suggested by Kvale will follow in the context of planning the interviews for this project.

#### 2.4.1 Thematizing the Interviews

The first step is a clear thematization of the interviews, which should relate back to both the short-term and main goals of the project. Thematizing the interviews here means explaining the “What?”, “Why?” and “How?” by formulating research questions specifically for the interviews and providing theoretical clarification for the theme being investigated (Kvale, 2013, p. 37). The first two questions must be properly answered before the design of the interview study can make any progress, as the “How?”, or the method, must be specifically adapted in order for any reliable knowledge to be produced (Kvale and Brinkman, 2014, p. 148). Denscombe also approaches this question using the term “research strategy”, which is described as an overarching idea shaping a custom design for answering a well-defined question that is answerable with the resources available for the specific study (2016, p. 22).

“What” the interviews are intended to give insight into is, as previously stated, the feedback and corrective processes employed by in-service English teachers in the Swedish upper-secondary school system in relation to student texts, as well as producing some understanding of why the teachers believe the frequent topics of feedback are occurring. The interview method is well suited for this as the qualitative data gained from it can be used to “illuminate the meaning” of the teachers’ experiences, both through patterns and differences (Patton, 2015, p. 13).

---

² *The Qualitative Research Interview* (My translation).
As to “why”, the interviews are intended to show whether or not the in-service teachers’ beliefs regarding common problems are the same as those found in the research on the topic, in an effort to understand how a digital diagnostic tool could fit within their routine and to gain an understanding of how the teachers attempt to solve them in practice. Beliefs as a concept were used here as defined by Fives & Gill (2014), meaning that the interviews explore teachers’ stated beliefs regarding their practices. Borg (2016) argues for beliefs as an aspect of the situation being studied but emphasizes the importance of taking the other aspects of the topic being researched into account, for instance, the context, external influences, and observable practices which the expressed beliefs function within. When researching teachers and teaching this would mean a perspective on the interview results that also take the educational context and its governing policies into account, per the perspective promoted by Educational Linguistics.

The interview approach is a good starting point for this as well, as the method is more likely to result in elaborate answers and allows for a more rewarding exchange between the interviewer and interviewee than less dialogic options, such as questionnaires (Denscombe, 2016, p. 30). The long-term goal is to produce a basis for further work that is as current as possible³, and that engages directly with the teaching staff by addressing issues as described by the teachers themselves.

### 2.4.2 Study Design

The design of the study revolves around the thematization and following Kvale’s method of having the research questions be used to formulate the interview instrument. The instrument is then to be further developed into the actual interview questions. Furthermore, the initial set of questions is evaluated and refined through two sets of pilot interviews. The transcription and analysis process of the pilot interviews are also likely to indicate improvements as they offer further insight into the interview technique and content (Kvale and Brinkmann, 2014, p. 221). This necessitates an evaluation at the end of the interview, which is further recommended by Kvale as it allows for the interviewee to elaborate or pose any further questions about the interview or project (Kvale, 2013, p. 56). The interviews followed this general structure:

1. Introducing the project and the interviewer
2. Introducing the drawing template
3. The interview questions
4. Repeating the main points of the interview to allow for elaboration
5. Discussing the drawing

---

³ Here it is important to note that the majority of the interviews ended up taking place during the COVID pandemic 2020 - 2021, and that the situational context at the time of material collection was quite extraordinary.
6. Evaluating how the questions were understood
7. Asking for any further questions or elaborations

The introduction includes a briefing consisting of a general description, an introduction of any recorders used, and an introduction of the interviewer (Kvale, 2013, p. 55). The interviewee is also given the opportunity to ask any questions they might have before the interview gets started. Kvale describes this as “setting the stage”, and puts emphasis on how important this first impression is for the interview to maintain good quality. It is connected to the quality checklist, as a good introduction enables the interview to remain focused and results in less material with more content (Kvale, 2013, p. 80).

The interview questions should be kept narrow and be “devoid of academic language” (Kvale, 2013, p. 57). The goal of the interview should be to produce a transcription with short and precise input from the interviewer and elaborate, content-rich answers from the interviewee, with follow-up questions posed throughout to make sure that the content is properly understood as further elaboration and validation might be impossible at the later stages of the study (Kvale and Brinkmann, 2014, p. 206). This is also why the main points of the interview and a summary of the answers are presented to the participant after the interview questions are completed as a debriefing (Kvale, 2013, p. 60). The opportunity to ask for elaborations and to immediately be able to validate the understanding of an answer is one of the major benefits of conducting face-to-face research (Denscombe, 2016, p. 32).

In order to achieve this narrow and to-the-point interview content, the interview consisted of three scripts moving from the initial research question to the interview-specific research questions before arriving at a final product with short questions posed in a clear manner and separated into different categories: The Teacher, The Student, Routines and Feedback, and Digital Tools and Experiences. Evaluating how the questions were understood ties into the quality refinement described in the previous paragraph.

2.4.3 Sampling and Respondent Recruitment

Before moving on to the interview guide, the question of the selection of participants must be addressed. The primary sampling frame was in-service teachers at upper-secondary schools who were currently engaged in teaching English at an upper-secondary level, which includes teachers active at adult education centers or any other context where the standard upper-secondary courses are taught. As the recruitment of participants relied on my network from previous workplaces and training, many of the participants were likely to be early-career teachers. Initially, local schools were contacted and asked to send an invitation to their EAL teachers. As this did not yield enough respondents and was unlikely to do so as the pandemic became more apparent, recruitment from my personal network took place as a type of convenience sampling.
The primary selection for the main sets was quite wide and the participants were chosen based on being actively engaged in teaching English in Sweden at an upper-secondary level without any other criterion. However, participants not fitting these criteria were also included in the first iteration of the survey instrument. This was done partially due to an interest in how the results could differ for participants from outside of Sweden, but also as a way of managing the limited number of respondents fitting the criterion in a way that would benefit the later iterations that are used for the actual analysis.

According to Grounded Theory, the selection must also be allowed to change as the data further develops the research subject, as developments might lead to new needs for the project (Denscombe, 2016, p. 163). The goal is to achieve a “saturation of knowledge”, meaning that the answers become repetitive and no “new” information is gained (Berntaux, 1981, p. 37). Limiting the number of interviews is also important due to the time constraints for the study, which is pointed out by both Kvale and Brinkman, and Denscombe as a valid consideration (2014 p. 156; 2016 p. 24). The number of interviews for this survey was dependent on the number of respondents available within the time allotted for the survey in the project’s schedule (1 year).

The number of interviews also needs to be considered for the role they play in this study as a whole. Both Kvale and Brinkmann (2014, p. 156) and Patton (2015, p. 16) warn against seeing a large amount of interview data as more valuable than a small amount when the focus should instead be put on the quality of the data produced, or the qualitative data rather than the quantitative. It is important to realize, at this early stage, that the low number of interviews, 12 in the survey, conducted within this project is only viable for a small subset of in-service teachers and should not be generalized beyond its intended limits. The initial interview study in the current dissertation functions as a compliment to the previous research meant to explore the applicability of that research in the Swedish setting, and should not be considered strong enough to stand on its own as a generalizable overview of teacher beliefs.

The likelihood of maintaining a higher number of respondents throughout the entirety of the study is also quite low. The recruitment of participants for studies on second language teachers is notoriously difficult and often involves many unpredictable issues appearing during the study (Hobbs and Kubanyiova, 2008). In addition to this, the participation of teachers in a study is difficult to maintain over time both due to the loss of engagement (Hobbs and Kubanyiova, 2008), participants finding the research more obtrusive than anticipated, or simply due to unforeseen changes in the lives of the participants, such as moving (Källkvist and Juvonen, 2021). As noted by Källkvist and Juvonen, redundancy should be considered when recruiting participants, if possible (2021, pp. 51-52).
2.4.4 The Interview Guide

As previously mentioned, the overarching research question for the interviews is: “What are the needs of Swedish EAL teachers in terms of focus-on-form written language evaluation and diagnostics?”. Kvale discusses questions as having a thematic property and a dynamic property. The thematic property is what deals directly with the research question for the interview, while the dynamic property concerns the question’s function in the discourse of the interview (Kvale and Brinkmann, 2014, p. 173). A good dynamic question serves to keep the discussion flowing, while also inviting further elaboration or follow-up questions. Patton also discusses the use of different questions, open and closed, as a way of introducing mixed methods in the interview by having the open-ended (dynamic) questions aimed at producing qualitative, elaborate data and closed questions (thematic) that generate a quantitative frequency on beliefs regarding certain topics among the teachers (2015, p. 15). The specific questions for the interviews are:

1. What error(s) in syntax and lexis are occurring in the students’ written production?
2. Are the same issues found in all modes of production?
3. Do they relate to factors like influences from outside of school or earlier language education?
4. What is the routine for supporting students with recurring impediments?
5. How do teachers work with corrections, feedback, and adaptation?
6. What are the background factors for the teachers’ perspectives?

What is described above are the thematic questions for the interviews, as they do not take the flow of the conversation into account or allow for elaboration. The questions were strictly concerned with diagnosing the errors experienced by the teachers in the students’ production, the teachers’ responses to these errors, and gathering data for categorization of the teacher’s experiences. Many of the questions above are also likely to get a simple yes or no reply, which would drastically diminish the quality of the data according to the quality markers described by Kvale (2013, p. 57). In order to create a functional interview guide, a dynamic dimension must be added, and interviewee interaction must be supported. However, by connecting the interview questions to the questions above, a clear connection to the aim of the study can be maintained through the different formulations and dimensions.

Working based on the thematic questions, the thematic properties were lifted into content categories, which were then elaborated on through questions that took the dynamic properties into consideration. This resulted in an instrument with a clear connection to the questions posed above, but which would also allow for a dynamic interaction in the interview situation. The interview guide
containing the questions used in the interviews and the interview structure is found in Appendix A.

### 2.4.5 Transcription and Analysis

The transcription process for these interviews will focus mainly on the content, as the goal is to create an overview of beliefs and experiences rather than being used for language analysis. This is in line with Kvale & Brinkman’s assertion that transcriptions should contain what is necessary for the analysis while being as coherent and clear as possible (2014, p. 221). The transcriptions must be shaped by the specific needs of the project and the intended analysis, rather than being perfect representations of the dialog between the interviewer and interviewee (Kvale and Brinkman, 2014, pp. 227, 230).

However, the transcriptions will still need to allow for some understanding of the spoken language they represent, as some information might otherwise be lost due to the transcriptions lacking the context for later analysis (Kvale and Brinkmann, 2014, pp. 223-224). This relates to the validity of the transcription as omitting repetitions or notes on the tone of voice will make it impossible to discuss uncertainty or the meaning of denials, while it is in the interest of the transcriber to present a transcription with a “literary style”, here meaning easily comprehensible, in order to convey meaning and nuances of statements to the reader (Kvale, 2013, p. 98).

With the above paragraph in mind, adapting the transcriptions to the needs of this study means that the transcript should have a focus on content and meaning presented in a literary style that conveys the statements, experiences, and beliefs of the interviewees clearly. The transcriptions should also include repetitions and pauses for possible further analysis (Kvale, 2013, p. 132). Any quotes used for emphasis on content and meaning in the main body of the text will be edited in a literary style for clarity.

The primary means for analysis of the interviews will be a qualitative reading and summary in the style of latent content analysis. Some of the interview questions are used to understand the contexts of the different teachers’ experiences and backgrounds. These questions in combination with information about the teachers’ work experience will provide the variables for sorting the data and context necessary for the qualitative analysis (Patton, 2015, p. 13).

Within Grounded Theory, the practice of using categories is presented as the use of “conceptual codes” which connect data to the resulting theory (Starrin, 1997, p. 35). The analysis initially makes use of two different types of code, substantive code, and theoretical code. The former creates categories of data based on variables relevant to the research topic, in this case, keywords and descriptions of language errors, while the theoretical code creates links between the substantive categories that allow an explanation to emerge from the pattern (Starrin, 1997, p. 36). Presenting the data in this way will provide a list of recurring errors, elaboration on those errors, and a selection of sorting variables...
that could show any patterns of interest in the data. The findings can then be forged into a narrative that introduces themes, patterns, and insights in a form that provides a general understanding of the material (Patton, 2015, p. 14). Once the contours of a pattern connected to the research topic start to become visible, the analysis can move into selective coding, meaning that the main categories of relevance are stated and that data is mainly sorted into these larger categories (Starrin, 1997, pp. 37 - 38).

When conducting the search for substantive codes there are two good reasons to worry about the validity of the data produced. Firstly, the interviews will, in most cases, be conducted in Swedish, while the data is presented in English. This creates a barrier between the raw interview and the representation of the data where the interviewer’s/transcriber’s translation plays a large role in the outcome. The focus of this dissertation is not the details of the language used by the respondents, but rather the content that they are communicating. Therefore, the translation is conducted according to sociolinguistic translation theory, where the main aim is to convey the message of the material (Ingo, 2007, p. 13). The focus is to create an equivalent understanding and reaction in the reader of the translation as would be experienced by a reader of the original, untranslated text (Nida and Taber, 2006, p. 12). Ingo summarizes the goal of the translation process as expressing the original content in a way that is functional for the situational factors in terms of pragmatics, stylistics, and semantics while still being structurally representative (2007, p. 15). These different factors are related to the purposes of the text, which were previously described by Kvale & Brinkmann’s description of a good transcription (2014, p. 221). The Swedish language transcription is provided alongside the translation.

The second issue has to do with synonyms and different ways of describing a similar situation, language error, or student impediment. This will be an issue when providing frequencies for the keywords when creating the codes, where one either uses a multitude of categories that make the data hard to comprehend and creates a situation where some information might be lost due to the same information being scattered into several categories, or one allows for interpretation about which keywords can go into the same category, which puts further distance between the representation of data and the actual interviews. To produce coherent and clear data this study will allow for interpretation and sort the data into a smaller number of categories, following Patton’s suggestion that a small number of strong insights is preferred to a large number of inconclusive ones, but it is yet another thing to be aware of when taking part of the results (2015, p. 16).

In order to support the coding process, Bengtsson’s (2016) process description of latent content analysis is used to transform the interview materials into sets of substantive codes. The process takes place in four stages, with each stage having a defined outcome that feeds into the next stage. The stages are
called decontextualization, recontextualization, categorization, and compilation (Bengtsson, 2016, p. 9).

The first stage, decontextualization, involves familiarizing oneself with the collected data to gain a holistic understanding of the contents. This is followed by breaking the content down into “meaning units” (Bengtsson, 2016, p. 11). Meaning units are the smallest unit containing insight relevant for the purposes of the analysis. Once identified, meaning units are labeled with a code relating to their content. Coding can be done either inductively or deductively. As the current study builds on Grounded Theory, inductive coding is used to capture emergent patterns in the materials.

Once meaning units have been identified and coded, the recontextualization stage begins. In practice, this stage consists of re-reading the original materials and ensuring that all of the relevant content is properly reflected in the previously created codes and meaning units. Bengtsson highlights the importance of letting go of non-coded text at this point in order to ensure the relevance of the results:

> When the researcher is deeply involved with the data, everything seems to be of importance. Nevertheless, a process of distancing is necessary, and the researcher must allow him or herself to let go of the unimportant information that does not correspond to the aim of the study. (Bengtsson, 2016, p. 12)

Categorization refers to the stage at which the codes derived from the meaning units are combined into thematic categories. However, before categorization begins, the meaning units must be condensed. Condensing the meaning units involves reducing the number of words without removing content relevant to the study (Bengtsson, 2016, p. 12). This is sometimes at odds with the style of transcription chosen for the dissertation, as clarity for the reader is emphasized. Categorization can take place on both a thematic and sub-heading level, and the act of moving meaning units between categories is considered progressive development of the category system. In connection to Grounded Theory in the current study, the sub-headings could be seen as sub-headings to the substantive codes, while the thematic categories end up becoming the substantive codes themselves. Once a reasonable outcome in the context of the aims has been achieved, the categorization step can be considered complete.

In the final stage, the thematic categories are compiled into a conclusion. This is similar to how the substantive codes are compiled into theoretical codes in Grounded Theory. The compilation process should be tied into the categories, and make the conclusions drawn explicit in regards to the connection to the meaning units. The focus should be on finding the essence of the materials, or as Bengtsson states: “In phenomenological and hermeneutical-based studies, the researcher focuses on exploring how the informants make sense of
experience and transform experiences into consciousness. The researcher must then attempt to find the essence of the studied phenomenon” (Bengtsson, 2016, p. 12). As both of the sections making use of content analysis in this dissertation can be considered phenomenological in nature, i.e., exploring teachers’ beliefs and experiences, the quote is fitting for the aims.

2.4.6 Verifying the Data

Before approaching the topic of verifying the data, the objectivity of data gathered from this interview study needs to be addressed. The first goal is for the data to properly represent the content of the interviews themselves, what Kvale calls “being adequate to the object” (Kvale, 2013, p. 121). This could be seen from a perspective based on the research aims for the study, meaning considering how well suited the data, if analyzed and presented as described above, is to adequately represent what is being researched, in this case language issues as experienced and described by teachers. Presenting the data in a quantitative manner, through connecting each topic to each respondent who has mentioned it in a content network, will represent which errors are commonly found and to which degree. The closer reading of data related to the errors found represents the actual issues. However, both of these representations suffer from the same issues that are discussed in the previous section, i.e., the translation barrier, and the possible difference between interpreted meaning and the respondents' intended meaning.

As a possible solution to the latter issue, Kvale also refers to Latour’s suggestion that a way of obtaining objectivity is by “allowing the object to object”, meaning that the object of study, in this case the teachers, should be allowed ample opportunity to protest against any conclusions or interpretations drawn from the data (Kvale, 2013, p. 121). In the context of this study, this means allowing the teachers access to the material and giving them the opportunity to challenge it once the study is completed. Doing so does not only increase the objectivity of the study by, possibly, decreasing any bias in the interpretation of the interviews, but it will also allow for further elaboration by the teachers on any questions that remain unclear. This also connects to the validation of the data as allowing for feedback and elaboration from the teachers makes sure that the interpretation has not strayed too far from what they intended to convey. Furthermore, since the results from this study will be presented with the participants being kept anonymous, this process will allow the participants to keep their “voice” in the findings (Kvale, 2013, p. 28). Respondent validation of analysis and results is also recommended by Bengtsson, albeit with caveats regarding the influence of time passing between interviews and validation amongst other things (2016, p. 12).

A common argument against the interview study as a method is the reliability of the data when it is used to generalize findings (Kvale, 2013, p. 126). The criticism has to do with the small number of participants, and the
often very different experiences of people in different regions, or in this case schools. This dissertation intends to use the interviews as a supplement to the literature review, which suffers due to a lack of relevant research in the Swedish school system. This is not a criticism of the research carried out on, and within, the Swedish system, but for the purposes of recontextualizing the previous international studies within this study the interviews must be specifically designed around the earlier research.

2.5 The Respondent Drawing

The use of drawings as a data collection instrument is not new within the field of linguistics and is also extensively applied by psychologists and psychiatrists. The use of templates in the form of a body silhouette has been used to produce language portraits that explore the use of different languages in young children, and to enable discussions about their backgrounds and relationship to the dominant language of the society they live in (Neumann, 1991; Krumm and Jenkins, 2001). The use of template drawings as a complement to verbal data collection has been used in a variety of studies and has been found to be a good addition to the classic format of interview studies (Busch, 2015). While the topics discussed in those studies are different from the ones engaged with here, they show the benefit of multimodal approaches to data collection (Busch, 2012, p. 8).

Beyond Educational Linguistics, Kearney and Hyle discuss the use of participant-produced, or respondent, drawings in respondent studies within educational settings when surveying responses to changes in organization and claim that the procedure helps with bringing sub-conscious ideas, thoughts, and emotions to the responses (2003, p. 236). This is also found to be true in several other organizational studies conducted in different environments for a variety of reasons, including research-oriented surveys (Zubroff, 1988; Meyer, 1991). Drawings as a way to “tap quickly into the emotional lives of the participants” were useful in the present study as they offered insights into the attitudes and underlying experiences of the teacher participants regarding different digital tools, but also their experiences of the feedback and support routines in place in their workplaces (Kearney and Hyle, 2003, p. 363). Likewise, Ward & Shortt explored the use of respondent drawings in interview situations and found that the use of drawings enabled respondents to focus the interview in accordance with their own agenda and granted the respondents more agency in their interactions with the interviewer (2012). As one of the main aims of this project is answering to the needs and experiences of the participants, such a tool is very useful in maintaining a beneficial focus on the teachers’ routines and experiences when collecting the data.

However, in a later text on visual data collection, Ward & Shortt (2018) also discuss the drawbacks and criticisms of respondent drawings. Beginning with
criticism of the method from the previously mentioned 2003 article by Kearney & Hyle, the authors suggest that the method could cause respondents to feel insecure regarding their ability to draw, which could act as a barrier to engagement. Previous uses of respondent drawings in research on more delicate topics have also been shown to make respondents feel more exposed than verbal instruments. This makes considerations about the nature of the topic being surveyed important (Guillemin, 2004; Morgan, 2009). The topic covered by the present study is not inherently delicate in nature, but Ward & Shortt’s warnings regarding awareness of where participants might feel uncomfortable or challenged must still be taken into consideration (2018, p. 272). As a partial aim of this survey is to gain insight into teachers’ experiences with digital tools, feelings of inadequacy or similar regarding computers could be a sensitive area for the respondents due to the quick pace of change in the area, especially within the school system.

The respondent drawings will play a supporting role in this survey and are meant to complement the interviews with additional information. This is similar to previous uses of the instrument (for instance Guillemin, 2004; Kearney and Hyle, 2003; Morgan et al., 2009). The instrument used for this study will be a structured drawing, meaning that the respondent will be provided with a template that they will populate with their own additions and notes (Meyer, 1991). Providing a template can also help lower the previously mentioned barrier of engagement due to circumventing some of the anxieties regarding graphical ability. The structured respondent drawing is similar to a semi-structured or structured interview, as the use of a template will also aid in keeping the data focused.

The template used for the present instrument contains three figures: two human silhouettes similar to the figure used by the Spracherleben research group at the University of Vienna and one stock-art stencil for a laptop computer without any symbols specifying manufacturer or operating system (Busch, 2012, p. 9) (Fig. 3). The three figures correlate to three of the categories found in the interview guide, with the fourth category (Routines and Feedback) manifesting in the relationship between the three figures on the paper.
The template was given to the respondent before the interview started, and the purpose of it as a mapping of routines was explained. The human silhouettes were not identified as teacher and student as the order might influence the teachers’ depiction of the workflow. The respondent was asked to make notes beside the relevant picture during the interview and was afterward asked to describe and draw the workflow between teacher, student, and computer between the template figures.

The drawings enabled the data to provide an idea of how the teachers imagined the flow of information through the new tool. The notes taken during the interview also showed what information the teacher experienced as important and indicated how the teacher related the items discussed to each other. The respondent was also given the opportunity to decline to use the drawing and just participate in the interview or to not submit the drawing after the interview session.

2.6 Ethical Protocol

The Swedish Research Council’s\textsuperscript{4} Codex website offers an entry regarding laws and principles for research involving humans in Sweden, which points out that sensitive topics require the approval of the regional ethics committee before being conducted. Denscombe provides a list of these topics which includes sexuality, religion, ethnicity, and illegal activities. Any research conducted on vulnerable groups such as ethnic minorities, children, or people with disabilities must also be formally approved (Denscombe, 2016, p. 425). Due to the fluid

\textsuperscript{4} Vetenskapsrådet
nature of interviews in terms of content, the survey portion of this study was submitted for ethics approval from the ethics committee in Stockholm, Sweden. The committee decided that the survey did not need formal ethical approval due to it not concerning sensitive materials, but provided feedback regarding which sources should be consulted for the ethical aspects of the survey design. In addition to the ethical considerations, the GDPR regulations regarding storing personal information were also considered, and the materials were stored in a locked safe in the university offices.

The current study follows the Research Council’s protocol for the ethical handling of personal information, consent, and confidentiality in order to be ethically sound. Furthermore, Patton points out the importance of confidentiality and information handling regardless of the interview topic as “...people in interviews will tell you things they never intended to tell. Interviews can become confessions, particularly under the promise of confidentiality” (2015, p. 495). Kvale recommends formulating an “Ethics Protocol” as a part of the design (2013, p. 99). The protocol relates to six of Kvale’s seven steps separately (Tab. 4), as the final step is the report. The protocol for this study is presented below, with additional input from other sources.

<table>
<thead>
<tr>
<th>Thematization</th>
<th>Design</th>
<th>Interviews</th>
<th>Transcription</th>
<th>Analysis</th>
<th>Verification</th>
<th>Report</th>
</tr>
</thead>
</table>

Table 1: Kvale’s seven steps of interview studies

Regarding the first step, thematization, the ethical questions to answer are whether or not the research is likely to produce anything beneficial for the participants or not, and if the anticipated results will contribute to any new knowledge (Kvale, 2013, p. 24). These questions are necessary due to the inherent risks of filing information on participants, and the further danger that such information might be used either to identify the participants or cause them any type of grief in the future (Denscombe, 2016, p. 424; see also Patton’s quote above). The risk involved for the participants must be weighed against the possible benefits of the research before any research is conducted (Kvale and Brinkmann, 2014, p. 110; Patton, 2015, p. 497). Denscombe discusses this in relation to the ethics principle “Beneficium”, which dictates that even though the results of the research might not benefit the participant directly, it must be

---

5 Etikprövningsmyndigheten Uppsala Dnr: 2019-02110
of use to a person in a similar situation (2016, p. 429). In the context of this study, the main goal is to benefit the participants by providing teachers with a new tool to help them cope with their growing workload and there is no explicit risk connected to the topic being discussed, meaning that the study is a low risk with a good chance of being beneficial to the participants if successful.

The second step, the interview design, is connected to ethical questions regarding informed consent, confidentiality, and personal consequences for the participants (Kvale and Brinkman, 2014, p. 99). The last question has already been elaborated on in the paragraph above. Informed consent is summarized by Kvale as “...informing the research subjects about the overall purpose of the investigation and the main features of the design, as well as of possible risks and benefits from participation in the research project” (2013, p. 27). It also includes informing the participants about the fact that participation is entirely voluntary and that they can withdraw at any point (Kvale, 2014, p. 107; Denscombe, 2016, p. 437). Participants should be informed about these things in both the briefing and debriefing during the interview, but it is also highly suggested that the written consent form should include this information as well (Kvale and Brinkman, 2014, p. 107). Apart from this, the written consent form should also specify how the interviews and the resulting data will be stored, used, and published (Kvale and Brinkman, 2014, p. 108; Denscombe, 2016, p. 438). As such, the information was included in the forms used for collecting the written consent.

The interview transcriptions resulting from this study were anonymized and coded to ensure that they cannot be tied to the identity of the participant. One physical copy of the transcript codes and corresponding names will be kept in a safe together with the recordings to ensure that only the researcher has access to the non-anonymized material (Denscombe, 2016, p. 441). This process is necessary due to the need for the material to be completely anonymous in order for the demand for confidentiality to be fulfilled, but it must also be possible for the researcher to identify the specific participants at the later stages of research if they were to withdraw from the study for any reason (Denscombe, 2016, p. 429). The recordings will also only be stored, or listened to, on hardware without any connection to the internet or any other form of network.

The next step of Kvale’s protocol is related to the interview situation and deals with the stress and anxiety the interviews might cause the participants (2014, p. 99). Denscombe also points to the risk of psychological trauma as one of the main considerations to keep in mind when designing studies involving interviews (2016, p. 428). Being honest and clearly introducing the interviewer, the project, and how the interview data will be handled after the interview is completed can alleviate some of the stress, which is why both a briefing and debriefing session is included in the interview script (Kvale and Brinkman, 2014, pp. 59-60). The power dynamic between the interviewer and the participant can also be a contributing factor to stress and anxiety experienced
during the interview (Kvale and Brinkman, 2014, p. 113). This asymmetry can lead to the data being unreliable or can negatively influence the quality of the interview content, apart from the obvious psychological harm it can cause the participant, and steps must be taken to minimize the impact of it if the study is to be considered ethically valid (Kvale and Brinkman, 2014, pp. 51 - 53).

In order to mitigate the power dynamic suggested by Kvale, the participants are given access to the interpretations, clear briefing and debriefing, and the opportunity to protest any interpretations or elaborate on answers before any material is published. None of the methods described above can in any way guarantee that a participant will not experience stress, anxiety, or a negative power asymmetry, which puts a lot of importance on the interviewer being receptive to the emotional situation of the participants throughout every part of the process (Kvale and Brinkman, 2014, p. 49). The remaining four steps are the transcriptions, analysis, validation, and reporting, which are tied to ethical considerations about confidentiality, objectivity, participant influences on the material, and the anonymity of the participants in the presented material (Kvale and Brinkman, 2014, pp. 99 - 100). These considerations have all been discussed above in connection to the earlier steps or as a part of the study design.
3. Background

This chapter provides a background of the main fields relevant to the current project. Previous research is presented, and possible problem areas are highlighted. The chapter starts with a brief introduction to the EAL subject in the Swedish context, followed by Educational Linguistics which provides the overarching field in which the dissertation is placed, and then moves on to second language acquisition (SLA), corrective feedback in second language acquisition and human-computer interaction.

Educational Linguistics is presented as a chronology of developments, which highlights it as a continuation of applied linguistics. The main ideas of the field and the reason for its branching off from applied linguistics are presented. Once defined, the field is put into context with the current study.

Second language acquisition is presented against a background of former TESOL president H. Douglas Brown’s overview *Principles of Language Learning and Teaching* (2000). The major models of the field are covered to some extent, with a focus on socio-cultural scaffolding and how it would function within the current project.

The background chapter will then narrow its focus further as it moves on to the topic of corrective feedback on student texts within SLA teaching. In order to remain critical of the application of corrective feedback, the section on the topic will trace a debate between Professors Dana R. Ferris and John Truscott as a way of accounting for the previous research both in support and against the use of corrective feedback. The main points of their exchange will then be relayed to the context of the current project.

Finally, a brief introduction to the field of human-computer interaction (HCI) is given. The main focus of the section is on how digital tools function within school-like environments, and on how tools can be designed in collaboration with the intended users. While the field of this dissertation is language acquisition, HCI is incredibly important as a tool that is not used does not benefit anyone.

3.1 The English Language Subject in the Swedish Context

The current Swedish teacher education for English language subject teachers follows the same guidelines as for other subject teachers, meaning that students pick two subjects and take obligatory courses in pedagogy, subject didactics, general didactics, scientific theory, assessment, and the school system together with students from other subjects. The teacher training programs also include 20 weeks of practice placements in schools. English can be selected either as a
primary subject of 120 European Credit Transfer and Accumulation System (ECTS) credits or as a secondary subject of 90 ECTS credits. In total, the teacher training program takes five years to complete and contains 300 ECTS credits of courses.

In 2013 teacher certifications came into effect, meaning that only certified teachers became eligible for permanent employment within the Swedish system, with exceptions in place for vocational subjects and mother tongue tuition (Skolverket, 2012a). There are ways of becoming certified as a teacher without going through the teacher training program, such as programs containing only the non-subject courses as a complement to previous subject studies. The license is granted by Skolverket and is not an immediate certification upon graduating from the teacher training program. As such, anyone fulfilling the requirements can become certified, regardless of whether those requirements have been met through a teacher training program or not.

The Swedish school system consists of ten years of mandatory schooling and 3 years of optional schooling. The ten mandatory years stipulated by the Swedish Education Act (SFS, 2010) are divided into the preschool year (age 6), primary school (age 7-9), middle school (age 10-12), and secondary school (age 13–15). The compulsory years can then be followed by an upper-secondary education (age 16-18), which is divided into programs based on students’ choice of national program and/or previous grades. These 3 upper-secondary years contain the courses that are the focus of the current project.

English for the first ten years is given in 1 course, Eng Grund (English Primary), but the core content is specified separately for years 1-3, 4-6, and 7-9. Knowledge requirements are provided only from year 6 and up. Instructions are divided into 3 categories: Content of communication, Reception, and Production & Interaction. The last 2 categories cover comprehension and production in written, spoken and interactional modes at a basic level. The content of communication guidelines mainly deals with the topics that material on the target language and student assignments should cover. In addition to the core content, the purpose of the subject is also specified, in which specific aims and abilities for the students to develop are mentioned.

Upper-secondary EAL is taught in 3 courses with Eng 5 being a mandatory core subject and the latter 2, Eng 6 and Eng 7, being offered as program-specific or optional courses. English courses in upper-secondary programs correspond to the B level of the Common European Framework of Reference for Languages (CEFR) framework and levels 5 to 6.5 of the International English Language Testing System (IELTS) framework (Skolverket, 2012b). As with Eng Grund, the upper-secondary course descriptions contain Content of communication, Reception, and Production & Interaction. The verbiage of the course descriptions is indicated to relate to affective abilities in order to make explicit the importance of maintaining students’ motivation and confidence, although
these aspects only appear in the SNAE’s comments to the course documents (Skolverket, 2012b).

The content builds on what has already been presented in Eng Grund, which is due to all courses being built from the CEFR guidelines regarding the inclusion of intercultural and socio-cultural awareness in the EAL instruction so as to positively influence possibilities for communication between nations (Europarat, 2020, pp. 100–102). However, the demands on formality, language adaptation, and comprehension skills increase with each course. The abilities intended to be practiced also become more demanding in the upper-secondary courses, such as the ability to revise their own and other students’ texts,

Since the year 2000 syllabus reform, GY2000, the content of the English courses in the Swedish system are tied to the CEFR. According to the SNAE steps have been taken to make CEFR more visible in the current curriculum, GY11, as can be seen in the content comments referred to in the previous paragraph (Skolverket, 2018). However, it is important to note that the courses within the Swedish system do not equal the levels of CEFR. Rather, the CEFR levels overlap across the courses as seen in Table 2 below.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Course</th>
<th>CEFR Level</th>
<th>IELTS Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>EngGrund (Primary)</td>
<td>A 1.2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>EngGrund (Middle)</td>
<td>A 2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A 2.2</td>
<td>4 (Basic user)</td>
</tr>
<tr>
<td>9</td>
<td>EngGrund (Secondary)</td>
<td>B 1.1</td>
<td>4.5-5 (Independent user)</td>
</tr>
<tr>
<td>10</td>
<td>Eng 5 (Upper-secondary)</td>
<td>B 1.2</td>
<td>5-5.5</td>
</tr>
<tr>
<td>11</td>
<td>Eng 6 (Upper-secondary)</td>
<td>B 2.1</td>
<td>5.5-6</td>
</tr>
<tr>
<td>12</td>
<td>Eng 7 (Upper-secondary)</td>
<td>B 2.2</td>
<td>6-6.5 (Proficient user)</td>
</tr>
</tbody>
</table>

Table 2: GY11 English course and corresponding CEFR and IELTS levels. Adapted from Skolverket 2012 and IELTS.org.

Themes that can be tied to the idea of focus-on-form assessment in the Swedish EAL context were identified by Tholin (2015), who summarized the Swedish dissertations published on the topic of foreign language education in Sweden between the years 2000 and 2009. Amongst the six themes identified by Tholin, “Assessment of students’ oral and written proficiencies” and “Error

---

6 Swedish National Agency for Education (Skolverket)
analysis and transfer” stand out as interesting for the topic of the current dissertation. It is important to note that Tholin’s review positions itself as exploring the decade leading up to the formulation of a new teacher education system, and is dealing with doctoral theses produced while language didactics was still a relatively young field in the country (2015, p. 77).

The theme “Assessment of students’ oral and written proficiencies” contains results indicating that upper-secondary EAL students are capable of functional self-assessment, which aligns with the inclusion of requirements regarding self-revision in the English subject description (Oscarson, 2009). Oscarson’s dissertation also indicated that in order for such self-revision to be beneficial, students would have to be given the opportunity to practice self-assessment. This aligns with the results on peer feedback in the lower years presented by Berggren (2019). While the idea of self-revision is directly tied to the core content through the ability to revise their own and others’ written production (Skolverket, 2012b), it could also be tied to the broader goal of building language confidence.

“Error analysis and transfer” does, as the name might suggest, focus on the issue of language transfer and its influence on errors in students’ written production. An included English subject dissertation by Köhlmyr (2003) explored errors in 400 national exam student texts, and found that concord and word order were the most common. Köhlmyr’s analysis also highlighted verbs as being the most frequent and serious errors, as their nature would likely be jarring for a native speaker reading the texts. The results highlight both errors due to transfer and due to overuse of certain forms. Tholin summarizes the dissertation as discussing the results from the perspective of correctness versus communicative competence, and arguing for grammar teaching and varied feedback as important aspects of teaching EAL in Sweden (2015, p. 87).

On the topic of transfer in English language education in Sweden, Hult describes a movement from teaching formal skills through transfer towards aiming to teach communicative skills through communicative teaching (Hult, 2012, p. 232). This can be seen in the move towards communicative teaching as a principle of the new curricula. However, he also notes that English within the system can play two separate roles: a core subject and a medium of instruction. The former of these roles is the main focus of the current study, as the dissertation deals with the language of the student production without engaging with the content of said production. However, developing language skills will also benefit students in acquiring knowledge where English is the medium of instruction.

After the year 2000, Hult notes a change in the way English as a subject was envisioned in Swedish policy and syllabi. A move was made from an EFL perspective aimed at transplanting the language from anglophone areas to allow students to communicate with native speakers towards English for the purposes of intercultural communication and for local use according to the language’s
status in the surrounding society. When writing about the then-current 2000 revision syllabus, Hult states that "In this sense, the current national syllabus for English attempts to encapsulate the transculturation of English—the notion that linguistic globalization is not simply about transporting a language as-is from one cultural space to another but about reconceptualizing it for new social spaces" (Hult, 2012, p. 223).

While the syllabus provides a picture of English as an especially strong second language in Sweden that shares the top segment of the language hierarchy with Swedish, Hult (2012) found that, in 2000, Swedish remained the natural language of communication within the classrooms and that English was only allowed to borrow this function when teachers actively facilitated a suspension of norms within the environment. The reconceptualization of English as a local language in Sweden could then be said to have moved further along in policy than in practice. This places the teaching of the language in a peculiar position in terms of focus-on-form feedback on written assignments, as the target becomes difficult to define.

### 3.2 Educational Linguistics

Hult (2008, p. 11) provides a chronology of the development of the term “Educational Linguistics” which includes its originator Bernard Spolsky’s activities within the Teachers of English to Speakers of Other Languages (TESOL) association during the creation of an American Association for Applied Linguistics during the 1970s, which emerged as today’s AAA. However, the AAAL, due to the width of the topic, was experiencing a crisis of identity from its inception, where the research’s connection to practice became a central topic of debate (Hult, 2008, p.12). Another central topic would become the relationship between linguistics and the field in which it was to be applied. Hult describes two initial areas of concern: the field becoming too broad to be meaningful, or too narrow to remain connected to the application (Hult 2008 p.12).

Hult & King (2011) describe the “intellectual diversity” of the field as being framed by two constant concepts: A problem-centered approach and a global outlook (p.XVII), fitting well with pragmatism. This draws on Hult’s description of applied linguistics, in which research is described as first identifying a problem and then applying the full transdisciplinary repertoire of the researcher to it (Hult, 2008, p.13). The idea of a transdisciplinary nature being inherent to applied linguistics is further described by Halliday:

I say ‘transdisciplinary’ rather than ‘inter-’ or ‘multidisciplinary’ because the latter terms seem to me to imply that one still retains the disciplines as the locus of intellectual activity, while building bridges between them, or assembling them into a collection;
whereas the real alternative is to supercede them, creating new forms of activity which are thematic rather than disciplinary in their orientation. (Halliday, 2001, p. 176, quoted in Hult & King, 2011)

Thematic becomes the keyword when the move is made from applied linguistics towards Educational Linguistics as a defined sub-field, where the theme becomes the constant while method, theory, and other components are allowed to vary. In Hult’s description of Spolsky’s motivations behind coining the term the broad nature of applied linguistics is given prominence, as a stricter focus, or thematization, would strengthen the research being conducted (Hult, 2008, p. 14).

Educational Linguistics was first named as a field by Spolsky in 1972 and was, as previously mentioned, an effort to further focus research taking place within the sphere of education with a base in applied linguistics. While the connections between applied linguistics and education were strong, the conceptualization of the connection and of the transdisciplinary work it implied was, according to Constant Leung’s reading of Spolsky, restricted (2010, p. 1). Spolsky also coined the term in reaction to what was experienced as a lack of “core” in the field of applied linguistics, which he argues lacked connections to practice (Spolsky, 2008, p. 1).

The new field is described to be broad enough to also include the new social and ethical implications of linguistics when applied in an educational setting. Spolsky draws on examples regarding language policies in colonial settings and the misuse of standardized language testing as a means of controlling migration to show the importance of these considerations when applying research in educational contexts (2008, p. 3). Hult identifies these areas as additions to previous understandings of the linguistics field, which started to appear once applied linguistics (and sociolinguistics) became more prominent parts of the field during the 1960s (2008, p. 12).

Hult also points to previous attempts to conceptualize more narrowly the connections touched on by Buckingham regarding applied linguistics, quoting Brumfit stating that “Applied Linguistics is emerging as an integrated discipline, feeding into Linguistics technically sophisticated statements about language in genuine social situations, on the one hand, and responding to the needs of practitioners, on the other” (Brumfit, 1996, p. 10). Educational Linguistics, then, would function as further narrowing efforts towards the social situations experienced while learning a, or through a, language, and the needs of practitioners engaging in teaching, or learning, (via) language.

---

7 However, the paper where the term was introduced, “The Navajo Reading Study: An illustration of the scope and nature of Educational Linguistics”, was published in 1975).
Nancy Hornberger’s 2001 reflection on 25 years of Educational Linguistics at the University of Pennsylvania provides a description of how the field has developed throughout the years since its inception while providing a further definition of what Educational Linguistics is intended to work with in practice. Through describing the development of the Educational Linguistics environment at the University of Pennsylvania, Hornberger intends to shed light on “…various aspects of the practice of Educational Linguistics at Penn and discuss them in relation to issues that have been raised in the literature about the definition, nature, and scope of the field” (Hornberger, 2001, p. 10). The definition, nature, and scope presented by Hornberger draw on work produced both by the graduates of the doctoral student environment being described and by other researchers active in the field.

Moving into Hornberger’s work, Hult describes her take on the field as: ”...Hornberger (2001) delineates three major dimensions that characterize it: it represents a reciprocal integration of linguistics and education, it provides in-depth analytical insight into a broad scope of issues related to language (and) learning, and it is problem-oriented in its focus on specific ways in which theory, research, policy, and practice inter-relate” (Hult, 2008, p. 16). In defining the field, Hornberger points to a definition produced by Spolsky in several of his texts, where he is summarized as stating that “…the discipline should focus on language education policy and implementation and that it should take a problem- and practice-oriented approach” (Hornberger, 2001, p. 4). Having a focus on problem-oriented issues that relate to actual practice strengthens the field’s connection to real-world educational contexts, while the inclusion of policy as a focal point recognizes the influences at work within educational systems. This final point also connects well to Spolsky’s examples regarding the possible negative outcomes of linguistics in the educational context, e.g., when used for gatekeeping or furthering segregation amongst students. Several others have provided similar definitions, making a strong case for Hornberger’s reading of the field (Shuy, 1981; Pica, 1994).

Hornberger also provides an overview of other types of definitions through implication, i.e., topics where scholars have included themselves under the Educational Linguistics term, such as mother-tongue learning, first language education, and second language acquisition within educational contexts (2001, p. 5). This provides some insight into the scope of Educational Linguistics, as it is shown to cover several sub-fields of applied linguistics as they relate to the educational context. Hornberger summarizes the definitions as describing “the same emphases on the integration of linguistics and education, close relationships among research, theory, policy, and practice, and a focus on language learning and teaching…” (Hornberger, 2001, p. 5).

When describing the scope of a study regarding EAL practices in England, Leung defines EAL in the context of Educational Linguistics research as “…the educational policy, provision, and curriculum practices associated with
language minority school students in England” (Leung, 2010, p. 6). This definition adds both a direct connection to this dissertation, as it concerns EAL, but also puts further emphasis on the study of policy documents and curricula within Educational Linguistics. However, it is important to note that there are many differences between the context described by Leung and the one dealt with in this dissertation, as the dissertation does not deal with an inner-circle context. Returning to Spolsky’s warnings regarding the unintended impact of research this finds a natural place in relation to the previous definition, but it also shows the importance of taking the influences on teachers’ practices into account. Even a well-researched and proven technique will be difficult to implement if it does not fit within the educational system as it is described from above, i.e in the documentation and policies controlling the practice.

Hornberger comments on the diversity of topics within Educational Linguistics by referring to the different journals on which the University of Pennsylvania Educational Linguistics researchers serve as editors or subscribe to, the diversity of conferences that their material is presented at, and the organizations they are active within (2001, pp. 17-18). Leung’s scope is also seen here, where several of the topics relate closely to policies and politics. Summarizing her overview of the field, and looking toward the future, Hornberger ends on a mission statement:

In Educational Linguistics, the starting point is always the practice of education and the focus is squarely on (the role of) language (in) learning and teaching. It is on those important differences that the argument for Educational Linguistics as a separate field rests, and it is in addressing those important challenges that the field of Educational Linguistics has its work cut out for many years to come (Hornberger, 2001, p. 19)

To summarize the scope, nature, and intent of Educational Linguistics, then, in many ways necessitates a summary of applied linguistics which is then narrowed down to Educational Linguistics before being further narrowed to the specifics of the task at hand. As has been shown, applied linguistics deals with language as it exists in many different contexts and practices outside of the strictly prescriptive and formal research areas, and signifies a move towards sociolinguistics and research into practice. Educational Linguistics moves the focus of applied linguistics into the practices within the educational sphere.

However, this must not be interpreted as the field solely working with the language being taught. Educational Linguistics also concerns itself with the role of language in the teaching, structuring, and social environment of the educational context. Much like applied linguistics, it remains broad in its focus in order to negotiate the complex situation of acquiring knowledge. Furthermore, research within Educational Linguistics cannot be allowed to
narrow its focus down to only the language, as Spolsky, Hult, Leung, and Hornberger each note the importance of the surrounding policies, hierarchies, and social factors in any given situation being studied. In fact, much of the materials referenced regarding the scope of Educational Linguistics in the descriptions previously cited deal mainly with these factors.

Moving on to the specific scope and nature of this dissertation, the final narrowing takes place based on the circumstances and intended goals. At a surface level, the current project deals with the language of upper-secondary students acquiring English as an additional language, but the previously presented nature of Educational Linguistics dissuades this simplistic perspective. When studying the texts of upper-secondary students one must also consider the contexts that influenced the production of those texts, such as the policies in effect at the school and the aims of the teachers, i.e., what are the reasons for the environment that the text was created in being the way that it is. Further complicating this situation, the goal of this dissertation is the evaluation of a tool intended to support teachers in their analysis of, and interaction with, student texts, which introduces workplace policies, curricula, syllabi, and course plans into the context. As such, it concerns language used by students, the language creating the criterion for student language use, language governing the teachers’ practice(s), and language used to efficiently communicate relevant information regarding the students’ language use to the teacher.

3.3 Second Language Acquisition

 […] We have seen that SLA is, amongst other things, not unlike first language acquisition, is a subset of general human learning, involves cognitive variations, is closely related to one’s personality type, is interwoven with second culture learning and involves interference, the creation of new linguistic systems and the learning of discourse and communicative functions of language (Brown 2000 p. 271)

Reading Brown’s summary above, Second Language Acquisition (SLA) in practice appears fairly straightforward in its goals and content. However, the second part of the quote provides some insight into the complex variables involved in actually providing a governing theory regarding HOW an additional language is acquired. Furthermore, Brown shows how this initial summary of the field is prone to be based more on common myths regarding SLA than it is based on research (2000, pp. 274-276). As a starting point, Brown identifies three viable, over-arching models for SLA in previous research: The Innatist Model, The Cognitive Model, and the Social Constructivist Model.
3.3.1 The Innatist Model of SLA

The Innatist model builds on the work of Stephen Krashen and was initially known as the “Monitor Model”, and later the “Acquisition-Learning Hypothesis”. In recent years the model has become more commonly known as the “Input Hypothesis” (Brown, 2000, p.276). In reality, Krashen proposed five different hypotheses that describe how an additional language can be acquired, amongst which we find both hypotheses that have been used as a moniker for the entire model as well as the monitor hypothesis after which it was first named (1981b). The original hypotheses were: The Acquisition-Learning hypothesis, the Natural Order hypothesis, the Monitor hypothesis, the Input hypothesis, and the Affective Filter hypothesis (Krashen, 1981a, p. 56). When writing these hypotheses at the beginning of the 1980s, Krashen was moving away from the previous audio-lingual or grammar-based classroom practices common in the U.S at the time, echoing the development previously described by Spolsky within applied linguistics just a few years earlier (Krashen, 1981a, p. 64; Spolsky, 2008, p. 1).

The Acquisition-Learning hypothesis makes a distinction between acquisition, which indicates a development of the “feel” of a language through immersion similar to L1 acquisition, and learning which here refers to formal training and knowledge of grammatical rules. Krashen states that acquisition is linked to the communicative function of a language, where a speaker can convey meaning but might not be able to account for the rules at play; they are unaware of what has been acquired, i.e. “in everyday terms, acquisition is picking up a language. Ordinary equivalents for learning include grammar and rules” (Krashen, 1981a, p. 64).

Following the line of thought apparent in the formulation of acquisition, Krashen’s second hypothesis states that learners acquire grammatical structures in a predictable order and that this order differs between L1 and L2 (Krashen 1981a p.69). However, learning also impacts grammatical prowess, as stated in the monitor hypothesis. The monitor hypothesis describes that fluency stems from acquisition, as acquired language allows the speaker to use it “easily and comfortably”, while conscious learning rather fulfills the function of an “editor or monitor” which makes minor adjustments before output (Krashen, 1981a, p. 69).

The Monitor hypothesis, however, comes with necessary conditions for it to be beneficial. In his overview of the evidence in support of his theses, Krashen finds that in order for the Monitor hypothesis to function as described the learner must (1) have the time to fully partake in self-correction, which is rarely the case in conversation, (2) be consciously focusing on the form of the language and (3) be aware of the rules at play to such an extent that they may be applied successfully (Krashen, 1981b). As will be shown later, this will come back into play when the focus of this dissertation turns to error correction in SLA.
Furthermore, Krashen states that all three conditions must be met simultaneously for the monitor to come into effect (Krashen, 1981a, p. 70).

Acquisition of grammatical rules, then, differs from the rules gained through learning apparent in the monitor. This acquisition is further defined through the Input hypothesis, which states that in order for new skills to be gained the input available to the learner must be slightly beyond the previous knowledge. The input must consist of the learners' current understanding + a new feature, often described as i + 1 (Krashen, 1981a, p. 58). This new information is acquired through the learner’s extra-linguistic knowledge; i.e:

We can do this, we can understand language that contains structures we do not "know" by utilizing context, extra-linguistic information, and our knowledge of the world. In second language classrooms, for example, context is often provided via visual aids (pictures) and discussion of familiar topics (Krashen, 1981a, p. 58).

However, in order for language to be acquired through i+1 input, the input must be comprehensible. In practice, this means that the level of input must be balanced between offering enough contextual information for the learner to be able to understand it, while also introducing something new (Krashen, 1981a, p. 62).

The fifth original hypothesis deals with the motivation and engagement of students and is named the Affective Filter hypothesis. Krashen identified three affective variables in previous research: Anxiety, Motivation, and Self-Confidence. According to Krashen, these variables play a much larger role in subconscious acquisition than they do in conscious learning, meaning that they become important for fluency in a new language (Krashen, 1981a, p. 62).

Since its inception, Krashen’s model has been quite heavily criticized on a number of points, but has also found support on others. While the baseline of the model is hard to disagree with; “...the effectiveness of providing a reasonable challenge (i +1) to students in a supportive, low-anxiety environment can hardly be denied by any teacher”, the use of sub-conscious acquisition and conscious learning has been criticized due to our lacking understanding of the actual process combined with the “fuzzy” distinction between the two in terms of definition beyond the connection to SLA provided by Krashen (Brown, 2000, p. 297; McLaughlin, 1990, p. 627).

A second point of criticism is Krashen’s suggestion that there is no overlap between acquisition and learning (Krashen, 1981a, p. 56). This would indicate that there is nothing to gain from learning towards acquisition outside of the limited function of the monitor, which would mainly be beneficial in situations such as grammar tests (Krashen, 1981a, p. 58). However, as stated by Gregg (1984):
If unconscious knowledge is capable of being brought to consciousness, and if conscious knowledge is capable of becoming unconscious – and this seems to be a reasonable assumption – then there is no reason whatever to accept Krashen’s claim in the absence of evidence (Gregg, 1984, p. 82)

Brown also refers to a multitude of studies carried out after Krashen’s formulation of the hypotheses that produced empirical evidence in support of the benefits of conscious rule learning (Lightbown and Spada, 1990, 1993; Ellis, 1997; Doughty and Williams, 1998; Swain, 1998). Rod Ellis would also argue that Krashen’s “zero option” of avoiding grammar teaching completely was not supported in literature or research in 1997, but found support for the opinion put forth by Gregg in 1984.

Regarding the Input hypothesis, criticism regarding the wide generalization of the statement has been made, and it is now preferred to discuss Uptake as the main factor in acquisition (Brown, 2000, p. 280). Uptake here signifies the parts of Input that are acquired by the learner, and as such is more specific to the actual benefits of listening to a language. Furthermore, the role of learner output is also often highlighted as missing from Krashen’s model. Swain and Lapkin (1995) suggested output is central for successful acquisition as “…it generates highly specific input the cognitive system needs to build up a coherent set of knowledge” (de Bot, 1996, p. 529).

In summary, Krashen’s model offers an approachable conceptualization of processes within a second language learner and their external factors, but the model is often criticized due to it's lacking specification and depth. It is, however, one of the preferred frameworks for language teachers due to it providing concrete and straightforward guidelines for classroom situations (Brown, 2000, p. 281). Brown summarizes the model’s role in terms of research favorably:

It is easy to see its appeal since, on the surface, the claims that are made seem to reflect accepted principles of SLA. But in their oversimplicity, the claims have been exaggerated. Nevertheless, in the final analysis, oddly enough, I feel we owe a debt of gratitude to Krashen for his bold, if brash, insights. They have spurred many a researcher to look very carefully at what we do know, what the research evidence is, and then in the process of refutation to propose plausible alternatives (Brown, 2000, p. 281)

Dayan Liu, in a 2015 review of Krashen’s Input hypothesis, argues that SLA has grown too complex for one, singular integrated theory to encompass all aspects of it (2015, p. 145). Rather, Krashen’s work survives as a starting point
for several other theories and models that are more limited in their scope but have a stronger connection to empirical evidence and are detailed enough in their descriptions to be properly tested (Liu, 2015, p. 145).

### 3.3.2 The Cognitive Model of SLA

Cognitive development as a theoretical position is often connected to the theories of Piaget and his *Origins of Intelligence in the Child* (1932). As one might infer from the title, Piaget’s work focused mainly on the general development of children, not on adolescent language learners. However, his theory is often referred to in SLA as well.

Later, an elaboration on Piaget’s initial theory explained the use of cognitive schemata as the structure containing knowledge, and changes of these schemata taking the form of assimilation or accommodation (Piaget, 1952). When a child encounters new knowledge, the information is accommodated in the existing structures and with time becomes assimilated into them. Here the process of acquiring knowledge became the focus, and in an educational context the role of the teacher became that of facilitating the learner’s exploration of new areas while supporting the process of acquisition.

The Cognitive model answers towards some of the criticism levied against Krashen’s loose definitions of consciousness and sub-consciousness by focusing on different types of processes theorized to allow humans to acquire language, while also moving away from seeing children’s acquisition as the ideal (Brown, 2000, p. 281).

The Attention-Processing model was originally proposed by McLaughlin in 1978 under the name Information-Processing model, but the version more commonly referred to was published by McLaughlin, Rossman and McLeod in 1983. The model proposes two types of processing mechanisms at play when learning a language, controlled and automatic, which manifest differently depending on if they are functioning in a peripheral or focal attention span (Brown, 2000, p. 282). Rod Ellis identifies the two central tenets of McLaughlin et. al’s model as “(1) the idea of processing limitations and (2) need for restructuring” (Ellis, 2003, p. 389).

The processing limitations of a language learner puts a limit on the amount of information that can be correctly recalled and correlated when using the language or comprehending input. Some parts of the input or output are selected as the focal point of the learner’s attention, while others are delegated to the peripheral focus (Ellis, 2003, p. 391). When using a new skill or structure the learner employs a controlled, focal attention that draws on formal rule learning, whereas a controlled peripheral attention would draw on implicit or analogic learning.

Automatic processes do not require the direct control of the learner, but instead have become routinized due to having been performed a sufficient number of times previously. Routinization thus adds to the learner’s available
language repertoire by “…making an increasing number of information chunks available for automatic processing” (Ellis, 2003, p. 390). This allows for an extension of the learner’s processing ability as chunks can be recalled and skills applied automatically, allowing the learner to focus on understanding new features in the input or applying new skills in the output. McLaughlin et. al presented the system as four cells in their 1983 publication:

<table>
<thead>
<tr>
<th>ATTENTION TO FORMAL LANGUAGE PROPERTIES</th>
<th>INFORMATION PROCESSING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal</td>
<td></td>
</tr>
<tr>
<td>Controlled</td>
<td>Automatic</td>
</tr>
<tr>
<td>(Cell A): Performance based on formal rule learning</td>
<td>(Cell B): Performance in a test situation</td>
</tr>
<tr>
<td>Peripheral</td>
<td></td>
</tr>
<tr>
<td>(Cell C): Performance based on implicit learning or analogic learning</td>
<td>(Cell D): Performance in communication situations</td>
</tr>
</tbody>
</table>


As can be seen in cell B above, automatic and controlled skills and recollections can both be carried out as focal or peripheral. However, the focus of the learner is determined by context.

McLaughlin (1990) also proposes restructuring as a second way of extending the information-processing capacity of a language learner, and as a trait often found in “expert” language learners. Ellis discusses routinization as a quantitative extension that adds “chunks of information” to the learner’s repertoire and restructuring as providing qualitative changes to the learner’s language (Ellis, 2003, p. 390). In this context, qualitative changes refer to knowledge becoming representational and the learner moving from exemplar-based performance to rule-based performance, i.e., moving from formulaic use towards simultaneously analyzing and applying rules at play.

The Implicit/Explicit model, as the name might suggest, is based on the differentiation between explicit and implicit linguistic knowledge (Brown, 2000, p. 285). Initially, this model might seem similar to McLaughlin’s Attention-Processing model. However, rather than dividing explicit and implicit by the attention and focus afforded to them by the learner, the explicit/implicit model defines the categories as explicit being “…the facts that a person knows about a language and the ability to articulate those facts in some way” and implicit as being knowledge applied “automatically and spontaneously” (Brown, 2000, p. 285).

The model also separates automatic and non-automatic processing of information, which builds on the definitions presented in the Attention-Processing model. Within the Implicit/Explicit model, the distinction between automatic and non-automatic is made based on the ease of information retrieval
for the learner. Automatically processable knowledge is easily retrieved without much effort, while knowledge accessed via the non-automatic process takes effort to reproduce and can be difficult to retrieve quickly (Brown, 2000, p. 286).

Both analyzed and unanalyzed knowledge can be accessed either automatically or non-automatically. The deciding factor here is processing time, which can manifest as fluency and response time in interactions (Brown, 2000, p. 286). Ellis argues that the combination of the Implicit/Explicit model’s definitions of explicit and implicit in combination with McLaughlin et. al’s focal and peripheral attention creates an interface between the types of knowledge where controlled knowledge can become automatic through practice (Ellis, 2003, p. 391). This would answer towards one of the major criticisms of Krashen’s hypothesis, as Krashen argued that there was no interface between learned and acquired knowledge.

McLaughlin’s description of restructuring, according to Ellis, implies the need for a third dimension to the model as exemplar-based and rule-based knowledge “…cannot be easily equated with the conversion of explicit knowledge to implicit knowledge (or vice-versa)” (Ellis, 2003, p. 391). Rather Ellis suggests that the progression from exemplar-based to rule-based should be seen as qualitative changes to the learner’s implicit knowledge over time.

Much like Krashen’s hypotheses, the Cognitive model(s) has received criticism regarding the definitions of some terms. Ellis lifts the example of ’practice’ in relation to McLaughlin’s Attention-Processing model where it is mentioned as a necessity for knowledge transfer between categories:

Does he mean ’opportunity to use a process under normal operating conditions’ (i.e. in communication) or ’opportunity to practice specific rules and items in contrived drills and exercises’? (Ellis, 2003, p. 391)

This distinction would be influential on the application of McLaughlin’s model in classroom practice as it dictates activity design on a base level. Furthermore, Ellis points out the inconclusive evidence regarding the effectiveness of exercises modeled on the second of the possible intended meanings (more on this in 3.3).

Both the Innatist and Cognitive models share a focus on the internal processes of the learner. While this perspective is important and has informed a lot of the research being conducted today, voices have also called for models incorporating external, social factors to a greater degree. In order to accomplish this, many current approaches aim to incorporate both interactions and relevant social identities into our understanding of SLA (Brown, 2000, pp. 286-287).


### 3.3.3 The Social Constructivist Model of SLA

Constructivism is based on Piaget’s description of the learner encountering new information and acquiring it into previous structures of knowledge through assimilation and accommodation, and shares a base with the cognitive model of SLA. However, as we enter into the social constructivist model, these ideas are put into a context derived from the socio-cultural perspective. It thus provides a model of acquisition that draws on previous ideas of learning, while also taking the external, social context into account. While the social constructivist model is introduced by Brown as a fairly modern model that arises from the gaps in the previously presented models, it builds on a core presented by Piaget (1952) with influences from Lev Vygotsky (1934/1986)\(^8\). However, while Piaget sees knowledge as individually constructed, Vygotsky argues that it is constructed socially in interactions.

Introducing the Zone of Proximal Development (ZPD), Vygotsky (1934/1986) argues that language becomes a requisite for development as development occurs through social interaction where the gap between the cognitive capacity and the potential development is such that new information can be understood and acquired. The gap in this description must be of a size where something new is gained, but not so large that the interaction is incomprehensible. This distance is the zone where development is likely to take place, similar to the i+1 of Krashen’s input hypothesis.

Vygotsky was interested in the mental development of children and argued that:

> Thought development is determined by language, i.e., by the linguistic tools of thought, and by the sociocultural experience of a child. Essentially, the development of inner speech depends on outside factors; the development of logic in the child, as Piaget’s studies have shown, is a direct function of his socialized speech. (Vygotsky, 1986, p. 94)

The social constructivist model of SLA draws on these prior statements by arguing that a learner creates their individual language in a socially constructed process, i.e., through interactions with themselves, other speakers, and their surroundings. As the second language is acquired socially, this puts the emphasis on these interactions as the explanatory factor in combination with the input available to the learner in any given situation.

Michael Long’s Interaction hypothesis (1980, 1991) emerged as a continuation of Krashen’s Input hypothesis and argues that “…comprehensible input is the result of modified interaction” (Brown, 2000, p. 287). Unlike

---

\(^8\) There is debate regarding whether or not Vygotsky’s theory is an addition to Piaget’s, or if it should be regarded as completely its own and disregarding Piaget (See: Tryphon and Vonéche, 1996).
Vygotsky’s general description of learning and development, Long brings the concept of socially constructed knowledge into the specific field of SLA. Modified interactions are defined as any type of modification to the input made by any speaker in order for a learner to understand and benefit from it, including slowing down speech, comprehension checks, paraphrases, and clarification/repair requests (Brown, 2000, p. 287).

According to Long’s hypothesis, input remains central to the acquisition of new language features and content, but the interaction plays a larger role in making the input processable. It also moves the setting for acquisition from the, often, artificial situations used for practice in language classrooms to the more natural setting found in everyday speech and communication. This led to what could be recognized as the modern curriculum, with interaction and communication at the core of goals and aims (Brown, 2000, p. 287). Brown (2000, p. 287) summarizes the change as: “Here, principles of awareness, autonomy, and authenticity lead the learner into Vygotsky’s (1986) zone of proximal development (ZPD) […] where learners construct the new language through socially mediated interaction”.

The construction of an L2 along the lines of the ZPD is certainly based on access to Long’s modified interactions, or already comprehensible input in the vein of Krashen. However, autonomy in regards to learners’ practice with language and awareness of what to look for in order to comprehend the input also play an important role. According to Brown, Long’s hypothesis pushed the SLA classroom of the 21st century into becoming a place where ”the contexts for interaction are carefully designed” (Brown, 2000, p. 288). However, returning to Vygotsky, this does not mean that Krashen’s ”zero-option” becomes viable and interaction is everything needed for language acquisition:

The speed of oral speech is unfavourable to a complicated process of formulation - it does not leave time for deliberation and choice. Dialogue implies immediate unpredicated utterance. It consists of replies, repartee; it is a chain of reactions. Monologue, by comparison, is a complex formation; the linguistic elaboration can be attended to leisurely and consciously. (Vygotsky, 1986, p. 242)

Because of this complexity, the genre of output for the learner could greatly influence the experienced difficulty. The quote above seems to encompass concepts such as processing time, attention type, and a monitor function of some kind.

The distinguishing feature amongst these approaches, then, is very much one of perspective. While touching on aspects of previous theoretical positions, the social constructivist position argues for the inclusion of further external influences on the learner’s language. This includes socially applied pressures,
modifiers at play in the interaction, or an internal need for identity negotiation and communication.

Scaffolding, a metaphor taken from construction work, refers to the temporary structures used by teachers when developing a student’s abilities. The metaphor was first used as a term for supportive guidance in education by Wood, Bruner and Ross (1976). It was used to describe the practices of parents who focused their children’s attention and kept them motivated, which was linked to successful language development (Hammond and Gibbons, 2001, p. 14). The main idea is to provide the student with a temporary support for completing a task, so that they are later able to carry the task out by themselves (Hammond and Gibbons, 2001, p. 15).

The connection to socio-cultural theory, Vygotsky, and Long is the use of the concept Zone of Proximal Development (ZPD). As described in the section on the Social Constructivist model of second language acquisition, the ZPD refers to the potential development of the student’s capabilities. Scaffolding then becomes the teacher’s consistent effort to support the learners’ exploration of their ZPD and support the learners’ acquisition of concepts within their ZPD.

3.3.4 SLA in the Current Project

The three over-arching theories presented above provide some idea of the diversity of thought found within SLA. However, some points of agreement can be found, albeit with different origins seen in the focus of the theories. As the focus of this dissertation is on the teaching of language patterns through automated overviews, these theories will mainly become relevant when suggesting implementations of the new tool. However, they are also relevant as guidelines for the thematization and analysis of the survey of teachers’ diagnostic needs.

The Innatist model provides an overview of the different functions that could be theorized to exist within a learner acquiring a new language. While lacking in details and containing some controversial statements, Krashen’s hypotheses allow the focus of this dissertation to be specified to the monitor function of the language learning mechanisms.

While Krashen initially argued that no overlap exists between explicit and implicit language knowledge, the cognitive models show that such processes are not only possible but rather efficient given the right circumstances. By facilitating a learning environment where learners are given the time to practice and engage with their own work, a transfer from explicit and controlled knowledge into automatic and implicit knowledge is feasible. This would also, with practice, result in students gaining the ability to restructure knowledge of language structures.

The teacher’s role as a facilitator to these processes also introduces external factors into the learner’s acquisition. The social constructivist model argues that it is beneficial. By engineering situations for the benefit of the learner,
knowledge can be constructed socially in the interactions between both teachers and learners, as long as these interactions are correctly modified. It is within this model that the work of this dissertation will take place. The tool should be situated as a way of enabling teachers to more efficiently manage the learning situation, and to modify it with greater access to detailed knowledge of their students’ language use.

However, the theoretical starting points for in-service teachers may differ, and the tool must be perceived as useful regardless. The application of the tool suggested within this dissertation will make use of the Social Constructivist model, focusing on the ZPD and the use of scaffolding language acquisition based on the data generated by the tool, but its applicability within other theoretical perspectives should also be considered in the design process.

The main function of the tool, the identification, and presentation of language issues in a group of texts will rely not only on the broader theories of SLA but must engage specifically with the research on how corrective feedback within SLA and L2 writing should be handled. The use-case for the tool explored further in this study will be the use of results from the tool in scaffolding for student texts. This allows for the tool to be used in accordance with the routines described in the interview study, and for the tool to be evaluated for use in existing practices.

3.4 L2 Writing

Second language writing (henceforth L2 writing) could be seen as the text-focused sibling of SLA, engaging with how writing and text production function within language learning. Historically, L2 writing has been a process-oriented field that has favored cognitive approaches to learning, perhaps due to the asynchronous nature of the draft-submit-feedback-revision cycle often seen in text production in the educational context (Atkinson, 2003). It has also received less attention than SLA but is increasingly being included among the perspectives in SLA research since the 2010s (Manchón, 2012, p. 6).

In addition to the inclusion of L2 writing perspectives in SLA research, theories traditionally included under the SLA umbrella, such as socio-cultural approaches, have seen increasing inclusion in research within L2 writing. This stems from cross-fertilization between the fields being an explicit goal for researchers active in either, as described by Manchón (2012).

Hirvela, Hyland & Manchón divide L2 writing into three specific orientations aimed at different perspectives (2016, p. 46). These orientations are designated as “learning-to-write”, “writing-to-learn-content” and “writing-to-learn-language”. While aimed at different aspects of the writing act, these three orientations are all connected to the purposes of learning a language in a context.

The “learning-to-write” (LW) approach has a focus on the writing itself, and is further categorized into three groups according to Hyland (2016, p.46). These
groups are categorized by the focus of the writing activity, and are “focus on the writer”, “focus on the text” and “focus on the reader”. The first category takes a process perspective on the writer’s activity and is aimed at teaching “good” writing strategies for the learner to model their own process on. The teacher facilitates the process by avoiding the initial focus on form and instead guides the learner through the text generation process, including drafting and revision.

“Focus on text” instead disregards the process in favor of focusing on the product of the activity. The text is seen as an object created by the writer’s ability to use grammatical rules and vocabulary to craft an expression of their message. Following this line of thought, the text is seen as an attempt to encode the writer’s intended meaning. In early takes on this perspective, texts were seen as non-contextual and independent, but a more current take views the texts as a discourse in which the anticipated reader, the social context of the writer, and the intention of the writing (such as a prompt) interact to create the structures seen on the page (Hyland, 2016, p.48).

Following the intention of the text fully, the “focus on reader” perspective embraces the influence of the intended reader of the text and sees the writer’s encoding of their intended meaning as entirely for the purpose of the intended reader(s)’s comprehension. This means viewing writing as an act of expectation and seeing the text as created for a specific community with caution taken regarding the norms of said community of readers.

The “writing-to-learn-content” (WLC) orientation relies on the concept of transfer, i.e., the ability to apply a learned behavior in a new area (Hirvela, Hyland and Manchón, 2016, p. 52). However, WLC is not to be seen as only a way for students to demonstrate what they have learned, but rather as a way of acquiring and synthesizing knowledge for different means. This includes syntax and lexis, but also the social constructs in which texts exist as well as the processes used in text generation. Furthermore, the concept of transfer is not seen as only applicable to text generation in an L2 classroom, but as something that can be applied in other subjects as well, such as notetaking in a history classroom or the preparation of a presentation in natural sciences. This is where one might consider the intersections between the WLC orientation of the L2 writing field and the concept of subject literacies, i.e., language use specific to subject disciplines (Danielsson and Bergh-Nestlog, 2022).

The perspective taken in “writing-to-learn-language” (WLL) more explicitly includes theoretical approaches taken from SLA and becomes the main orientation for the previously mentioned transdisciplinary efforts (Manchón, p. 55 in Hirvela et. al. 2016). The lynchpin argument regards the temporal aspect of text generation as a central component to the benefits of including writing as a part of the language learning experience. As a learner engages with text generation, ample time is available for experimentation and revision as the engagement with the reader is asynchronous. However, for spoken production
the production and comprehension take place simultaneously, leaving no room for such processes as described by Vygotsky.

As the current dissertation deals with the design of a tool that interacts only with the text, and since the results are intended to be communicated to a learner only through the facilitation of the teacher, the dissertation could be said to fall within the “focus on text” perspective. However, this should not be construed as disregarding the other perspectives from a didactic or pedagogical standpoint. Rather, the position taken in the current dissertation is a result of recognizing the limitations of a computer parsing human language, and an emphasis on the importance of a human mediator negotiating the results produced by the tool.

3.5 Corrective Feedback

The use of corrective feedback (CF) regarding grammar in second language acquisition is a contested topic, with the two opposing sides in current research showcased here through the debate between John Truscott (1996) and Dana R. Ferris (1999), although it was discussed earlier as seen in the description of the innatist model of SLA. Truscott’s 1996 article “The Case Against Grammar Correction in L2 Writing Classes”, published in Language Learning, argues three main points: (1) CF is shown to be ineffective, (2) CF should be expected to be ineffective based on theory and (3) CF has harmful effects on learners. The article is contextualized as a counter-reaction to the assumption that SLA writing instruction must contain grammatical corrections.

The ineffectiveness of CF on grammar is argued for based on previously conducted studies by Cohen & Robbins (1976), Semke (1984), and Rob, Ross & Shortreed (1986). The results from those studies indicate that no improvement in accuracy or fluency could be found in groups where grammatical errors were corrected, regardless of target language (Truscott, 1996, pp. 330-331). Sheppard (1992), Semke (1984), and Kepner (1991) all experimented with different types of CF, such as sentence level feedback, correction codes, and teacher-student conferences with different focus areas, but found that the groups that received corrections on grammar did not develop a more accurate language than control groups who were only given feedback on content (Truscott, 1996, p. 332).

Furthermore, Sheppard’s study indicated that the students who received detailed grammatical corrections employed avoidance tactics in order to avoid corrections, resulting in students limiting their language use to simpler structures (Truscott, 1996, p. 333). Semke’s study also showed significantly weaker results in the student groups who were given grammatical corrections, with the group where errors were only marked for self-correction performing at the lowest level (Truscott, 1996, p. 332). This is in line with the study performed by Frantzen & Rissel (1987) as well, as they showed that students without explicit information about the grammatical error in a text are unlikely to identify
its nature, even with the location marked (Truscott, 1996, p. 332). Gass (1983), Seliger (1979), and Sorace (1985), however, found that students’ intuitive understanding of a language’s grammatical rules far outweighs their formal understanding of the same system; i.e. students can apply grammatical rules without being able to state the actual rule (Truscott, 1996, p. 348).

Truscott acknowledges that the timing of these corrections could be problematic, as they do not take the natural learning sequence of grammar into account or are unable to produce a learning environment that adheres to them9 (1996, p. 345). Although Truscott attributes this issue of timing to earlier studies on morphemes, the natural acquisition sequence is today mostly related to Processability theory (See: Pienemann, 1998).

However, arguments are also put forth as to why grammar correction in SLA is ineffective in general according to acquisition theory. The main issue here is that L2 learning is a complex sequence of acquisition, as seen in the previous section, while grammar CF is often based on a simpler model of knowledge transfer between teacher and student:

Consider what is probably the standard view of correction: Learners find out that they are wrong in regard to a particular grammatical structure and are given the right form (or directions for finding it); they then have correct knowledge about that structure, so they should be able to use it properly in the future, assuming only that they understand and remember the correction. (Truscott, 1996, p. 342)

The idea of language learning depicted in the quote above is in contrast to language acquisition as depicted in theory, where it is described as a gradual process of understanding, application, and interim solutions for communicative purposes. Grammar acquisition, more specifically, is also understood by Truscott (via Long, 1991) as “…a gradual process, not a sudden discovery…” (Truscott, 1996, p. 342).

Finally, Truscott arrives at the issue of the aims of language learning via Rod Ellis’ (1988) distinction between modeled and communicative learning, continuing onto Ellis’ (2003) later 1994 distinction between explicit and implicit knowledge of grammar (Truscott, 1996, p. 346). These are connected to Krashen’s (1987) distinction between learning and acquisition, where the first applies to formal knowledge of grammar rules and the second signifies actual fluency and accuracy in production in the target language through familiarity. Ellis’ (2003) text proposes that explicit grammar knowledge can aid in the acquisition of implicit grammar knowledge, thus leading to better fluency in the

---

9 Truscott’s argument here is based on that not enough was known about developmental sequences at the time of writing for any meaningful teaching to be based on them.
language, while Truscott was critical of the idea of formal, grammatical knowledge being useful (Truscott, 1996, p. 346).

Ellis (1998) himself responded to Truscott’s essay in the discussion regarding negative feedback, i.e. error correction, in his article “Teaching and Research: Options in Grammar Teaching” published in ESOL Quarterly. The aim of that article was to create a discussion regarding the different types of knowledge employed by teachers: Technical and Practical. Technical knowledge is the theoretically justified type of declarative knowledge gained through deliberate studies and research on a topic. It proposes general statements, but is difficult to apply in the rapid taking of decisions involved in teaching (Ellis, 1998, p. 40). This type of knowledge is connected to the explicit knowledge referred to by Truscott, and it stands to reason that difficulty of implementation would apply similarly to the teaching techniques described by Ellis and the grammatical rules discussed by Truscott. Practical knowledge instead refers to the procedural, intuitive knowledge that is drawn upon for quick decision making, for instance during a lesson or while partaking in conversation. Although hard to state declaratively, it makes up the knowledge of practice in a situation and guides correct procedure intuitively.

In his review of studies conducted on CF regarding grammar, Ellis finds that corrective feedback has been found to promote interlanguage development in classroom settings, through student uptake, the use of recasts, or through the use of contrastive strategies (1998, p.53). It is important to note that Ellis draws from material concerning different modes of student production (spoken, lab reports, written) while Truscott only discusses CF on written production. However, in summary regarding CF, Ellis states that the transfer of the teachers’ technical knowledge into practical knowledge and practice is difficult due to the complex situation in the classroom, and the general nature of the framing theories: “Indeed, given that error correction involves attending to a variety of social and affective factors (See: Allwright, 1975), technical knowledge about what works best for language acquisition can never provide a complete basis for correcting errors” (1998, p.53).

In 1999, Ferris published “The Case for Grammar Correction in L2 Writing Classes: A Response to Truscott (1996)” in the Journal of Second Language Writing, where she critiques Truscott’s paper on grounds of lacking definition of grammar errors and overstating the negative implications of previous studies. It draws partially on Ellis’ 1998 study discussed above, but also introduces several new studies performed during the second half of the 1990s to the discussion. Ferris also agrees with Truscott on several points, for instance, the practical issues of applying effective error correction: “Even if teachers do recognize an error, they may not be able to explain the problem to the students, either because of lack of knowledge or lack of time” (Ferris, 1999, p. 3). This argument is continued by both Ferris and Truscott to mean that teachers often
do have the theoretical knowledge, but rarely have the time needed to apply it when correcting student writing (Ferris, 1999, pp. 6-7).

Aside from Ferris’ assertion of Truscott overstating the negative implications of previous research, she also argues that the heterogenous samples represented in the studies make a contextualization necessary. Truscott acknowledges the diverse samples of the studies, but does not connect to this in his conclusion. Ferris also states that the previous research cited by Truscott involves diverse groups of subjects and comes from different paradigms, both in terms of research philosophy and the instructional ideas amongst the teachers involved (1999, p. 4). This diverse sample might also cause issues due to the necessity of individualized intervention based on the students’ prior experiences, L1, and motivation (Ferris, 1999, p. 7).

The temporal aspects of the studies summarized by Ferris (1999) also varied, as some covered longer durations and several lectures, while others attempted single interventions. Several of the studies were also not operationalized to support or disregard grammar error correction, but instead only provided incidental data on the topic. To summarize, Ferris writes that “because of these key differences in subjects, research design, and instructional methods, it is virtually impossible to support any generalization other than the cliché ‘Further research is necessary’ from this group of studies” (1999, p. 5). Ferris also points out multiple studies that have indicated that CF on grammatical errors can be beneficial for second language learners, such as Fathman & Whalley (1990) and Lalande (1982).

Ferris argues for grammar error correction being useful in some aspects of language learning, based on previous studies and experiences (Ferris, 1995a, 1995b, 1997). However, the effectiveness of students’ self-editing of written production, according to Ferris, hinges on three criteria: (1) That students are focused on the importance of editing, (2) are trained to identify and correct patterns of frequent and serious errors and (3) have been given explicit teaching on the grammatical rules relevant for those errors (1999, p. 5). The argument is concluded with the assertion that indirect error correction, as in performed by students self-editing, is preferable to direct correction by the teacher. However, Ferris recognizes that this type of intervention is only functional when students’ errors occur in a rule-governed pattern, while other errors appearing randomly need other types of intervention. The examples given for rule-governed patterns of error by Ferris are subject-verb agreement, run-ons, comma splices, missing or incorrect articles, and verb forms (1999, p. 6). The examples are derived from first-semester EAL students at university during a language proficiency course. Returning to Ellis’ conclusion that the transfer of the teachers’ technical knowledge into practical knowledge and practice is difficult, Ferris (1999) proposes that the gap should instead be bridged within

---

10 Bold for emphasis in original
the students, as they convert their technical knowledge about grammatical rules to procedural knowledge through self-editing.

The future research proposed by Ferris in 1999 is aimed at defining the context and impact of CF on L2 learners’ grammar errors. The specific questions posed were:

Do teachers respond accurately to students’ errors? Does training and practice help them to do so more effectively?
Are students more able to make progress in monitoring for certain types of errors than others (e.g., morphological or syntactic errors versus lexical errors)?
Which individual student variables affect learners’ willingness and ability to benefit from error correction, and can student problems be mitigated by thoughtful pedagogical practices?
Which methods, techniques, or approaches to error correction lead to short- or long-term student improvement (assuming that student, teacher, and contextual variables are adequately controlled for)? (Ferris, 1999, p. 9)

In 2004 Ferris published the position paper “The ‘Grammar Correction’ debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime...?)”, bringing the questions of CF into the 21st century. The current state of grammar error correction in 2004 is summarized by Ferris as three observations: “(1) the research base on the ‘big question’—does error feedback help L2 student writers? — is inadequate; (2) the previous studies on error correction are fundamentally incomparable because of inconsistencies in design; and (3) existing research predicts (but certainly does not conclusively prove) positive effects for written error correction” (Ferris, 2004, p. 50).

Although still considering the research inadequate, Ferris makes three predictions based on theoretical knowledge from SLA and L2 writing research, in combination with her own experience: (1) Adult learners will fossilize and experience stalled L2 development without explicit instruction and error feedback, (2) students who receive CF on grammar will be more likely to self-correct, which results in demonstrated uptake that will likely benefit long term language competence and (3) students are likely to attend to and appreciate feedback on their errors, which can motivate them to make corrections and work harder on improving their writing (Ferris, 2004, p. 56).

---

11 Truscott published a rebuttal to Ferris (1999) in 1999, but it mainly reiterates Truscott’s interpretation of previous studies and acknowledged the agreements of research being inconclusive and not fit for the purpose of answering whether or not CF on grammar is beneficial. As Ferris (2004) elaborates on the same issue with new support, Truscott’s 1999 rebuttal has been omitted from the literature review.
Half a decade later, Ferris (2010) provides an updated summary of the progression on the topic, which has been actively researched by both SLA and L2 Writing researchers, thus providing parallel sub-disciplinary studies (pp. 183 - 184). Drawing on work conducted by Ashwell (2000) and Chandler (2003), as well as one collaboration between herself and Roberts (2001), Ferris provides evidence for the beneficial effects of written CF in controlled forms. Similarly, studies performed within SLA by Bitchener & Knoch (2008), Ellis et. al (2008), and Sheen (2007) also indicate positive effects on students’ written production. Summarizing the new studies Ferris arrives at the conclusion that “as a body of work, they provide consistent and compelling evidence that written CF, under the right conditions, can facilitate L2 development and help students improve the accuracy of their writing, at least for the particular features under consideration” (2010, p. 186).

The different starting points of SLA and L2 writing explain some of the differences in operationalization that make it difficult for literature reviews to arrive at a conclusion regarding grammar-oriented CF, for instance many of the studies indicating positive results involved students revising their texts after receiving feedback (Ferris, 2010, pp. 188-189). The efficiency of such approaches when it comes to long-term language development and development carried over to new pieces of writing becomes difficult to prove, but Ferris argues for self-editing and revision as an important step towards acquiring new language features (2010, p. 189). This is supported by the added benefits of indirect CF, including developing the students’ meta-cognitive skills and ability to improve their own language once outside the classroom (Ferris, 2010, p. 190).

The practical issues related to CF on grammatical errors are highlighted as one of the major problems with implementation (Ferris, 2010, p. 193). These include the energy, time, and expertise required to fulfill the prerequisites for CF to be effective according to the research summarized by Ferris. Although Ferris considers the research available now, from both SLA and L2 writing researchers, to show how CF can be used beneficially, the question of how to implement it in a way that actually works in practice remains (Ferris, 2010, p. 194). Drawing on the research from both sub-disciplines, Ferris proposes a research design built to discern the effects of the types of CF indicated as beneficial for students’ language development:
The design shown in Table 4 can be seen as describing how to answer the questions posed by Ferris in 1999. In 2010, Ferris tentatively answers the first question, “Do teachers respond accurately to students’ errors? Does training and practice help them to do so more effectively?”, with the tentative answer that training is not the missing part for teachers to be able to provide beneficial CF, but time, energy and expertise is. The fourth question, “which methods, techniques, or approaches to error correction lead to short- or long-term student improvement […]?”, has its answer indicated as being tied to the learners receiving relevant feedback in a comprehensible amount, so that they can properly engage with the new information. This also means that the feedback must be given in relation to the level of the learners.

Ferris (2010) further indicated that this engagement should take place through revision and self-editing, while possible positive effects cannot be reliably measured until the learners compose a new text. This means that the design in Table 4 requires two separate pieces of text to be written by the learners in order for feedback to be shown to have had any effect on language development. This issue of ascertaining efficiency is an interesting point to consider regarding feedback, as how participant efficiency or perceived benefit is measured might alter the outcome quite drastically (Price et al., 2010).

Table 4: Design adapted from Ferris (2010, p. 194)
However, efficiency is also a very important component in the argument regarding whether or not teachers should invest the large amounts of time and effort needed to provide focus-on-form feedback. To quote the title of Price et al.’s paper: “Feedback: All that effort, but what is the effect?” (2010).

Self-revision stands out as interesting due to its use in Ferris’ framework for text interactions with CF (2010). CF as scaffolding would rely heavily on the teacher acting in a supporting manner when providing the information used for self-revision, as the information would need to be curated and sorted in accordance with the learners’ individual ZPDs, i.e., timing stands out as important.

In summary, the research on CF in SLA and EAL indicates that in order for the feedback to be effective several requirements must be fulfilled in terms of form, timing, delivery, and evaluation. The corrections must be given in context of the students’ own production, must be explained in terms of applicable rules and opportunity must be given for students to re-engage with their writing to correct the errors within a fairly short time frame.

These results indicate the importance of scaffolding throughout the CF process and fit well with the socio-cultural perspective on language learning chosen for the current study. The results also indicate how CF could be implemented through the tool, as the connection to scaffolding informs the design about what a beneficial implementation should look like, beyond the nature of the computational results. In order for the tool to be successful, it must not only be functional in terms of text analysis, but it must also support the teacher in scaffolding the students’ learning experiences.

### 3.6 Computer-Assisted Language Teaching

Computer-assisted language learning (CALL) is defined by Michael Levy as “…applications of the computer in language teaching and learning” (1997, p. 1). The field has been active since the 1960s, and initially focused on programmed instruction with call-and-response-type exercises facilitated by a computer (Davies, 2016). From the 1980s forward, CALL has branched into other types of learner interaction and exercise types. Graham Davies, editor of Information and Communications Technology for Language Teachers (ICT4LT.org)\(^\text{12}\), divides CALL into four broad categories: Traditional CALL, Multimedia CALL, Web-based CALL, and Explorative CALL (2016).

Traditional CALL fits with the description of early CALL practices in the previous paragraph, and is presented as being a “…teacher-centered, drill-based approach to CALL” (Davies, 2016). Multimedia CALL came about with the computer boom and improved accessibility to recordings through personal computer systems.

\(^{12}\) ICT4LT.org is a collection of web-based training modules for ICT aimed towards language teachers and was initiated through funding from the European Commission.
computers. This category also includes other ventures into media for the purposes of language acquisition such as video games and automated interactive exercises.

While initially seeming similar to Multimedia CALL, Web-based CALL endeavors to exploit the world wide web for teaching and learning purposes. Davies’ categorizations show their age here, as one of the main distinctions made between Multimedia CALL and Web-based CALL is the ease of access to audio and video afforded by DVDs and CD-ROMs (Davies, 2016). This distinction can no longer be said to hold true due to the massive expansion of material available online via streaming and video-hosting services. However, the other main distinction holds true in that web-based approaches allow students to engage with a wide selection of interactive materials that are maintained, curated, and shaped continuously.

Finally, Explorative CALL moves away from the teacher-centered and drill-based approach towards a learner-centered application of language analysis software (Davies, 2016). Tim Johns (1991) describes this as Data-Driven Learning (DDL), where software is used to create data, which is interpreted to guide the learning environment.

The idea of automating language analysis was around even in traditional CALL. Davies presents the Computer-assisted Lessons for French (CLEF) developed by Camsoft as an early example of automated error detection and feedback (2012). CLEF works by presenting the learner with exercises on clearly defined topics and then comparing the answers provided by the learner with a library of feedback messages to provide correct output.

Error detection performed as it is in CLEF becomes limited to the exercises included as the software can only anticipate the common errors occurring in scripted language use-cases, for instance fill-the-gap type exercises. Error detection and feedback have developed further since the release of CLEF, for instance through the application of artificial intelligence within Intelligent CALL (ICALL), but the use and development of such applications have met with some resistance. The debate during the 1990s was summarized by Davies as “…there is a gulf between those who favor the use of AI to develop CALL programs (Matthews 1994) and, at the other extreme, those who perceive this approach as a threat to humanity (Last 1989)” (Davies 2016).

Returning to Data-Driven Learning, Johns (1991) describes the placement of the computer in the educational context not as a surrogate tutor, but as an informant. The issue of how to create software for that type of approach is also highlighted:

If computers are to act as informants the problem is how to get the machine to respond to learner-generated questions. The obvious

---

13 This is further elaborated on in regards to how the multi-mediality of web-based environments impact learners in Godwin-Jones (2010).
answer is that we should try to make it as intelligent as possible: that we should build up a system of rules (morphological, syntactic, semantic, sociolinguistic ...) that would allow the computer to act as an "expert system" on the language in question. (Johns, 1991, p. 1)

However, Johns deemed the results of such applications disappointing at the time of writing. The core issue being that “…the ‘rules of language’ cannot with any plausibility be specified in computer terms to allow the machine to carry out as apparently simple an ’informant’ task as correcting the English of a piece of student writing” (Johns, 1991, p. 2).

A later overview of CALL by Chapelle (2003) brings up assessment as an issue to do with routines, i.e. an assessment tool that necessitates a departure from the learners’ routines of production accomplishes little and technological solutions are rarely “invisible” (Chapelle, 2003, p. 28). The issue is whether assessment procedures that are different from the normal procedures of production can actually provide entirely valid information about the production.

In an earlier paper on the position of CALL in relation to SLA research paradigms, Chapelle (1997) discusses how learner output can be engaged with through CALL and highlights how corrective feedback would function in such a setting drawing on Lightbrown & Spada (1990). Chapelle (1997) shows examples of two applications drawn from classroom assignments. The examples are placed within Long’s Interaction model of SLA (1985), and the second is summarized as a “…type of interactional modification believed to facilitate SLA is one that interrupts the normal interactional structure which is working toward a communicative goal to focus on linguistic form, that is, overt correction of linguistic errors learners produce while they are working toward communicative goals” (Chapelle, 1997, p. 29). Reading Chapelle (1997) in context is important here as the paper was published during the 90s, and one of the main conclusions is the importance of allowing actual language use by the learners during CALL activities rather than giving them the possibility of communicating entirely through mouse-clicks, as was the case in one of Chapelle’s examples.

Moving towards the focal practices for this dissertation, i.e., text interactions, Sauro (2009) explores how CALL could change with the development of synchronous feedback, comparing recasts and meta-linguistic feedback forms. A meta-linguistic feedback form includes explicit information regarding the error while a recast is a response using the same form correctly. Both of these feedback forms build on Schmidt’s (1990) Noticing Hypothesis, which suggests that for language learning to occur the learner must be made aware of and engage with a difference in their output and target-like input. This fits into corrective feedback as “corrective feedback, by juxtaposing learning output with input, can assist the acquisition of certain hard-to-learn forms by increasing
the likelihood that they will be noticed” (Sauro, 2009, p. 97). The results of Sauro’s study indicate that the meta-linguistic form resulted in a higher efficiency during an immediate post-test, but that both meta-linguistic responses and recasting yielded similar efficiency at a delayed post-test.

Drawing on the history of CALL, this dissertation goes back towards a traditional approach to computational applications in the classroom. The focus on longer student-produced texts of different types is the main departure from previous efforts in the vein of CLEF, which focused on shorter answers that could more easily be anticipated. As Johns (1991) pointed out, the issue with approaches similar to the one taken in the current project has been the inadequacy of software when it comes to handling the unpredictability of student texts in combination with the complex patterns that shape language. In the thirty years since Johns’ publication, software has developed rapidly, and the analysis of student texts is no longer an impossibility.

Boulton’s (2010) overview of DDL and corpus-linguistics in language teaching highlights several issues that stand in the way of bringing these practices into the language classroom. Drawing on Römer’s (2006) wish list on the topic of widening the influence of corpora and DDL on language teaching, Boulton concludes that DDL needs to be included in pre-service training and be made more accessible, but most importantly be normalized as a teaching practice and demystified (Boulton, 2010, pp. 3-4). However, even when it is included in teacher-training programs students tend to focus on topics more directly related to passing their program requirements (Boulton, 2010, p. 4). This is in addition to the technical and pedagogical issues apparent in introducing corpus methodology as a classroom practice (Johns, 2002).

In order to counter these problems, the tool designed for this project must make use of current, up-to-date technologies capable of handling the complex issue of interim language analysis. Another important aspect of the solution proposed herein is to focus on fewer patterns that often result in feedback rather than attempting a complete automation of the language analysis. The current project also shifts the role of the computer from being an informant for the learner to being an informant for the teacher. Furthermore, the tool must not in itself pose a hindrance to the teachers by being inaccessible and demanding too much technical competence to operate.

Mukherjee’s (2006) selective overview of the intersection between corpus linguistics and language pedagogy provides 3 categories of application: (1) Producing learner corpora, (2) using corpora in the classroom for DDL and (3) using learner corpora (Mukherjee, 2006, p. 1). Local learner corpora, especially, are highlighted as an approach that is likely to be fruitful in introducing corpus linguistics to classroom practices in a way that the teachers would be responsive to (Mukherjee, 2006, p. 7). Local learner corpora are corpora compiled from the students’ written production and are used to create a dataset representative of the learners’ language use.
It is important to remember that the tool designed for the current project is not proposed as a stand-alone solution, but as an augmentation of a classroom practice carried out by a teacher. This design decision was also intended to enable the tool to work in combination with already applied pedagogies leading to a more efficient pedagogical integration, highlighted as necessary for automated writing assessment implementations by Ware (2018). The main goal is to enable the teachers to do their work more efficiently and with greater ease, but the application of the tool still relies on their individual expertise and judgment.

3.7 Teacher-Computer Interaction

Human-computer interaction (HCI) research deals with how humans communicate with, and through, computers, and as such is a part of the broad field of computer science. Exploring more explicit definitions within the field of educational technologies, Guney (2019) summarizes two related definitions by Ipek (1995) and Earnshaw et. al (2018): "HCI deals with the study of people, computer technology and techniques, which influence each other" (Guney, 2019, p. 732) and "an area of research that seeks to understand the interaction between technology and the people who use it" (Guney, 2019, p. 732). The former of these definitions takes the impact of computer technology and techniques on an existing practice into account, while the latter is more narrowly focused on the interactions themselves. In summary, the perspective afforded by HCI is not about what computers can do, but rather on what humans can do and what is needed for them to be able to do so. The two perspectives on HCI highlighted by Guney’s summary are the usability and the user experience (2019, p. 732).

As the current project engages with HCI in a very specific context, that of the upper-secondary language classroom, the human in question will be a teacher. The current project will refer to the narrowed version of the field as Teacher-Computer Interaction. Drawing from the descriptions of HCI above, the area of interest becomes how teachers could interact with a computer for the purposes of diagnostic feedback and how this interaction could influence already existing practices. Beyond the general definition of HCI provided in the previous paragraph, this section will narrow the focus down to research dealing directly with teacher-computer interaction. The focus will then be broadened again as the section moves into methods for designing digital tools.

A meta-study produced by Teo & van Schaik suggests that the most powerful influence on the intention of using a new technology is the teachers’ attitude towards the tool (2012). Intention to Use is here to be understood as the likelihood of a teacher engaging with the tool without explicit policy to do so top-down. Due to this, the positioning of the tool as a supporting function for teachers within their already established routines and preferences is key to
implementation being successful. Similarly, perceived usefulness and ease of use are important influences on attitude and must be considered both during design and field tests.

The Interaction Design Foundation provides guidelines on how to achieve User-Centered Design (UCD) through an iterative process. The process involves both generative methods, such as brainstorming and related previous examples, and investigative processes, such as interviews and surveys (Interaction Design Foundation, 2019). The schematic below is adapted from the definition of UCD on the Interaction Design Foundation’s website and the investigative/generative tag has been added.

![Diagram](image)

*Figure 4: Adapted from the Interaction Design Foundation definition of UCD 2019*

The full-lined arrows in between the blocks indicate the workflow, with the dotted lines connecting the blocks from outside of the workflow showing iteration through the processes. After the tool has been evaluated against the user requirements, either through experimental testing or field studies, the development process iterates back through the previous blocks in order to adapt them to new information gained during the evaluation.

As can be seen in Figure 4 above, the connection to the practices surrounding the intended implementation is the first step of the development process. This aspect must be at the forefront of the design, and must be accounted for rigorously (Lazar, Feng and Hochheiser, 2010, p. 4). Once the context is understood, user requirements can be collected and interpreted from their context. In the current study this is achieved through a combination of interviews and previous research. By drawing from these two information categories the design of the tool can begin.

The design could be split into two parts: Function and Interface. The function of the tool relates to what is needed to perform tasks with sufficient precision and efficiency. In the project described within this dissertation this means text analysis and language error identification, but also the creation of output material that communicates this to the user in a comprehensible and actionable form. All of this must be designed in accordance with the user requirements and the context. For the current study the main functionality will be inherited from the language checker chosen for the tool, which will then be narrowed down to
accommodate only the elements that are indicated as relevant by the interview study and the previous research on SLA and CF. As previously mentioned, the tool will also need functionality that allows the results of the program script to become visible and exportable, i.e., comprehensible and actionable for the teacher.

In order to create an idea of the function of the tool to iterate over during the design process, a flowchart that details the intended functions and how they flow into each other was created. By using the top-level flowchart as a guide, more detailed flowcharts can then be created for each of the components separately. The flowcharts in the current study are created using the DRAKON modeling language, as its emphasis on transferability and readability fit well with the motivations behind the project (DRAKON Labs, 2018).

Returning to the interface as the channel between human and machine, Treu (1992) describes interfaces as a combination of conceptual, logical, and physical patterns. Guney (2019) summarizes the role of the interface as supplying “a direct visual and language-based interaction between user and computer” (Guney, 2019, p. 733). As such, creating an interface for the tool in the current project must draw from the teachers’ needs, experiences, and routines in order for the intended interaction to function well, and must also be created using both visual and textual signage to clearly indicate functions. This signage must adhere to pre-existing patterns in the interaction with student texts, so that the conceptual pattern of the interaction is represented in the physical pattern of the interface, and thus fits within the logical pattern that the teacher is accustomed to when performing the task.

In order to design an interface that invites teachers to use the tool and allows for interactions that are useful, the interface must build on both the function as intended and the routines of the teachers. This will be explored using wireframes to break the interface design into smaller components that can be shaped individually with explicit connection to which function influences the form of each component. An example of a wireframe based on the Google search window is shown in Figure 5.
According to the Interaction Design Foundation’s entry on the topic of wireframes, they can be defined as an overview of an interactive product that serves to “…establish the structure and flow of possible design solutions” (2019). It could take the form of a sketch on paper, or be created using a digital tool. The sketch is created by first drawing a mock-up of the window within which the interface is placed, and then segmenting that window into the different components that make up the interface. Each segment is then labeled according to the function it governs by the use of a symbol, and a list detailing the function, input, output, and purpose of the segment is then created. By doing so the interface is tied to the functions in a comprehensible way that allows the designer to adjust the design and make sure that all user requirements are fulfilled through the interface’s options of interaction. While the interface can create a channel of communication between human and machine, it must also steer use towards the intended purposes and allow for the human to intuitively perform the tasks needed. Nielsen (1994) elaborates on usability in a way that brings these aspects together, which will be the perspective on interfaces used in the current project. This makes the metrics used to evaluate the design important, as classic metrics such as error rate, task fulfillment and time per task are not entirely fitting if they do not relate directly to the main points of the project (Lazar et.al, 2010, p. 6). For this project, metrics that relate to the perceived usability of the tool and the experienced applicability in the classroom will be used during field tests. These results were collected through debriefs to allow for participant elaboration.

A concept within HCI that is gaining in popularity is the use of Participatory Design, which means that the target group for a project is involved in the design and development (Hayes, 2011). In the case of the current project, this means involving in-service teachers throughout the process, both while planning and
evaluating the diagnostic tool. The main influence will be the use of interview materials in the design, but changes to the design will also be made during the field tests of the tool through the previously mentioned debriefing sessions. This process also fits well with the definition of User-Centered Design (UCD) as described in Figure 4.

Furthermore, the Substitution, Augmentation, Modification, and Redefinition (SAMR) model adds an education-specific taxonomy for the classification of digital tools and activities, which enables insights into the impact of new tools in the school environment (Puente, 2010). Puente’s SAMR model supports the placement of new technology in the teaching situation into categories based on the impact of the new technology, which allows for a formalized understanding of the change and impact the addition could result in, which ties into considerations regarding the usability and intended use-cases. Substitution, as the name suggests, categorizes technologies that perform the same tasks as would have been carried out without a computer with no functional changes, for instance the use of a computer to print out a worksheet. Augmentation retains the function of the task, but technology is used to more efficiently carry it out. A prominent example of augmentation is the use of digital documents for submissions and feedback. Modification and Redefinition both refer to additions of technologies that do not supplement traditional classroom practices but transform them. Following the UCD process based on the interview study, the tool can be placed within the SAMR model and the practical usage can be further understood. In addition, the placement of the tool within the SAMR model as indicated by actual use can provide valuable information for future designs based on a similar concept.

3.8 Background Summary

This dissertation fits within Educational Linguistics in regards to thematization, being problem-centered, and applying a transdisciplinary approach. While focused on learner language, the context of the upper-secondary school system requires further breakdown into its components in order for an analysis to be functional. Those components include the policy documents and syllabi within the Swedish school system in general, the documents governing the English subject in particular, the experiences of current teachers, and the theoretical positions within second language acquisition and corrective feedback within SLA and L2 writing. It is also important to take the warnings regarding possible unforeseen implications to heart, as the tool proposed in the current study could result in an oversimplification of the feedback process if implemented improperly.

All of the previously described models, to some extent, allow applications within corrective feedback. Although the theories apply to much more than just written production, they all offer different entry points regarding it. Reading
through Ferris’ exploration of the topic, it seems that the specific requirements for corrective feedback are discernable in the SLA theories. While the final verdict regarding the efficiency of CF in SLA writing has yet to be agreed upon, the traditional use of the method and the research in support of its implementation indicates that it will remain relevant in the foreseeable future of SLA education.

Ferris argues that CF in SLA is likely only effective regarding rule-governed errors, such as sentence structures, articles, plurality, and tense. From the perspective provided by cognitive and social constructivist theory, this makes sense, as long as students’ acquisition of these patterns is facilitated in accordance with the learning processes. The students also require time for practice, interactions designed for their benefit, and opportunity to focus on applying the patterns.

Regarding the interactions, the meta-study by Biber et. al (2011) indicates that feedback is more effective when provided by peers. The social constructivist model, however, emphasizes the importance of knowledge being within an individual learner’s ZPD and that interactions should be modified accordingly. Accomplishing this with peer feedback puts high demands on the teacher facilitating the environment. The research on CF also indicated self-editing based on comments as effective, as long as comments were made in relation to the learners’ levels.

The combined model of evaluation proposed by Ferris fits well with the case studies that make up the majority of empirical evidence in this dissertation. By using Ferris’ combined design, the case studies will be able to show the possible effects of the tool on students’ writing that can later be compared to future studies, while also encompassing the requirements indicated as important for efficient CF.

Situating this background in the Swedish system quickly becomes interesting due to the situation of the English language in the nation’s educational system (Hult, 2012). As the current study is aimed at developing tools and methods for the upper-secondary courses, the expected language level is equivalent to CEFR B or IELTS 5 and up, meaning intermediate, independent users. However, as Hult describes in his overview of the policy documents and syllabi, English is undergoing a transculturation into the Swedish society, and is currently peculiarly placed between a global and local language. In addition, the move from English as a Foreign Language (EFL), i.e., English transplanted from a native-speaking country, towards EAL for communicative purposes globally means that the current project must adjust accordingly and focus on what is communicatively viable internationally as well as locally, rather than what is formally correct in any given variety. This also separates the Swedish context from that of the previous research, and places Swedish EAL education close to EFL education in some of the native speaking contexts described previously in terms of policy and syllabi, but also close to the EFL education in other previous
studies depending on the level of each individual in the classroom. This indicates that, in order for a tool to be successfully developed, it must be broad enough in its application to encompass both situations.

It must also be broad enough to offer usability regardless of the theoretical starting point of the teacher. The implementation suggested in the current project will rely on social constructivist theory and situate the tool as an aid in creating social interactions at the appropriate level for both group and individuals. However, applications from other theoretical perspectives will also be considered during the design, and supplementing implementation plans in accordance with other theories post-facto should both be possible and encouraged.

The insights into what makes a tool engaging to employ gained from HCI research provide some starting points for design and development. Furthermore, the use of dynamically generated local learner corpora is likely to be found more appealing to teachers than the use of reference corpora, according to Mukherjee (2006). In order for a tool to be perceived as useful it must align itself with the needs and beliefs of the teachers, and it must be positioned not as a radical shift in practice but as an aid functional within already existing routines. This is accomplished both through the UCD process and the evaluation of the tool in accordance with the SAMR model in order to highlight functions that would move beyond augmentation.

Drawing from previous research on CF and SLA, the design of the tool must account for best practices as proposed by Ferris, suitability for finding the relevant language issues and output that can be used in support of planning as dictated by SLA theories. Relevant metrics, then, are fit with expressed routines, the efficiency of relevant tagging, experienced ease of use, and experiences of the tool as support in planning on both a group and individual basis. In addition, as highlighted by previous CALL efforts regarding automated language diagnostics, the ability of the tool to identify and present relevant concordances and data based on authentic student texts is of great importance.
4. Interviewing EAL Teachers Regarding Diagnostic Needs

This chapter presents an interview study on in-service upper-secondary EAL teachers’ diagnostic needs, feedback routines and student text interactions. The interviews are intended to indicate the fit of international research in the Swedish context. The interviews will also form the starting point for the development of the tool. The study consists of 3 sets of interviews, with the interview instrument being iteratively formulated through the 2 pilot sets.

The interview materials are then processed according to Bengtsson’s (2016) description of content analysis. The meaning units are presented in a narrative fashion, with analytical context provided for the quotes. The substantive codes resulting from the meaning units are then combined into theoretical codes, i.e., condensed results for the interviews on a holistic level.

The results are expressed through network graphs, which link the interview content topics with each respondent. The main rationale behind their use is to visualize recurring topics amongst the interviewees in a clear way. Topics connected to several respondents can be considered more general, while topics with fewer connections are considered individual.

4.1 Interview Analysis

The interviews were carried out in three intervals. The respondents were named T1 or T2 depending on whether they were part of the first or second set (Test 1 and Test 2) followed by a number corresponding to the chronological order of the interviews, e.g., T102 for the second interview of the first set and T203 for the third interview of the second set. The respondents in the third set were instead given fictional names, as the analysis is more in-depth and the focus shifts from method development to content analysis. The purpose of the initial pilot was the evaluation of the questions, the guide, and the design in general. The interview design and content were then revisited and combined with the respondent drawing before a second set of pilot interviews were carried out.

The second set of interviews served the purpose of testing any changes made to the interview guide, while also providing feedback regarding the respondent drawing. As in the first pilot, respondents that did not fit the inclusion criteria were picked so as to not exhaust the limited number of respondents available. After the second set, the interview guide needed no further revision, and the respondent drawing was shown to benefit the analysis. The results of the two initial rounds of interviews were compiled into a control dataset with which the interviews conducted with respondents fitting the inclusion criteria could be compared.
One of the interviews in the second set were carried out online using Zoom for a video call in combination with a shared browser whiteboard (AWWapp) for the respondent drawing. The use of Zoom and a shared browser whiteboard did not seem to impact the efficiency of the interview instrument or the quality of the recording. While this test was initially done so that a secondary way of completing the interviews would be available if respondents did not have the time for a physical meeting, the outbreak of COVID-19 in early 2020 led to most of the interviews being carried out online. This is unfortunate, but was unavoidable given the guidelines on social distancing and several of the respondents having been quarantined in accordance with national guidelines.

The third set of interviews was carried out during 2020 and data was collected through both the interview and respondent drawing. Respondents were also asked to indicate interest in participating in the field tests planned for the evaluation of the diagnostic tool.

The meaning units are presented in a narrative fashion, with analytical context provided for the quotes. The substantive codes resulting from the meaning units are expressed through network graphs, which link the interview content topics with each respondent. The substantive codes making up the individual topics are then condensed into compilation topics that describe the results of the combined interviews. These combined topics fill the role of theoretical codes. These theoretical codes are considered relevant for the tool design, as they represent more widely found results in the interview study.

4.1.1 Analysis: Pilot Interviews

The first set of interviews focused on refining and trimming down the interview guide, while the second also employed the respondent drawing while testing the changes made to the interview guide. In September of 2018 three pilot interviews were conducted in Växjö, Sweden, in order to assess the design of the interview script and questions, and to evaluate the transcription and analysis process (Kvale, 2014, p. 221). The respondents were recruited by the author either through previous workplaces or studies.

The results of the interviews will be presented in the form described previously. This section will contain two intertwined types of analysis, one focusing on the interview results, where the method for the analysis described above is used (section 2.1.4), and one focusing on improving the interview design before moving on to the next set of pilot interviews, which will include the respondent drawing.

The three interviewees had different educational backgrounds and different amounts of experience, but all belong within roughly the same age bracket (20 – 35). They were all actively engaged in English language teaching at an upper-secondary school, but were working at different schools. The courses they were currently teaching are logged amongst the base variables. The base variables are presented in Table 5 on the next page.
The interviews took between 35 and 45 minutes each, including the introduction and debriefing. Some questions stood out as needing elaboration in all of the interviews, and some questions elicited very similar answers. The questions were revised for clarity, and questions resulting in repetitive answers will be removed. The background variables, however, seemed to fill their purpose nicely in terms of participant categorization.

All three participants pointed to L1 interference in lexis, specifically the use of anglicized Swedish words and phrases, as being the most common issue, followed by articles and subject-verb agreement errors. The individual issues were more different in nature and did not seem to stem from one specific category. T101 described individual issues as often being the result of a flaw in the student’s previous school environment, perhaps making oral production difficult due to a harsh classroom climate or direct intervention-type corrections. According to T101, this creates a student whose language practices are based around the idea of using language as little as possible, preferring to communicate in short bursts of familiar phrases.

(1) “Jag tycker en väldigt enkel sak är kanske mest psykologiskt egentligen men… det är ju att de inte provar… de har inte en vana att uttrycka sig och inte en trygghet i att uttrycka sig och osäkerheten leder till också mer brister och en ovilja till att ens uttrycka sig på engelska leder ju till att man faktiskt inte ens kan hjälpa dem att utvecklas inom det…” (T101)

My translation:

(1) “I think a simple thing is mainly psychological really, but, …they do not try. They are not used to expressing themselves and do not feel confident in expressing themselves, and that uncertainty leads to more issues. Their reluctance to express themselves in English makes it difficult to help them develop the language.” (T101)
While this is mainly found in oral production, it would also have an impact on the student’s written production due to hindering language development.

T102 pointed to spelling as the most common individual issue, and from their description the more specific answer seems to be that students tend to spell English words by sounding them out, rather than relying on formal knowledge. This leads to issues with homophones, such as “there”, “their”, “witch”, “which” and so on. Both T102 and T103 were of the opinion that most individual impediments were remnants from earlier courses where the students had not fully understood what was being taught.

(2) “Det är på gymnasiet orättvist hur olika undervisning eleverna har fått tidigare” (T102)

My translation:

(2) “In upper-secondary it is unfair how different the education the pupils have received previously has been” (T102)

Because of this, the errors in syntax could not really be summarized or generally described by the teacher beyond being the result of a lacking language background, often in terms of basic English grammar rules and terminology. It is worth noting that T103 was currently teaching the two more advanced English language courses available at the upper-secondary level, which might have had some influence on this perspective. However, the perspective was also shared with T102.

On the topic of feedback, all three respondents preferred individual feedback but only T102 had the time to sit down with each student in practice. All respondents used written feedback regularly, and communicated it via notes and comments in the student texts. The digital environment was said to play an important role here as the ”shared” editing available in, for instance, Google Docs became a central component of the process.

Possible substantive codes drawn from the interviews would be L1 interference in the lexis through sounding out words, missing steps from earlier courses, and minor grammar issues related to verb forms. Prior school experiences could also be linked to avoidance strategies in production. All respondents reported that marking and providing feedback on student texts influenced their lesson content when time allowed for it.

In Figure 6, both respondents and categories are shown as nodes. Respondents are highlighted in yellow. A respondent mentioning a category is shown by a line drawn from the respondent node to the category node. The nodes found in the outer circle of the graph are connected to one of the respondents, the nodes in the same circle as the respondents are connected to
several respondents and the nodes found in the middle are connected to a majority of the respondents.

Figure 6: Network of categories drawn from first set of pilot interviews

The interview guide seemed to be a viable method for collecting relevant data, although some changes are necessary before moving on to the next set of tests. The project has also moved along since the writing of the first interview guide. This enables the guide to be shortened as the survey is becoming more focused as it progresses, as was described in the section 4.1 regarding Grounded Theory and conceptual coding.

The second set of pilot interviews was carried out in early May of 2019 with a new group of participants to further evaluate the accuracy and reliability of the interview guide. There were three participants from different backgrounds than the first group, but all were previously known to the author. Two out of the three had received their teacher training in a country other than Sweden, and were not involved in teaching within the Swedish system at the time of the interviews. The decision to include these two participants was mainly motivated by trying to ensure that the participants available for the main set would all fulfill the inclusion criteria. As the pilot sets were intended to iterate over the interview instrument and to test the respondent drawing, prioritizing the respondents that did fulfill the inclusion criteria for the main set became
necessary as the number of interested respondents was limited. The base variables for the second set of interviews are seen in Table 6.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Experience in years</th>
<th>Eng. 5</th>
<th>Eng. 6</th>
<th>Eng. 7</th>
<th>Primary/Lower-secondary experience</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>T201</td>
<td>8</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>T202</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>T203</td>
<td>4</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

*Table 6: Base variables for second pilot participants*

The interviews yielded similar information to the first set of pilot interviews, with the addition of an added focus on the teachers’ experiences and beliefs regarding digital tools. The interview guide does not seem to have suffered from the removal of the questions yielding repetitive answers, nor did the drawing seem to have influenced the sessions in terms of dynamic discussion or fluidity. The drawings produced are presented below (Fig. 7).

*Figure 7: Respondent drawings from the second set of pilot interviews*
T202 and T203 produced similar drawings that represent the flow of information and text in their preferred interaction between student, teacher, and digital tool (Fig. 7 A & B). The flow can be described as focusing on the teacher as a mediator of information, as both teachers emphasized the importance of picking only the relevant language issues for feedback to the student during their interviews. T203 also included a digital interface, separate from the digital tool being discussed, as they preferred students to have ready access to the feedback via, for instance, Google Docs. Both T201 and T202 instead preferred individual oral feedback for every student but cited lack of time as a reason not to implement such a routine in practice.

T201, however, drew a more complex interactional flow (Fig. 7 C), where the teacher was removed from the student’s work to a larger extent, and the circles were said to represent the student workflow while the bars represent different errors or difficulties the student experiences. The teacher described only wanting to interfere once the student encountered such a difficulty, so as not to disturb the work. The teacher still wanted a mediating role in the context of information about the student texts though, as (s)he did not want the student to feel overwhelmed. This fear of access to a large amount of information about the texts being available to the student being overwhelming, or in other ways counter-productive, was also touched upon by T202 and T203 in the interviews.

The drawings were not used for taking notes by any of the three participants, although all were given explicit permission to do so, but the drawings became the centerpiece of the discussion of student-computer-teacher interaction during all sessions. Two out of the three participants also elaborated on the categories “Routines and Feedback” and “Digital Experiences and Tools” while engaging with the drawing after the interview questions had been asked. The drawing thus seems to fill the functions suggested by the previous research in this new context as well.

T201’s background contains no teaching in the Swedish system, but they have been active in EAL education for the relevant age groups (students aged 16 – 19) and have an equivalent certification to the Swedish teaching license. Recurring areas of feedback included articles and verb tenses, with punctuation errors being apparent in a minority of students. T201 had also experienced issues that were not found in informal language but appeared when students were entering into more formal forms of writing. The main reason for this was believed to be previous experiences with formal English usage in school, where students who had not become comfortable with the higher register were more prone to errors due to stress.

As with previous respondents, written comments in the student texts were the main form of feedback, but T201 also employed peer review when the hours available allowed for it to take place in the classroom. The material was mainly commented on after submission or when the student encountered difficulties, as previously described in relation to the respondent drawing. The influence of
corrections on the lesson planning depended on how strict top-down control was in the different contexts where T201 had been active. In some cases, teaching was to deal only with content from a textbook, in the order of appearance in the material.

The interview with T202 was carried out via Zoom and used a shared whiteboard for the respondent drawing. T202 has mainly been active with L1 English speakers, but has also taught advanced EAL in the same classroom due to the local system. Like T201, T202 pointed out attempts at formal language use as a large contributing factor to students not performing well in writing. This was linked to word choice, as students would tend to look for “fancier” synonyms without understanding fully what the new word means or how it is used. Being in an English-speaking environment helped the students acquire a functional understanding of the language, but the formal usage brought them out of their comfort zone. As a result, the students produced language errors that were reliant on rules that they did not have access to: “They know that it [the sentence] is wrong, but they don’t know why it is wrong” (T202). This could, according to T202, be linked to previous education, as the students had not been given explicit education on how to apply grammar to new words. Outside of the structural issues brought on by demands for formal language, T202 described issues with verb tenses as a common area of correction, together with run-on sentences and punctuation.

Feedback was mainly given through comments in the student texts, both during the writing and a summary after submission. While preferring peer-review sessions and oral individual feedback these were deemed too time-consuming. Feedback was also routinely given within a matrix or rubric describing the grading criteria and course content.

As can be seen in Figure 8, the addition of the second pilot set has split the respondent nodes into two clusters, each connected to its own sets of categories. As respondents from both sets are found in each cluster, this seems to be due to the clustering of beliefs rather than due to changes in the interview instrument.
The second set of respondents added meaningfully to the number of categories brought up during the interviews. This is important to notice as the number of categories added to the list as more interviews were conducted relates to the idea of attaining a “saturation of knowledge” when conducting analysis according to Grounded Theory. When no new categories are added to the network while new respondents are being interviewed, this criterion can be considered fulfilled. The number of categories per interview is shown in Figure 9, together with an example of how a “saturation of knowledge” would look once fully attained.
In summary, the drawings in combination with the refined interview guide seem to yield relevant data with some reliability and accuracy, and both instruments seem to benefit from each other. The data indicates that some categories return amongst most of the respondents regarding patterns in student texts (Avoidance Strategies, Spelling by Sound, Homophones, and Previous Teaching). Similarly, the content regarding feedback and routine indicates similar wishes among the respondents, with time being the main obstacle to achieving their desired feedback practices.
4.1.2 Analysis: Main Interview Set

The main set of interviews was carried out during the end of 2019 and the first half of 2020 with four respondents. They were recruited through my acquaintances and e-mails to upper-secondary schools. The names of the respondents have been changed. With the exception of Billy, the respondents were not previously known to the author. Due to the COVID-19 outbreak, three out of the four interviews were carried out online via Zoom and a shared, online, browser whiteboard (AWWapp) was used for the respondent drawing. The shared whiteboard also gave respondents access to typing text, which features in several of the drawings. This section will follow the analysis outline seen in section 2.4.5, but with a full focus on the content. The base variables for the main set are shown in Table 7.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Experience in years</th>
<th>Eng. 5</th>
<th>Eng. 6</th>
<th>Eng. 7</th>
<th>Primary/Lower-secondary experience</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim</td>
<td>4</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Billy</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Robin</td>
<td>2 + 6</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Mischa</td>
<td>4</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 7: Base variables for the main set of interview participants

4.1.2.1 Kim

Kim works at a school in a medium-sized city in southern Sweden and has been teaching for four years. They graduated and received their license 2 years ago, but worked as a supply-teacher throughout their studies. They teach Swedish and EAL at an upper-secondary level. Within EAL they are currently teaching in English 5, English 6, and English 7, and thus cover all of the standard EAL courses at an upper-secondary level in the current curriculum (GY11).

They describe their groups as having a generally high level of communication, but Kim often has to put considerable focus on textual details and language norms. There are instances of issues linked to the students’ L1 that appear when direct translations are made by the students, mainly word order issues and articles. This is especially common among students who are simultaneously acquiring both Swedish and English due to the languages being similar enough to become confusing in terms of structure.

According to Kim, many students have difficulties with punctuation, sentence structuring, and using capital letters, for instance at the beginning of sentences and when using the first-person singular pronoun, as a result of previous teaching:
(3) "Man har blivit lite förslappad, överlag, med hur man ska skriva. Språknormer, språkriktighet och sådär [... ] Stor bokstav, punkt, stor bokstav vid namn och sådant där. Man orkar nog inte säga till om det till slut, för det märker man jättemycket här. Det är sånt man struntar i för att man oftast skriver på nätet, och där behöver man inte det." (Kim)

My translation:

(3) "People have started slacking, in general, with how one should write. Language norms, rules and such [...] Capitalization, full stops, capital letters in names and that sort of thing. One probably gets tired of pointing it out after a while, because that is very noticeable here. It is the kind of things that one does not care about because one mainly writes online, and there those things are not needed." (Kim)

A contributing factor regarding these issues is also said to be the lack of proofreading and self-revision. Many students do not read feedback on grammar in their texts, but only look at the summary and grade.

Kim also links these issues to changes in the larger language norms in society, especially online, taking place rapidly between generations. This becomes very visible when students encounter the more rigid norms demanded by upper-secondary grading criteria and the formal use of English in academic writing.

(4) "Jag får ofta förklara att jag vill att de skriver hela meningar, med stor bokstav och punkt. Det beror nog på att de lever i en annan värld än lärarna vad gäller språknormer." (Kim)

My translation:

(4) "I often have to explain that I want them to write full sentences, with capital letters and full stops. This is probably due to them living in a different world than the teachers as far as language norms are concerned." (Kim)

There are also major differences in the groups in regard to previous knowledge of the language and confidence, according to Kim. Some students are at a high communicative level from the beginning while others barely fulfill the criteria for the basic courses in year 9. Amongst the weaker students, avoidance strategies are common and cause issues in conversation practice.
Kim adapts their lesson plans after what is seen in the student texts, but also keeps a close eye on how projects are progressing to adapt planning accordingly. Interventions are common in the beginning in order to provide students with a starting structure and motivation to come through with the assignment. Feedback is given through comments in the student texts through Google Docs after submission and a summarizing statement is also provided. In the respondent drawing below Kim has used numbers to annotate every part of the process.

![Figure 10: Kim’s respondent drawing](image)

1. Kim reads the submitted text and provides comments, both in-text and summarizing.
2. Kim creates a presentation that brings up recurring issues combined with examples from student texts handed out individually.
3. Students revise according to presentation and individual examples.
4. Students re-submit their texts with revisions.
5. Final version is submitted to the digital platform.

According to Kim the feedback needs to be focused on one or two rules at a time so as not to overwhelm the student. It is important to create a feeling of motivation, not failure, in order for the feedback to be effective. This is especially important for bigger assignments where Kim follows the idea of 2 wishes and a star; providing two corrective areas and one area of praise.

The general, recurring issues need to be discussed before the individual examples and texts are handed out. If this is done in a different order, students
are likely to disregard the presentation on general issues in favor of their individual texts and examples. Due to this, the interaction labeled number 3 in the respondent drawing signifies three sub-processes taking place in a specific order.

Feedback must also be given in close temporal proximity to the assignment being submitted. Partially Kim does this for their own benefit as the reading of the student texts benefits from having the students’ writing process fresh in memory, but it is also for the benefit of the students. Being given feedback creates a feeling of progression, and allows for revision to take place while the students are still familiar with their writing.

If given more time, Kim would like to give more in-depth individual feedback through the use of longer, more detailed summarizing comments. Kim would also like to provide a larger number of individual examples for the students to work through. Kim believes that having more regular and longer discussions with individual students regarding their goals and aims within the subject would also be beneficial, as well as allowing for more adapted planning.

### 4.1.2.2 Billy

Billy works in a small city located in southern Sweden, and teaches EAL and Religion at an upper-secondary level. They have worked as a teacher for 1 year and are licensed for both the subjects they are currently teaching. Billy is currently teaching EAL in the courses English 5 and English 6.

A lot of time is spent on correcting spelling as providing rules for the issue that can be applied beyond the word commented on is difficult. Homophones are especially problematic, hinting at spelling by sound. Billy also spends a lot of time providing feedback on the use of apostrophes, both in the possessive and in verb contractions.

Grammar, as a topic, is considered difficult by Billy’s students and when providing feedback Billy takes special care not to overwhelm the students. Common issues that are experienced as fruitful to comment on are plural -s, subject-verb agreement, and the previously mentioned structures involving apostrophes.

According to Billy, this could be due to how English is taught at earlier stages in the students’ schooling. As the focus in the earlier years is mainly communicative, the concept of grammar can seem intimidating and hard to engage with.

(5) “Jag får en känsla av att grammatikundervisningen på högstadiet blivit åsidosatt. Man kan göra sig förstådd, men kan inte reglerna. Det man kan försvinner i det skrifliga. Man har aldrig fått det svartvita i engelska.” (Billy)

My translation:
 Billy finds the same issues in both written and oral production, with the addition of pronunciation issues in unprepared oral production such as informal conversations around semi-familiar topics. The biggest problem for Billy’s students is mentioned to be a lack of proofreading and self-revision, as many students can correct issues once their attention is brought to them.

In English 5, Billy is a part of the entire process and provides comments throughout the drafting stage through Google Docs. For students in English 6, however, Billy is more hands-off in order for students to be able to creatively explore the language on their own. Language issues are brought up during lessons where Billy strategically set up students who have shown understanding of the rule to help students who do not during lesson time.

The work on written assignments is mediated through digital platforms where the teacher and students can exchange comments and drafts, ideally creating a cycle of writing, feedback, and self-revision as seen in the process described in Figure 11 below.

Figure 11: Billy’s respondent drawing

As seen in Figure 11, Billy makes use of multiple resources to communicate feedback to their students. Written comments, as previously mentioned, are
provided in the text via Google Docs, while direct questions can be engaged with in person or via mail. Urkund is used to check for plagiarism and results are sent to the teacher.

4.1.2.3 Robin

Robin works at a school in a big city located in southern Sweden, and teaches EAL and Religion at an upper-secondary level. They have previous experiences teaching at an upper-secondary level and have also taught upper-secondary courses within adult education. Robin is licensed to teach both EAL and Religion, but worked as a supply teacher for 6 years before becoming licensed as a subject teacher 2 years ago.

Amongst Robin’s students, subject-verb agreement errors stand out as the most common topic of feedback on grammar, while other common topics are coherency and vocabulary use. The last issue, according to Robin, likely stems from students relying on digital synonym resources when writing without fully understanding the connotation and meaning of the synonym they have chosen.

Other issues related to grammar include the use of capital “I”, punctuation, and run-on sentences. Sentence structures are brought up as something Robin would like to have more time to focus on individually with each student as the levels differ a lot within Robin’s groups. The differing levels of language acquisition are mentioned as one of the challenges when planning.

(6) “Jag kan ju ha vissa elever som knappt klarar grundläggande engelska och vissa som i princip är på universitetsnivå, men det är ytterlig heter.” (Robin)

My translation:

(6) “I can have some pupils that can barely manage basic English and others that are basically at a university level, but those are extremes.” (Robin)

Robin notes that the informal learning is very visible in their students, as students who interact with English through film, music or games during their free time often gain access to more complex features faster. There is also a difference in the types of issues found in texts and in oral production, as many of Robin’s students can hear that something is wrong when speaking, but do not notice issues in the same way when writing.

Feedback is mainly given through in-text comments using Google Docs, unless the students contact the teacher and book an appointment for oral feedback. There are also lessons planned that deal with recurring issues in the group, as a way of providing feedback to everyone at once. For students who are experiencing difficulties with EAL extra time with a teacher can be made
mandatory at Robin’s school through the afternoon study hall available 1 day every week.

Robin would prefer giving individual oral feedback to every student, but cites lack of time as a reason not to. The assignment design is, however, put together in such a way that Robin has ample opportunity to provide comments throughout the students’ writing process. The writing assignments are also designed to force self-revision of student texts in accordance with comments as a way of acquiring new knowledge.

As can be seen in Figure 12 below, Robin draws on the access to the digital learning platform and shared document editor in order to interact cyclically with the students around their written work with an emphasis on formative interventions in multiple drafting steps before the assignments are submitted.

![Figure 12: Robin’s respondent drawing](image)

The terms “scaffolding” and “coaching” appear as a categorization in the drawing, and are found again in Robin’s description of how they would like to conduct feedback if given the opportunity. Optimally, Robin would like to sit down with every student for 15 minutes and read through their draft together while discussing the student’s personal experience of the assignment in terms of goals and content. This would allow for coaching regarding motivation and
mindset, while also providing scaffolding for each student’s individual language issues and structuring.

4.1.2.4 Mischa

Mischa works at a school located in southern Sweden, and is teaching in the adult education environment. Mischa teaches EAL and Swedish as a second language (SAS and SFI), being licensed in both subjects. Concerning EAL, Mischa teaches in all of the English courses outlined in GY11.

The context Mischa is active in is different from the upper-secondary context of the other respondents as Mischa is currently active only within adult education. The student group consists of adults who are either supplementing their studies from upper-secondary or have recently arrived in Sweden without documented grade equivalents for the Swedish system. Many students in the second group are also involved in Swedish as a second language education in parallel with their EAL studies.

Mischa describes a very varied teaching environment in terms of previous knowledge, learning styles, and language issues. For instance, Mischa can see that students from the Swedish system have acquired a greater fluency in spoken production, but are still weak in syntax and formality. However, some recurring issues stand out across the groupings, such as: tenses, numerus, sentence structures, run-on sentences, and subject-verb agreement.

Outside of syntax, vocabulary causes issues as students struggle to incorporate new words into their repertoire. Mischa brings up the sentence “I go dog tomorrow” as an example where new words have been acquired, but the communication is unclear due to syntax and missing words. This also serves to highlight the difference in communicative level in Mischa’s groups as the students exemplified by the quoted sentence are put into the same group as students who have achieved a much higher level and can produce language with lesser and/or fewer communicative issues.

Students incorporating written material from different sources into their own production without references is a common issue, and leads to students reciting texts from memory during prepared oral production without fully understanding the content. Due to this, plagiarism is a common problem among Mischa’s students. The form of feedback follows along the line of previous respondents, where the teacher adds comments in a shared document and students are then encouraged to make self-revisions in accordance with the comments.

When narrating the respondent drawing (Fig. 13), Mischa wishes to create a cyclical process of drafting, feedback, and self-revision. During the interview, Mischa produced a drawing based on the template with several iterations between the teacher, computer, and student signifying this process. However, they later opted to erase the drawing and template in order to instead draw a new piece emphasizing the cyclical structure as the most important aspect of the exchange between student and teacher.
However, some issues regarding applying the cyclical approach in reality have become apparent to Mischa during the last semester, mainly due to students not revising their texts. As a way of encouraging revision, Mischa’s school will transfer to a seminar-driven process during the next semester. This would also include peer review to a larger extent than the current process.

Mischa would prefer having more 1-to-1 interactions with their students, where their texts could be discussed more in-depth. Mischa discusses a previous experience where circumstances had placed them with only 1 student during a lesson, and where they experienced themselves having a much more direct and fruitful impact on the student’s production. However, the time involved in such interactions makes them hard to accomplish on a regular basis.

### 4.1.2.5 Main Set Summary

Summarizing the main set of interviews provides a similar picture to the one provided by the pilot sets, with the addition of an explicitly mentioned intention to create a cyclical feedback process involving self-revision through the use of shared documents. However, such a process could be seen in the previous interviews as well, once the concept had entered the theory formulation.
As can be seen in the network graph for the main set (Fig. 14), most of the topics are recurring for several respondents and could be roughly divided into classifications of student issues and teaching, or feedback, methods. Furthermore, nodes such as “Individual Oral Feedback” might not be applied for some of the respondents, but it is a part of their description of desired feedback procedure. Similarly, some respondents are likely to have included “Verb Tenses” in the concept of “Sentence Structures”.

![Network of substantive codes drawn from the main set of interviews](image)

As with the pilot interviews, nodes were only added for topics explicitly mentioned by respondents with a further categorization of the combined sets by the author taking place in section 4.1.3.

The feedback process desired by the respondents includes more time for each individual student and oral interactions around student texts to a larger extent than what is possible today. As was seen in the pilot sets, both the desired feedback process and text interactions are hindered by the limited time each respondent can spend on each student. Time constraints were also brought up in relation to the reading of student texts when respondents were explaining their routines throughout.
4.1.3 Analysis of Combined Interview Sets

The pilot sets provided 15 topics while the main set provided 19. This could partially be due to the continued alterations made to the interview guide since the first pilot set, but one cannot rule out the influence of the move to online communication due to COVID-19 as the face-to-face interactions were highlighted as important in the background to the interview method. The topics produced by the interviews can be seen in Figure 15, where they are linked to their set of origin. Topics found in both sets are placed in the center.

![Diagram of interview topics](image.png)

*Figure 15: Substantive categories for all interviews linked by set*

As can be seen in Figure 15, the majority of the topics were recurring after the pilot sets with 12 out of the 19 found in the main set having already been logged. Building a theory based on these categories, teachers seem to be favoring 1-on-1 interactions around student texts, while emphasizing the importance of the teacher as a mediator for feedback on grammar. Two wishes and a star was highlighted method of structuring written feedback, and it was used to avoid creating a feeling of failure in the student. Another important function of the teacher as a mediator is to make sure that the student is not overwhelmed, and that the feedback is actually fruitful for the student. In the context of this project, the concern regarding overwhelming the students indicates that the teacher is
the primary audience for language overviews so that a selection of appropriate feedback topics can be made.

Teachers provided feedback through shared documents, most often Google Docs, and made use of in-text comments, with a majority also providing a summarizing statement. Three out of four respondents in the main set reported that students rarely made use of the in-text comments and instead made revisions based on the summary. Furthermore, some students did not make revisions at all.

Previous teaching is also a recurring topic for the respondents where three out of the four main set teachers believe that some of their students have not been properly prepared for the upper-secondary courses when it comes to grammar. Similarly, issues with the use of formal English are also brought up as a result of previous education. However, positive experiences with previous teaching are also brought up in relation to students who feel motivated within the subject.

Moving on to specific issues with the language, punctuation, and sentence structures were pointed out as problematic areas by three out of four respondents. Kim connected this to L1 interference, simultaneous language acquisition, or lack of proofreading. While the reason for these issues appearing is not of central importance for the use of the interview data in the tool design, it is interesting to note that all of the respondents had put consideration into why these issues were appearing in their students’ written production.

Some of the final categories were compiled by several categories, as the categorization was mainly due to implied causes for the issues, which do not concern this project. One example would be Spelling, which initially could be said to have 3 categories:

![Figure 16: Example of a combined category in the dataset](image-url)
Categories that could be drawn from the interview data and transferred directly into the design of the tool were verb tenses, subject-verb agreement, run-on sentences, articles, punctuation, spelling, and sentence structures. Other categories will impact implementation, such as the different types of peer review brought up by respondents, the need for structured self-revision, and proofreading. The list of relevant categories after compilation is found below in Table 8, sorted by sphere of influence.

**Tool Development:**
- Punctuation
- Sentence Structures
- Articles
- Verb Tenses
- Register
- Spelling
- Subject-verb Agreement
- Apostrophes

**Application:**
- Avoidance Strategies
- Overwhelming
- Written Feedback
- Oral Feedback
- Self-revision

*Table 8: Categories from Interview Study*

The additional categories found in the main set when compared to the pilot sets appeared in the first two interviews of the main set. This indicates that some measure of saturation of knowledge could be achieved, as seen in the graph below where the non-combined categories have been plotted against the number of interviews (Fig. 17).
4.2 Conclusion of Interview Study

Examples of scaffolding taken from the interview study include work with drafts and resubmissions, as well as one-on-one discussions with learners regarding their text. While these discussions were not common in actual feedback, they were found in most of the descriptions of desired routines. Of course, scaffolding in a school setting can also take many other forms, but since the focus of the present project is student text production the examples remain in that realm.

Ferris, as described previously, puts emphasis on errors having to be rule-based in order for feedback on grammar in written production to be fruitful, as students will otherwise encounter difficulties when memorizing the new knowledge. This seems to be the opinion of the respondents as well, as the categories brought up as common to comment on are all rule-based phenomena, arguably with the exception of “Register”. There were also answers that relate more closely to content, such as cohesion and disposition, but these will not be included in the goals for the tool development as they are outside of the scope of the project.

The categories linked to the tool development agree with Ferris’ description of viable foci for corrective feedback to a large extent, and indicate that the ecological validity of previous international research on CF in SLA and EAL is high within the Swedish system, as experienced by the respondents. Similarly, the categories linked to the application of the tool fit well with Ferris’ suggested framework for beneficial CF, both in its cyclical nature and in the emphasis on individual feedback adapted for each student. Due to this, the current project will align its efforts with the research presented in the background section without modification except for adapting to the level of the Swedish upper-secondary courses.

As HCI research indicated perceived usefulness and ease of use as the main factors influencing attitude towards a digital tool, and thus the main influence on the intent of using such a tool, the routines regarding feedback and text interactions presented by the respondents are of extreme importance. Following the UCD design patterns described previously, the interview data provided this study with an entry point to the tool development in the following chapter.
5. Tool Development

Following the understanding of the text interaction and feedback routines gained from the interview study, this chapter deals with the design of a tool that fits within the routines described by the interview participants and is able to carry out the relevant diagnostic analysis. The chapter describes the process chronologically by first introducing the resources that were used to build the tool, while describing the rationale behind their selection. The resources used to perform the language analysis and complexity scoring are presented separately.

The section following this introduction describes the design process based on the frameworks introduced in section 4.5 on Teacher-Computer Interaction. Interfaces are presented as wireframes, and the functional overview is presented via a flowchart. Finally, an overview of the intended workflow for the tool is accounted for in the context of a scaffolding approach to SLA.

The chapter is closed by initial testing in a controlled environment created by compiling an EAL learner corpus from an online repository of Swedish student texts. In addition, a small-scale user-interaction and engagement experiment is conducted in order to evaluate the instructions and usability of the tool before the dissertation moves on to field studies.

5.1 Resources

The programming language chosen for the current project is Python 3, as it allows access to the resources needed for the handling of documents, the language analysis, and the visualization of the results. Furthermore, Python will also allow for the interface to be created within the same script as the analysis, visualization, and parsing processes. The versatility and ease of implementing external resources are the main strengths of Python, along with its relative ease of readability when compared to most other programming languages (Sarkar, 2019). Python itself will not supply the main functionalities needed for the project, and so the external resources used will be shortly introduced before moving on to the actual functions in a practical context.

The external resources mentioned in the previous paragraph are what are known as Python libraries. A library is a set of functions written by a third party that is imported into a Python script to extend its possible uses. A typical Python program will make use of multiple libraries in order to carry out its intended purpose. In many cases libraries are tied to a specific area, for instance image manipulation, work involving a certain file type or, as relevant for the current project, text analysis.
5.1.1 Language Analysis

The core functionality of the tool relies on the Java LanguageTool (JLT), which is an open-source text analysis tool written in Java based on the work of Daniel Naber, presented in his 2003 diploma thesis. The current version of JLT supports 31 languages, although the amount of support and available pattern descriptions differ. JLT is mainly aimed toward implementations through websites and word processors, but implementations through programming languages in custom software are also available.

In addition to JLT, the tool also makes use of the Natural Language Toolkit (NLTK). NLTK was conceived in 2001 at the University of Pennsylvania and has since then been continuously expanded by multiple contributors (Bird, Klein and Loper, 2009). The functions offered cover many of the commonly used methods within corpus linguistics, such as tokenization, stemming and part-of-speech (POS) tagging, but within the tool developed for the current project NLTK mainly plays a supporting role. The main analysis of texts is carried out by JLT, while NLTK supplies quantitative data on the content, for example average sentence lengths.

The tool also includes functionality for complexity and readability analysis. The benefits of including an analysis of complexity are that the measurements correlate to the criteria in the Swedish description of the EAL subject, and as such are of interest to teachers, but also because syntactic complexity is a predictor for a formal register (Larsson and Kaatari, 2019), which was explicitly requested in the interview study. There is, of course, a danger in introducing quantitative scores into the tool as it could be seen as giving users an incentive to use the tool as a foundation for grading, which is not the intended purpose as it would require a different set of functions and measurements. Including complexity and readability measurements also puts a higher emphasis on the teachers’ understanding of the concepts involved in such scoring, as goals such as academic or formal language use are not represented in full by measurements of sentence level complexity or number of syllables per word. Such measurements also ignore features such as the phrasal complexity found in scientific writing (Biber and Gray, 2019).

The current project will make use of the TextStat library for Python, which includes several different readability and complexity scores. In addition to the scores, TextStat includes tokenizers for syllable, word, and sentence counts, which will also be used. The readability and complexity scores from 8 different readability formulas are combined into a consensus score.

14 https://dev.languagetool.org/languages for a complete list of languages and resources for each language (19/3 -19).
15 Not as a way of utilizing the tool for grading, but as a way of gathering data relevant for what the students will later be assessed on.
16 It is important to note that Larsson & Kaatari are mainly discussing syntactic complexity, but do reference the global complexity measurements used within this project.
17 https://pypi.org/project/TextStat/ (23/3 -19)
5.1.1.1 Java LanguageTool

As previously mentioned, Java LanguageTool (JLT) is based on Naber (2003) and is a rule-based text analysis tool. The aim of JLT during development was to build on previous spellchecking software in order to allow for analysis of spelling, grammar, and style. While JLT is intended for use in Java, the current project makes use of the Python library implementation by Jack Morris.\(^{18}\)

Rule-based approaches to text analysis rely on a library of errors that are defined manually based on patterns of Part-of-Speech-tags (POS-tags) and phrase chunks. An alternative to rule-based analysis would be the use of probabilistic approaches to predict error tags. Probabilistic approaches calculate the likelihood of an item being tagged as one of the possible alternatives and assign the most likely tag, while rule-based approaches look for matches in the data based on the manually defined rules.

Spelling errors are found by JLT by comparing the individual words with a reference dictionary, and highlighting the words that do match. In addition to this, JLT also provides words that are close to matching as a way of suggesting possible corrections. Due to being based on a dictionary, the analysis can only identify words that are included in the dictionary and is unable to identify, for instance, names that are not included in the list of proper nouns. The dictionary approach also creates issues with false negatives, as spelling errors resulting in a different word will not be indicated, even if the context makes it apparent that there has been an error.

Grammar errors are defined as structural or non-structural errors in JLT. Structural errors refer to non-grammatical sentences that can be made grammatical by inserting, deleting or moving one or several words. Non-structural grammar errors refer to non-grammatical sentences that can be corrected by replacing a word with a different one. As grammar errors are detected through pattern matching, the number of available patterns becomes central to the efficiency of the tool. The collection of error patterns used by JLT is referred to as its “rule set”.

While the manually defined rule set demands a higher initial investment in terms of time and effort, Naber (2003) evaluates it as more fitting for the language tool as the rules are easier to understand, define and extend later in the development cycle. There are still probabilistic elements to the JLT, such as the POS-tagger and phrase chunker, but the grammar analysis remains rule-based. In addition to the extension of the rule set, Naber (2003) also suggests that rule-based approaches would fit better for broad implementation as they are not dependent on the nature of the training data used to formulate probabilistic models. This results in JLT being a prescriptive tool for language analysis, as the rule-set behind the error tagger is dependent on prescriptive grammar.

\(^{18}\) https://pypi.org/project/language-tool-python/ for an overview of Morris’ implementation of JLT into a Python library (20/3 -19).
patterns rather than being trained on authentic language, i.e., a descriptive approach.

The rule set used in JLT is a community effort, and is continuously updated as the users identify more errors, false negatives, and false positives. As Naber suggested in 2003, the decision to make use of manually created rules following a standard formulation has made extensions of the rule set both possible and approachable. As of writing, the JLT rule set has been extended to 5,569 patterns and exceptions for the English language\textsuperscript{19}, and rule sets have also been created for a number of other languages.

Under normal conditions, JLT would communicate with a server hosting an updated collection of rules for the analysis via an Application Programming Interface (API). Using a brief analogy, an API fulfills the same task as a waiter in a restaurant, where you order a dish and the waiter (API) brings it to the kitchen (system server) which prepares what you have requested, and then the waiter (API) brings the finished order back to you. However, given the nature of the texts of interest for this project, sending the information back and forth outside of the local PC is not preferable. Furthermore, relying on the official API would also mean that the tool would need an internet connection in order to work. Due to this the tool in the current project will instead make use of a local server, i.e., a static collection of patterns on the user’s computer. The use of a local server to supply the rule set for the tool means that users will not benefit from the community development of the rule set unless they manually update their installation. The rule set used for the tool is downloaded during the first installation, and updating the rule set could be done by removing the tool and re-installing it.

\subsection*{5.1.1.2 Text Complexity and Readability Scores}

The Flesch Reading Ease Formula was initially formulated by Rudolph Flesch (1948) as a way of determining the comprehension difficulty of reading a text. The formula was initially developed to measure the difficulty of newspaper articles. The idea was later developed to measure the difficulty of newspaper articles. The Flesch-Kincaid Grade Level Formula (Kincaid et al., 1975), which did away with the tables needed to interpret the Flesch Reading Ease Formula in favor of a result that related directly to the U.S. grade levels. Both formulas are based on sentence length in words and word lengths in syllables.

The Gunning Fog Index also outputs a result linked to the expected reading level of a U.S. grade level. It was developed by Robert Gunning (1952) as a way of assessing his own writing and is focused on clarity and simplicity. Much like the Flesch Reading Ease Formula and the Flesch-Kincaid Grade Level Formula, the Gunning Fog Index works based on sentence lengths. However, instead of relying on an average syllable count per word, the Gunning Fog Index

\textsuperscript{19} The rule-set can be browsed at https://community.languagetool.org/rule/list?lang=en&offset=0&max=10 (12/9 -22)
divides the wordlist into complex words with 3 or more syllables and non-complex words.

In 1969 the Flesch-Kincaid Grade Level Formula was further developed into the Simple Measure of Gobbledygook (SMOG) by Harry McLaughlin. Like the Gunning Fog, the SMOG calculates readability through assessing the number of polysyllable words, i.e., words with 3 or more syllables. However, the SMOG needs a minimum sample of 30 sentences in order to produce a valid result. The SMOG measurement samples 10 sentences from the beginning of the text, 10 from the middle and 10 from the end.

The Automated Readability Index (ARI) was developed for technical writing within some branches of the U.S. military, as manuals were found to be beyond the reading level of the staff (Kincaid et al., 1975). Unlike the previous readability scores, ARI calculates word complexity based on character count rather than syllable count or categorization. A stand-out feature of ARI, during the 70s, was that the index was built for the purpose of automated use rather than calculation by hand. The Coleman-Liau Index (Coleman and Liau, 1975) was created with a similar purpose in mind, but more focused on use with computers to analyze physical materials. Both the Coleman-Liau and the ARI generate a result in the form of a U.S. grade level.

The Lensear Write Formula (O’Hayre, 1966) calculates a score based on sentence count and the number of monosyllabic words, and was created to improve readability in government communications. The score is intended to influence writers towards plain, speech-like language that is easy to read. The score is given based on readability to the average adult U.S. reader. The results take a number of stopwords into account and are intended to account for syntactic complexity to some extent.

Finally, the Dale-Chall Readability Score assesses texts based on a list of common English words combined with words per sentence. The original Dale-Chall Readability wordlist (Dale and Chall, 1948) contained 763 words, but the TextStat library makes use of the extended 3000-item wordlist from the new Dale-Chall Readability Formula (Chall and Dale, 1995).

In summary, the readability formulas focus on different audiences, features, and text types. As previously mentioned, TextStat runs all the formulas and then produces a consensus score which provides the likely level of complexity for the text. This should balance the foci of the different formulas enough for a valid measurement. The consensus score produced by TextStat is, as for many of the individual scores, expressed as a U.S. grade level. This means that the score must be transposed from a reader or writer in an inner-circle native speaking context to the L2-like context of English in Sweden to fit within the current project.


5.1.2 User Interface and Information Handling

The most important library for the tool’s data management is Pandas (Reback et al., 2021), which is a spreadsheet manager for Python. The Pandas version of a spreadsheet is referred to as a dataframe, and is a spreadsheet stored within the program as it runs. The dataframe can be easily manipulated, sorted, and curated depending on the needs of the tool. Once the desired data has been located, it can be visualized into a digestible overview for the teacher.

The visualizations will be done through the graphing library Plotly (Plotly Inc, 2015), which supports both 2D and 3D graph types. It enables the tool to provide fitting visuals for the different diagnostic measurements, while also offering some customizability according to teacher preference. However, while the visuals offered by Plotly provide excellent materials for overviews, they do not enable the teacher to inspect the individual elements being presented as the images are static representations. This is solved by making the graphs interactive through a browser dashboard.

The dashboard is created by building a Dash browser application using the Dash library. Much like JLT, Dash can be run locally to ensure that the students’ data is not at risk. Showing the visualizations through the Dash dashboard will allow the teacher to interact directly with the data through Dash’s interface and carry out any desired sorting of the materials. It will also enable the teacher to interact with the visualizations to gain a deeper insight into areas that appear interesting in the overview.

In order to load texts into the tool and to control the language analysis, an interface was to fulfill those roles. This interface in many ways fulfills the role of a launch client. The interface was constructed using the TKinter library, which is the standard interface library included with the Python base installation (Lundh, 1999).

5.2 Tool Design

As mentioned in the section on TCI (3.7), the design takes the form of two separate tracks of development: Function and Interface. The description provided in the Interface portion of this chapter will describe the functions of the tool in the context of the intended use-case for the field studies, i.e., scaffolding from a Social Constructivist perspective (Wood, Bruner and Ross, 1976).

The functions of the tool are described using a flowchart, providing an overview of the data processing, analysis, management, and presentation. As described in the section on resources (5.1) this includes language analysis and language complexity scoring, but also practical aspects such as parsing documents and managing information in spreadsheet-like data structures.
The interface design process makes use of wireframes in the context of the flowchart. The tool will make use of two interfaces, a launch client for loading the documents into the tool and a dashboard for interacting with the results of the analysis. The launch client is as basic as possible and exists only to provide the teacher with a transparent way of handling the texts. The dashboard, on the other hand, is the main interactive feature of the tool. As such, the dashboard needs to be evaluated against the categories provided by the interview study.

Finally, an instructional document for the use of the tool must be authored. As the instructional document will play a central role in both testing the tool in controlled environments and in the field studies it will be presented as an essential part of the design. The document will be made available to the participants in printed form, and digitally for on-screen reference during use.

5.2.1 Tool Function Design

The tool comprises four main functions: Document parsing, Language analysis, Complexity analysis, and Visualization (Fig. 18). Simply put, student texts go in one end and data visualizations come out the other. In this section the functions are presented individually, and at the end a complete flowchart for the tool in its entirety is presented.

![Flowchart for the tool's main functions](image)

The starting point for the tool is the input format. As seen in the interview study, Google Docs is the most commonly used text editor among the respondents’ students. In addition, the .docx file format used by Google Docs is easily created using other text editors, is the default format for MS Word, and can be utilized regardless of the operating system. The current study will assume the .docx format for student texts, and the first step of the design process will be creating the functionality for parsing text from the format (Fig. 19 on next page).
The .docx format is essentially a .zip archive containing the different modalities and formatting for the file. The method used in the first function works by unzipping the archive and extracting the text element. The text is then joined back together to create a purely textual version that can be passed on to the analytical function. At this stage the text is also cleaned from in-text citations, as the JLT analysis flags these as incorrect phrase structures. The texts are then passed on to two separate functions.

The second function passes the texts on to the local server running JLT on the user’s computer. JLT runs the American English (AmE) set of spelling and
syntactic rules on the texts, and outputs a .JSON\textsuperscript{21} structure containing the pattern codes and the sentence in which the pattern occurs. The reason for AmE being chosen as the standard setting is that TextStat is optimized for AmE lexis. The pattern codes indicated as relevant in the interview study (Fig. 20) are identified and their information is extracted to create a spreadsheet file in the tool’s directory.

![Diagram of JLT categories](image)

*Figure 20: Interview results (top row) and corresponding categories in JLT output (bottom row)*

As can be seen in Figure 20 above, focal points from the interviews do not relate directly to the categories in the JLT analysis. Patterns that are relevant for the focal points can, in most cases, be found in several of the JLT categories depending on the context they appear in. While the main overview of the tool is based on these broad JLT categories, their different sub-categorizations will also be made visible to the teacher and offer a more detail-oriented type of overview. This is likely to be especially useful when interacting with individual students over time, as the sub-categorizations allow for a more in-depth view of progression. An important thing to note here is that categories mentioned during the interviews, such as L1 interference, are not identified as separate from, for instance, spelling errors by JLT. Since JLT can only access what is on the page it is unable to differentiate between patterns based on possible origin and can only provide an indication of the nature of the error tied to the likely intended word. More complex categories thus become swallowed by the more rule-focused JLT categories.

The third function (Fig. 21) runs the TextStat consensus readability/complexity analysis on the texts, and provides the word count and sentence count. The consensus score and the counts are then passed on to a spreadsheet file (.CSV\textsuperscript{22}) in the tool’s directory. The reason for these two functions outputting an external file is so that the data can be saved externally.

\textsuperscript{21} JavaScript Object Notation, which is an information structure containing categories and their corresponding information for multiple objects.

\textsuperscript{22} Comma Separated Values, an information structure for tabular data. It is in many ways similar to an Excel file, but simplified.
by the user and then interacted with again after the initial analysis is done. The text analysis is quite time-consuming, and being able to revisit the data without going through the entire process is likely to be appreciated. However, these files are overwritten once the analysis is performed on a new set of student texts, so data must be downloaded through the Dash interface in order to be stored by the user between batches.

![Flowchart for Analysis](image)

*Figure 21: Flowchart for Analysis*
The fourth of the main functions, Visualization, feeds the results from the previous functions into the graphs generated by Plotly, and to the Dash dashboard (Fig. 22). The dashboard is labeled “Display” in the flowchart, and will be elaborated on in the section on interfaces (5.2.2). Once initialized, Dash relies on callbacks to the dataframe in order to alter the display based on user interaction. In terms of functionality, Dash also provides the functions needed for the user to be able to download the displayed data for future reference or alternative usage.

Figure 22: Flowchart for visualization
In summary, the combination of the four main functions provides a pipeline through which data regarding the relevant language patterns can be attained. As can be seen in the flowchart for the full tool on the next page (Fig. 23), the connections between the functions take place on different levels depending on user input and input demands for the sub-processes of the function following. The differentiated output before the merge into a single dataframe between functions two and three (Language Analysis and Complexity/Readability Scoring) is referred back to in the Dash dashboard to provide the teacher with the ability to see tabular data when performing the overview.
Figure 23: Flowchart for the entire tool
5.2.2 Tool Interface Design

The launch client interface that will be used to load documents into the tool includes the necessary functions to choose a folder of documents to analyze, load said documents into the script, and perform the analysis. Finally, it enables the user to generate the dashboard through the web browser. These functions are tied to a button each. In addition to the interactive components, the launch client will provide information to the user regarding which documents have been chosen and how the analysis is progressing. The design can be seen in the wireframe below (Fig. 24). Please note that the labeling refers to the function, and is not the actual labeling of the buttons.

![Wireframe design for launch client](image)

Using the tool in the standard way would mean pressing the buttons in the order indicated above, which to the author seems like the intuitive way. While the two text outputs (“Chosen files” and “Files that have been analyzed”) will provide similar information, the accumulative addition of texts to the second text output will show the user that the tool is working, and will be used to post a message indicating that the process is completed. Once this message is posted, the user...
can click the last button (“Generate report”) to open the Dash dashboard and interact with the results.

Clicking the “Generate report” button will either start a new browser instance, or open a tab in an already existing instance, which takes the user to the Dash dashboard interface. As seen in the workflow, the dashboard interface contains two different displays, “Overview” and “Individual Overview”. The user can choose which display to view by selecting a tab in the dashboard. The layout of the dashboard and the overview tab can be seen in Figure 25 on the next page.
Figure 25: Wireframe design for dashboard and overview tab

The overview tab pictured above in Figure 25 is intended to provide the information necessary for assessing the texts on a group level and for use as support for lesson planning. Due to this, the visualizations and tables are focused on providing data regarding the texts as a group, while also visually highlighting how individual texts stand out as relevant for individual analysis. The two tables have export functions which makes it possible to save them for later use, such as learning trajectory mapping and showing progression.
The use case for the first tab is mainly scaffolding on a group level through lesson planning and broad assignment feedback. The first graph indicates JLT hits, i.e., language patterns tagged by JLT, per sentence weighted against the readability/complexity score. This graph is intended to enable the teacher to see texts operating at the further extents of the students’ ZPDs in terms of writing, and help in identifying both issues for a majority of the group and the next steps of progression for texts exhibiting fewer JLT frequencies.

The table provided below the graph in the dashboard shows the teacher some general statistics, such as sentence counts and lengths, word counts, and TextStat consensus score. In addition to this, the table provides JLT frequencies per sentence, which makes them comparable across texts of different lengths, and average sentence lengths. This table is very suitable for exporting for later use as it is more comparable in nature than the more detail-oriented data found in the later tables. This is followed by a grouped bar chart which shows the categories found in each of the texts. The intended use of this graph is to identify categories suitable for lesson plans and categories suitable for individual exercises (see table in the Individual Overview tab). Categories found at higher frequencies throughout a majority of texts should be considered fitting for lesson plans, while categories that spike in frequency in fewer texts would be suitable for individual exercises.

The sunburst graph provides additional information on the sub-categories for the results, and data can be sorted by clicking the graph. The sunburst graph is intended for a more in-depth look at categories found at high frequencies in the texts, while also providing an overview of the sub-categories found in the texts as a group. Below the sunburst graph, a table containing the absolute frequencies of all the categories in each of the texts can be found.

Moving on to the next tab of the dashboard, the individual overview (Fig. 26), the first graph appearing on the tab is a bar graph of the absolute frequencies of JLT hits in the individual texts. The intended use of this bar graph is to indicate the amount of work needed for each of the texts so that the teacher can plan accordingly. Above this graph, texts can be selected from a dropdown menu.
Once a text is selected, the graph is exchanged for a sunburst graph to provide a more detail-oriented overview. Below the graph, a table detailing the JLT patterns detected in the texts is found. This table is tied to the text selector, and is updated according to which text is chosen. Unlike the previous tables, the table on the individual overview tab has some columns selected for hiding via toggle. This is due to the larger amount of information available in the table,
and the density of the table making it more difficult to comprehend. The table is exportable to Excel via a button in the dashboard, as were the previous tables.

While serving as an access point to the details of the JLT hits for the teacher, this table is also intended to form the basis for generating individual grammar exercises. The teacher has full editing possibilities within the table, and can freely edit the text within the individual cells. Columns can also be deleted, sorted or edited. The table can then be exported to Excel for final touches.

5.2.3 Instruction Document

The instruction document for the tool is divided into two sections: Installation and Use. The version used for the interventions only ran on Windows. It is found in appendix C. While the tool does not require a traditional installation, it does require the installation of additional dependencies for JLT to run. JLT will also perform an additional installation when the tool runs for the first time, which is likely to be experienced as a crash if the user is not warned in advance.

Ideally, the instruction document should serve as a quick start guide, but it should also provide enough of an overview to function as a primer for actual usage. The approach for the instruction document in the current project was to use a minimalistic checklist for installation and a functional overview based on visual materials. This will be supplemented by a video that will explain (1) installation and setting up, (2) the group overview tab and (3) the individual overview tab. These recordings are hosted on YouTube, and accessible through the instruction document through QR, for users working from a printed version, and direct URL, for users working from a digital version.

The contents of the instruction document are limited to what is needed to use the tool, and will not go into the details of the functions or the metrics. While these are important for the user to understand, they would expand the document beyond what most will be willing to engage with. However, due to these limitations, the accessibility of the dissertation manuscript, i.e., the text you are reading right now, should be considered during testing and later use of the tool.

5.2.1 What’s in a Name?

Up until this point this dissertation has mainly discussed a “conceptual tool”, i.e., a concept not yet attached to a real-world counterpart. As that counterpart has now been created, albeit in an experimental form, a name could be attached. There are also practical reasons for providing a proper name, as it will ease discussions with the respondents in both the user testing and the later intervention studies. For these practical reasons, I have decided to name the version of the tool evaluated in this dissertation “FirstContact”, as it allows the users to infer the intended usage vis-à-vis their reading of students’ texts.

However, the main text of the dissertation will continue to discuss it as “the tool”. My reasoning for this is the placement of the tool within the design of the current project, and the perspective that I wish to provide on it. “FirstContact”
has not been designed as a tool for large-scale public usage, it has rather been
designed in order to test the conclusions drawn based on the previous research
and the information gathered from the in-service teachers in a practice-oriented
way. The placement of the tool is as an extension of the methodology, rather
than it being the conclusion of the dissertation. One could consider the tool the
experimental setting in which the earlier conclusions are elaborated on and
tested.

Furthermore, I believe that a perspective on the tool as a complete,
independent object could create a level of abstraction between the results of the
evaluation and the choices made regarding the resources presented in this
chapter. “FirstContact” is the sum of its parts, and in order for the evaluation to
be productive it must be discussed at the level of affordances and limitations of
each of those resources, rather than discussing the tool as a singular object.
While the term “FirstContact” will be used for user-facing communication, the
term “tool” will be kept in the dissertation text.

5.3 Initial Testing

Before bringing the tool into a real-life scenario, the efficiency of the tool and
instruction document must be tested within a controlled environment. The initial
testing was conducted in two stages, aimed at measuring different aspects of the
tool. These tests will also indicate areas that need improvement before the case
studies are conducted.

The first section attempts to mimic a real-life scenario by evaluating the tool
with samples taken from a learner corpus containing student texts from the
Swedish school context. The corpus is based on the Mimers Brunn repository,
and has been tagged for text genre.

The final section will evaluate the actual use of the tool and the efficiency of
the instructions. This will be done through a small user-interaction experiment
using standard questionnaires for evaluating usability (SUS scale) and user
engagement (UES scale). Both of the questionnaires are based on Likert scales,
in which participants rate their experiences through incremented measurements.

5.3.1 Evaluation in a Controlled Environment and Text
Genre Suitability

As previously mentioned, this section will evaluate the tool using an EAL
learner corpus compiled by a well-known repository of Swedish student texts
called Mimers Brunn23. The goals are identifying errors and issues in the tool’s
design and analysis, as well as evaluating which text genres the tool can be used
on efficiently. The Mimer corpus is introduced before the evaluations are
presented.

23 https://www.mimersbrunn.se/ (01/05 -21)
5.3.1.1 The Mimer EAL Learner Corpus

Mimers Brunn is an online study community which hosts a forum, a list of course providers, and a repository of student texts in a wide range of subjects uploaded by the community. The texts are uploaded by the students so that they can receive feedback, input, and support, as well as the texts themselves providing inspiration for other students writing on the same, or a similar, topic. While there is a repository for lower-secondary and secondary school level texts, this project will focus on the upper-secondary repository. While English is not the largest section in the repository, it contained 470 texts spanning nine text genres.

For the purposes of the current project, the English subject texts were collected using a web scraper built in Python based on the BeautifulSoup library, which is a resource used to collect text from webpages, amongst other things. These were then tagged for text genre and content\textsuperscript{24}. The distribution of text genres in the Mimer corpus is found in Figure 27.

![Content Distribution](image.png)

Figure 27: Content text genre distribution in Mimers Brunn’s English subject upper-secondary repository as of September 2020

As is seen above, the corpus contains an uneven distribution of text genres, with reports making up the largest amount of material, followed by essays, fiction, biography, and argumentative essays. Essays here are referring to the text type sometimes known as informative essays, which are distinct from argumentative essays. When selecting samples for the tool evaluation, this must be taken into account.

\textsuperscript{24} The tagging was performed with excellent help from Mai Nabawy, who was a student of the LNU MELL Master’s program at the time.
5.3.1.2 Tool Evaluation

The evaluation of the tool’s efficiency and precision will be performed by creating five samples of texts from the Mimer EAL learner corpus. Each sample will be processed in the tool in accordance with the previously described procedure. The results are inspected through the dashboard, and any issues and errors are accounted for in order to create usage guidelines for the tool.

In order to evaluate the tool against different texts from different genres, samples from the Mimer EAL learner corpus will not be entirely randomly selected. Using the random sampling function included in the Pandas library for Python, sampling took the text genre category into account. As can be seen in the generated example sample (Fig. 28), the content listings for the Mimer EAL learner corpus also contain notes from the categorization process. The evaluation will make use of three samples, each containing between 18 and 20 texts.

<table>
<thead>
<tr>
<th>Text</th>
<th>Type</th>
<th>Keywords</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MimerEngGym315.txt</td>
<td>Report</td>
<td>Myths, Imaginary Creatures</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym207.txt</td>
<td>Report</td>
<td>Celebrities, Fashion</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym392.txt</td>
<td>Biography</td>
<td>Marilyn Monroe</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym105.txt</td>
<td>Biography</td>
<td>Samuel Clemens, Mark Twain</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym430.txt</td>
<td>Argumentative Essay</td>
<td>Immigration, Sweden</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym104.txt</td>
<td>Essay</td>
<td>History, The Apartheid, South Africa</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym205.txt</td>
<td>Letter</td>
<td>Travel, Sydney</td>
<td>NaN</td>
</tr>
<tr>
<td>MimerEngGym107.txt</td>
<td>Fiction</td>
<td></td>
<td>NaN</td>
</tr>
</tbody>
</table>

Figure 28: Example of sample texts generated from Mimer EAL learner corpus

In sample 1, the initial graph of the Group Overview Tab indicates high frequencies of Typo pattern JLT hits in four texts (Fig. 29 on the next page). When inspecting the Individual Overview Tab many of the hits are due to JLT incorrectly flagging place names and proper nouns, which seems to cause the majority of false positives within the Typo category. The majority of Typo category hits are indicating as intended.
Figure 29: JLT Categories in sample 1

MimerEngGym107 (indicated above) stands out here due to containing the highest frequency of Typo hits. The text is about a fantasy setting in a video game franchise, and contains a large number of unusual character names and places (Fig. 30). This is likely to be an issue with other texts containing names and places not included in JLT’s word list.

Figure 30: “Misspelling” pattern hit in MimerEngGym107

In addition, MimerEngGym104 is shown to have a high frequency of Typography pattern hits. The reason for this is the template used for subheadings in the document containing double spaces, which triggers the “whitespace” error flag in JLT (Fig. 31).

Figure 31: “Whitespace” pattern hit in MimerEngGym107

In sample 2, MimerEngGym142 highlights the inherent issues that come with the tool being locked to AmE. While this is necessary to comply with the usage guidelines for the TextStat readability/complexity consensus scoring, it also flags the use of BrE spelling as spelling errors (Fig. 32 on the next page).
On the topic of the consensus score in sample 2, MimerEngGym318 stands out in the initial overview graph that visualizes JLT hits per sentence combined with the consensus score (Fig. 33). However, upon inspection the JLT hits appear correct and are not due to BrE spelling or named entity recognition issues. Instead, this seems to reflect a situation described in the interview study where an author is performing at a high level of lexical complexity, but lacks syntactical awareness, i.e., “the grammar” or “black-and-white of English” referred to in the interview study.

In sample 3, the JLT frequencies and consensus scoring cluster the texts a lot closer together and only MimerEngGym166 stands out with a high JLT frequency. The hits are mainly to do with syntax and stem from the unorthodox structure of the text, which is a piece of poetry. This is likely to be an issue with any piece of creative writing that does not conform to the strict prescriptive definitions that JLT operates based on.
5.3.2 Pilot from Practice Placement

As a part of my upper-secondary teacher training during autumn 2021, I was able to partake in a 10-week long practice placement (VFU\textsuperscript{25}). This placement took place during the autumn semester, and provided an opportunity for me to test the tool in practice. This first outing of the software was intended to aid in the design of the instructions and guidelines, as well as inform a second iteration of the design of the tool itself, in accordance with the iterative design framework.

During my placement, I had the opportunity to teach in the English 5 and English 6 courses, and perform one teaching design based on written production in each course. The output of the tool was not shared with the students, and the focus was on how the results were able to inform my teaching plans and text feedback. The results were also shared with my mentor at the school, as they guided me in the planning during the placement.

While the tool functioned as intended, the volume of patterns found made the results difficult to interpret through reading the visual representations and often required extensive engagement with the tables. Upon inspection it became clear that much of the “noise” came from categories that were not deemed relevant for the upper-secondary courses, while most of the relevant patterns came from the categories “TYPOS”, “GRAMMAR” and “PUNCTUATION”. However, patterns from other categories were more likely to become relevant as the courses progressed.

A second concern brought up during the placement was the lack of references and explanations within the tool itself. While the version of the instruction document created for the dissertation project’s midway seminar during 2021 contained explanations of the different functions and graphs within the dashboard, these would only be accessible through having the quick start document open while using the tool. These explanations were also rudimentary and required a reading of the manuscript in order to be transparent. Adding further explanations to the quick start document would make it longer and denser, in terms of information, which would likely impact the experience of setting the tool up negatively. Finally, the quick start document being a local file distributed alongside the tool itself means that it cannot be updated as more is understood about the tool and user experience throughout the evaluation.

5.3.2.1 Second Design Iteration

The second design iteration is quite minor in nature, and is mainly concerned with improving the user experience and answering the concerns brought up during my practice placement. As this was the final step before sending the tool

\textsuperscript{25} The Swedish upper-secondary subject teacher training program includes two 5-week in-service placements and one final 10-week in-service placement. These are credited under the term “Verksamhetsförlagd utbildningsdel”, or VFU, meaning “in-service part of education”.

116
out for testing by someone not involved in its creation, and who is intended to use it without active participation from the author, the focus was on improving transparency and information access for the new intended user group.

The changes in the graphical user interface (GUI) were tied in with desired functional changes. A checkbox was added that would open a browser tab with expanded instructions at the launch of the dashboard when checked. This is intended to mitigate the issues with navigating the instructions in a separate file during use, and allows more information to be accessible for the end user. In order to make the navigation of the document intuitive, screenshots of the actual dashboard were used to indicate where information regarding the different sections of the dashboard can be found in the document. The expanded instruction document is hosted via Google Docs so that alterations can be made continuously without necessitating a new download for the user.

In addition, pattern selection was implemented through a pop-up window activated via a button in the user interface. The patterns are selected via a checkbox, and the intention is for the function to allow the user to produce a less noisy and cluttered dashboard. The categories indicated as the most frequently interesting during the practice placement are selected by default, with additional categories being selectable by the user. This will also allow for changes in the focal areas of the analysis according to what the class is currently working on. The interface with the pop-up pattern selection window is shown in Figure 34.
As for the dashboard, changes were made to axis labels and visualization titles. Changes were made from titles and labels that indicated functionality in the script running the tool towards titles and labels that serve a more descriptive purpose for the user. The initial descriptors were chosen to simplify error location while building the tool, and the change towards practice-oriented readability marks a general change within the dissertation project, as the focus is now turning towards the teachers as users rather than the development of the tool as such.

5.3.3 Tool Distribution

Due to the circumstances surrounding the testing and evaluation phase of the dissertation project, with the ongoing COVID pandemic and distance teaching, the distribution model for the tool had to be robust and easy to use for the participants. The Google suite of tools was prevalent in the practices of the teachers who took part in the interview study, and therefore Google Drive stood out as a good choice for the distribution of the tool. Ideally, distributing the tool via Google Drive should provide the teachers involved in the testing and evaluation with a familiar way of accessing the materials, while also fulfilling the requirements of the types of materials that need to be distributed.

The primary material is a .zip archive containing the executable, the dependencies, and an instruction document. The Google Drive folder also contains two sets of samples from the Mimer EAL Learner corpus, which are intended to help the teacher in running the tool for the first time, as well as provide materials for user engagement and usability testing. In addition, the drive also contains a folder with older versions of the tool, should there be a need for a roll-back, as well as a patch log so that the teachers can orient themselves amongst updates if a new version is released during their evaluation periods.

During the winter of 2021, the distribution model was tested using five home computers, and was found to function well in all cases except for one. The case that was not able to produce a running copy of the tool on the PC was hindered due to the firewall blocking the installation of the Java Development Kit (JDK) needed for the JLT instance to run, and the user not having the correct privileges to make alterations to the firewall exceptions.

It is also important to note that downloading the .zip archive from the Google Drive will produce a warning message due to the file size. The size of the file makes it impossible for Google Drive to perform the virus scan normally done for all downloads, and it is important that respondents are notified about this as it might lead to some discomfort and worries.

Both of these results are likely to become more important for the actual field evaluations than they were during the testing of the distribution model, as the tool is likely to be used on work computers by the teachers. The system privileges employees have on their work machines vary, which will make
dealing with things like firewall exceptions difficult. Furthermore, the warnings from Google Drive are likely to cause more alarm when issued on a machine not privately owned, especially one likely to contain important materials from work. For participants who did not wish to use Google Drive, USB memory sticks were made available.

5.3.4 Usability and User-engagement Testing

In order to test the user interface and instructions, an iterative study was conducted using questionnaires. The questionnaires were composed from the Short-form Usability Scale (SUS) created by Brooke (1995) and the Short Form User Engagement Scale (revised SF UES) by O’Brien, Cairns and Hall (2018), which were chosen due to their prominent usage within user experience research and ease of use. The former is intended to provide an idea of experienced usability which is an important factor for actual use according to the HCI research, while the latter collects information regarding how engaged users felt while using the tool.

The current study will make use of both the SF UES and the SUS scales in order to explore the perceived usability of the tool, as well as the experienced engagement. The SF UES and the SUS will also be distributed during the end of the field tests, so that changes after prolonged use can be identified by comparisons with the shorter evaluation period of the initial tests.

The SUS is described as a “quick and dirty usability scale” by Brooke (1995), and was created during his work at Digital Equipment Co via Redhatch Consulting. Usability is defined by Brooke as “appropriateness to purpose” (Brooke, 1995, p. 1). This ties this testing back into the previous discussion on ecological validity, and indicates that the SUS will also make a meaningful contribution to the debriefing of the field test respondents.

When shaping the SUS scale questionnaire, Brooke started out with 50 statements, each tied to a Likert scale running from 1 to 5, i.e., “strongly disagree” to “strongly agree”. These were then applied through 50 corresponding questions during the testing of 2 software tools, one considered easy to use and one considered near impossible to use. The 10 questions that indicated the most distinguishing usability experiences were then selected for the actual SUS. The resulting statements were:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.

26 This chapter is built from excellent advice given to me by Nico Reski, who pursued his PhD at LNU at the same time as me, and took the time to discuss how to explore the UX side of the project.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Half of the statements were selected due to indicating strong agreement and the other half due to indicating strong disagreement. Once filled in, the SUS score is calculated with the contribution of items 1, 3, 5, 7, and 9 being the scale position minus 1, and the contribution for the remaining items being 5 minus the scale position. The sum of the contributions is then multiplied by 2.5 to get the overall SUS score, ranging from 0 to 100. According to the best practices described by usability.gov (U.S. Dept. of Health and Human Services, 2006), a score of 68 is to be considered average. The score spectrum can also be expressed through adjective ratings, as in the subsequent matrix adapted from UIUXTrend.com based on work by Bangor, Kortum & Miller (2009):

<table>
<thead>
<tr>
<th>SUS Score</th>
<th>Grade</th>
<th>Adjective Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 80.3</td>
<td>A</td>
<td>Excellent</td>
</tr>
<tr>
<td>68 – 80.3</td>
<td>B</td>
<td>Good</td>
</tr>
<tr>
<td>68</td>
<td>C</td>
<td>Okay</td>
</tr>
<tr>
<td>51 – 68</td>
<td>D</td>
<td>Poor</td>
</tr>
<tr>
<td>&lt; 51</td>
<td>F</td>
<td>Unsatisfactory</td>
</tr>
</tbody>
</table>

*Table 9: SUS score and adjective rating*

The UES is intended to measure user engagement, which differs from the usability measured by the SUS as it is focused on measuring the “quality of user experience characterized by the depth of an actor’s cognitive, temporal, affective and behavioral investment when interacting with a digital system” (O’Brien, Cairns and Hall, 2018, p. 29). Combining previous scales used for User Engagement metrics, such as Jacques (1996) and Webster and Ho (1997), O’Brien et. al. (O’Brien and McLean, 2009; O’Brien and Toms, 2010) initially created a questionnaire containing 31 items, spread across six factors of engagement:
1. FA: Focused attention, feeling absorbed in the interaction and losing track of time (7 items).
2. PU: Perceived usability, negative affect experienced as a result of the interaction and the degree of control and effort expended (8 items).
3. AE: Aesthetic appeal, the attractiveness and visual appeal of the interface (5 items).
4. EN: Endurability, the overall success of the interaction and users’ willingness to recommend an application to others or engage with it in future (5 items).
5. NO: Novelty, curiosity and interest in the interactive task (3 items).
6. FI: Felt involvement, the sense of being ‘drawn in’ and having fun (3 items).

Once introduced, the original UES saw a lot of use within HCI research, and in 2018 O’Brien, Cairns and Hall were able to draw on the applications of the questionnaire in order to refine it. This was done in accordance with two common trends seen in the research that had made use of the UES: (1) The length of the original (31 items) was deemed too long and led to studies often only using one sub-set (factor) in their analysis, and (2) the results of studies applying the full UES indicated that there were in reality four independent factors at play, rather than the six used for the design (O’Brien, Cairns and Hall, 2018, pp. 30–31). The statements for each of those four factors are shown below:

FA-S.1 I lost myself in this experience.
FA-S.2 The time I spent using Application X just slipped away.
FA-S.3 I was absorbed in this experience.
PU-S.1 I felt frustrated while using this Application X.
PU-S.2 I found this Application X confusing to use.
PU-S.3 Using this Application X was taxing.
AE-S.1 This Application X was attractive.
AE-S.2 This Application X was aesthetically appealing.
AE-S.3 This Application X appealed to my senses.
RW-S.1 Using Application X was worthwhile.
RW-S.2 My experience was rewarding.
RW-S.3 I felt interested in this experience.

The four factors chosen for the short-form (SF) UES were FA, PU, and AE, with the remaining 3 being combined into a 4th factor named Reward (RW). The short-form UES consists of 12 items, with 3 items from each factor. Like the SUS, the SF UES makes use of 5-point Likert scales. When compared to the results of the full UES, the derived statements in the SF UES were found to correlate on a factor level. Scoring the SF UES can be done on both a factor-
and holistic level, with the former being done by adding together the score within a factor and dividing by 3, while the latter is done by dividing the complete sum for the questionnaire and dividing by 12 (O’Brien, Cairns and Hall, 2018, p. 39).

5.3.4.1 User Testing Pilot

Sampling for the user testing pilot tests were done through convenience sampling. An invitation to partake in the study was sent out to the Corpus Methods in Practice (CMiP) course at the advanced level, to the upper-secondary English subject teachers’ program at LNU through their course in sociolinguistics, and to the respondents involved in the initial distribution tests. Since this final group of respondents contains people who have not had contact with corpus linguistics or the English subject after upper-secondary they will be considered as a non-linguistics user group for usability and engagement outside of English subject university students. The invitations received 8 answers (2 CMiP, 3 Subject teachers, 3 Non-linguistics), and 3 completed responses.

Distribution was done through a Google Drive folder linked in the quick start document, as well as through USB memory sticks. The instructions were hosted on Google Docs, and the default setting opened them when the respondent started the tool. Except for the addition of the SF UES and the SUS, the respondent experience was the same as the distribution model used for the field tests.

In order to fulfill the criteria for informed consent posed by the ethics Codex (Vetenskapsrådet, 2019), new participants were introduced to the project through a PowerPoint presentation during the recruitment stage. In addition, emphasis was put on the voluntary nature of participation and the rights of participants in both the presentation and the questionnaires. The information given regarding the project during the presentation was mainly focused on the form and purpose of the tool. This was done so that the data from the interview studies and the specifics of the SF UES and SUS scales would not influence the opinion and actions of the participants.

Two user tasks were distributed together with the evaluation copy of the tool. The formulation of the tasks is important, as their results must verify whether or not the participants were able to complete them, while also engaging them with the information produced by the tool. The tasks must also have explicit end states, so that participants have a firm idea of when they have completed a task. This is especially important as testing is carried out remotely, without direct contact with a test facilitator. Finally, the tasks have to be grounded in the reality of the intended use-case scenario, i.e., analysis of student produced texts in a school setting for the purposes of feedback or lesson planning, so that the data collected from the user experience corresponds to a scenario similar to anticipated reality.
The tasks formulated for the tests were:

- “Using the FirstContact tool, can you pick 3 areas that you would focus lesson plans on for sample 1”

- “Using the FirstContact tool, can you suggest 2 focus areas for 3 texts from sample 1 or 2”

The first task is intended to focus the users on the general overview tab, and have them perform analysis on the full sample as a group. In order to avoid leading tasks, explicit mention of the tabs or overview labels has been avoided. However, these are included in the on-screen instructions that are opened when the report is generated. The intention here is to see whether or not the current model of instruction distribution is functional.

The second task is intended to move the focus to the individual overview tab. Key elements of the task are the use of the document selector and the use of the overview graph for the individual texts. As both samples are suggested for the second, the change of sample between the tasks is left up to the user. This is due to the relatively long time taken up by the analysis on some hardware systems.

Google Forms were used to collect the results of the SF UES and SUS. Three separate forms were created, one for each set of respondents. All three forms contain the exact same content, but using multiple forms allows this study to separate the results of the three sets of respondents. Results were collected anonymously, as no data regarding the individual respondents were collected in the form.

The response rate for the pilot user tests was 37.5% (3 out of 8), which is an interesting result when seen in relation to the distribution method. When recruiting for the tests, the main point of discussion with individual respondents was how to manage access to the tool and the setup procedure. This was further indicated as an issue through later e-mail conversations with the respondents, with almost all correspondence being on the topic of setup issues and issues in accessing the files for the tool.

For the set of respondents recruited from the subject teacher program, with a response rate of 33% percent for the group, corresponding to a single respondent, the single SUS score indicates a lower-than-average experienced usability with a SUS score of 62.5, with 68 being considered average. This provides an adjective rating of “Poor”, with the main issue indicated as being the stand-alone software nature of the tool, when compared to other solutions within the same realm. This is indicated by the free-form answer quoted below from one of the respondents:

“I believe the tool is excellent and has potential, but I would prefer it to be a web-based program similar to Grammarly, as many
people will avoid downloading external programs to check for errors when everything can be done online in a few clicks.”

Moving on to the CMiP respondents, the group had a 50% response rate or 1 participant. The SUS score indicates higher perceived usability at 72.5, with the adjective rating being an above average “Good”. However, the free-form answer indicates a slightly different focus from the respondent, where the key interest is the ease of use rather than ease of deployment. The setup is explicitly mentioned by the respondent as a weakness, as can be seen in the quote below:

“I liked it! I tried to be scrupulous in my evaluation, and while I think there is still room for improvement in the setup phase and UX, I have found the overall experience very intuitive and gripping. As FirstContact is by nature an assistant software, I believe teachers would find it very helpful and come to grips with the technicalities in a short amount of time as there are not many steps to learn in order to get to the results.”

In terms of activities carried out within the tool, the quote above also indicates that the respondent finds it likely to be seen as useful due to the small number of actions needed in order to produce the results. It is important to note that the quote is from the CMiP course respondents, which means that the technicalities mentioned are things that resemble experiences the respondent is likely to have recently had in the course, and that this might influence the evaluation of the difficulties.

For the non-linguistics group, the response rate was 33%, corresponding to 1 participant. The completed response indicated a “Poor” usability rating with a SUS score of 52.5. As with previous respondent groups, correspondence with the participants indicated the main barrier of entry as being the distribution model and overall impression of the setup process. In addition, the participant expressed a lack of understanding regarding what was being measured, which made the overview more difficult to navigate. No free-form answer was provided by the respondent.

While the SUS scores provide a general idea of the user experience, the SF UES provides scoring on different factors. This allows for a more granular perspective on the respondent data. The factors of the SF UES, as well as the total score, are shown in Table 10 on the next page. A neutral SF UES score would be 3, while the highest score is 5.
Table 10: SF UES score per respondent group and factor

<table>
<thead>
<tr>
<th>Factor</th>
<th>Teachers</th>
<th>CMiP</th>
<th>Non-Linguistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused Attention</td>
<td>3</td>
<td>3.3</td>
<td>2.33</td>
</tr>
<tr>
<td>Perceived Usability</td>
<td>3,66</td>
<td>4,66</td>
<td>3,33</td>
</tr>
<tr>
<td>Aesthetic Appeal</td>
<td>3</td>
<td>3,66</td>
<td>3</td>
</tr>
<tr>
<td>Reward</td>
<td>3,33</td>
<td>4,66</td>
<td>3,66</td>
</tr>
<tr>
<td>SF UES Score:</td>
<td>3,25</td>
<td>4,08</td>
<td>3,08</td>
</tr>
</tbody>
</table>

As with the SUS score, the CMiP respondents scored the tool slightly higher than the subject teacher students throughout. Both respondent groups gave lower ratings for focused attention and aesthetic appeal, and both rated the tool higher for perceived usability and reward. The general score for the subject teacher and CMiP groups indicated a slightly more positive than neutral experience. As for the non-linguistics group, the overall score indicates a neutral experience. However, the per factor score indicates that this is due to a higher experienced reward, and a lower experienced focused attention.

Lessons to be taken from the pilot into the actual user testing are the issues with the distribution, and the importance of insights gained from the free-form answers in the questionnaire. In order to mitigate the issues with the distribution system, user tests will be done via TeamViewer using a virtual machine that has already been set up with the tool and samples. This approach allows for remote testing, while also navigating around the participant having to go through the setup process, which was too engaged for the short user tests. As the free-form answers gave insights that were helpful in the analysis of the SUS and SF UES questionnaires, the user tests expand the free-form aspect of the evaluation through the use of a think-aloud protocol being in place during the testing (Jonasson, 2007). This would also mean that the researcher joins the participant during the test, and is available for technical support should the tool run into any issues.

5.3.4.2 User Testing

Sampling for the user tests was done through a Facebook community for English language teachers in Sweden. Permission to recruit via the group was collected from the head administrator of the study circle which operates the community, and the invitation post is available in appendix D. As the asynchronous nature of this type of recruitment makes the use of presentations
difficult, the project site for the current dissertation hosted at the LNU website was used to inform about the project. The respondents were also asked for any questions or thoughts at the beginning and end of the session.

3 participants were recruited, and the form designed for the pilot was used unaltered. Due to the national reach of the community and the lessons taken from the pilot, sessions were conducted on Zoom with me present throughout in order to provide technical support. As previously mentioned, a virtual computer with the tool already installed was accessed remotely by the participants in order to mitigate issues surrounding distribution and setup. In order to provide an additional understanding of the results, a think-aloud protocol was employed. Respondents were asked to talk the researcher through their thoughts while using the tool, and the researcher occasionally asked questions to further the interaction. The results were recorded as anonymized field notes.

Respondent 1 scored the tool at 72.5 in the SUS, indicating an adjective rating of “Good” or letter grade of “B”. The think-aloud protocol indicated that the teacher had mainly been interested in the tool as a way of finding relevant areas for planning lessons aimed at English 6. The reason for this was that the teacher had recently been given their first teaching assignment for that course, and worried about “keeping up” with the higher-proficiency students in terms of lesson themes. They also wanted a way of providing the students with support “beyond the textbook” in terms of grammar and form.

In terms of influence on their workload, the time spent on making corrections to students’ texts was brought up as the main perceived benefit of the tool. While the respondent believed that it would take a considerable amount of time to become comfortable with the tool and the report it provides, while also believing there could be long-term benefits in terms of workload.

Scoring and rating texts was also brought up as a consideration by respondent 1. Using the tool to survey error patterns over time, respondent 1 believed a system could be put into place where grammar exams would be less necessary and proficiency could instead be tracked through written assignments. Specific grammar assignments could then instead be focused on whichever patterns emerge as frequent, rather than examining a full range of patterns.

Respondent 2 scored the tool at 85 on the SUS, corresponding to a letter grade “A” and an adjective rating of “Excellent”. Respondent 2 also indicated an interest in using the tool mainly to explore student texts for areas of interest to incorporate in lesson planning, much as respondent 1 did. In addition, respondent 1 experienced that the workload would often impact the quality of feedback, and was interested in how the tool could be used to facilitate the creation of individual feedback.

---

Feedback became the main topic of the interaction at this point, and the tool was discussed from this perspective. Further functionality desired by respondent 2 included the possibility to indicate error patterns in the actual texts and for the tool to be able to produce error annotations in the form of comments in the .docx documents. While the current output of error patterns in a table was considered interesting, the respondent believed that the function needed to be further extended in order to have an impact on workload.

Respondent 3 provided a SUS score of 80, indicating an adjective rating of “Good” or a letter grade of “B”. As with the other user testing participants, the main draw of the tool was the ability to identify areas for lesson planning and focus assignments. However, respondent 3 raised some concerns regarding the technical aspects of using the tool and the amount of information available. Respondent 3 worried that the effort needed to navigate the dashboard would demand more time than a teacher would be given to get acquainted with the output in a real-world scenario.

Moving on to the SF UES results for the user testing group, the results range from slightly positive to very positive (table 11). The reward factor stands out as being the most positively rated by all participants. The “Focused Attention” stands out as neutral across the board, except for respondent 3 who rated all factors as positive. The uniformity of respondent 3’s responses indicates a low rate of engagement with the content of the form. Respondent 1 rates the “Perceived Usability” as close to neutral, while respondent 2 rates the “Aesthetic Appeal” as neutral.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Resp. 1</th>
<th>Resp. 2</th>
<th>Resp. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused Attention</td>
<td>3</td>
<td>3</td>
<td>4,66</td>
</tr>
<tr>
<td>Perceived Usability</td>
<td>3,33</td>
<td>5</td>
<td>4,66</td>
</tr>
<tr>
<td>Aesthetic Appeal</td>
<td>4</td>
<td>3</td>
<td>4,66</td>
</tr>
<tr>
<td>Reward</td>
<td>4,33</td>
<td>5</td>
<td>4,66</td>
</tr>
<tr>
<td>SF UES Score:</td>
<td>3,66</td>
<td>4</td>
<td>4,66</td>
</tr>
</tbody>
</table>

Table 11: SF UES score per respondent and factor

5.4 Conclusion of Tool Development

From an implementation standpoint, the document format limitations of the tool make it vulnerable to changes in the rapidly developing digital resource landscape utilized by Swedish schools. However, the .docx file is accessible and conforms to what is now available via Microsoft, Google, and Apple. The tool
can also be used with plain .txt files, but this approach would likely result in an additional step for the teachers when preparing their students’ texts for analysis.

The use of a local JLT server means that the tool will not receive future updates to the JLT resources, which could include patches to fix incorrect pattern matches or introduce new patterns altogether. However, using an API to connect the tool to JLT’s own server would introduce a cost to the tool, while also necessitating an internet connection. In addition, the sending of student texts over the internet in this manner would not be considered an entirely ethical practice, or even legal from a GDPR perspective. Having the Dash dashboard initialized from a local server follows a similar rationale, but does not lead to drawbacks in the same way.

TextStat’s consensus score being reliant on AmE as a default is, as was seen in the initial tests, likely to cause issues. This is something that must be pointed out clearly in the user guidelines, so that the teacher can view the results accordingly. In practice, this means that the teacher must look through the results to identify false positives for the “misspelling” category when AmE or BrE has not been specified in an assignment.

Other points highlighted by the initial tests are, as with many traditional spell-checkers, to do with names and place names. This was especially visible in the sample text MimerEngGym107, which dealt with fantasy, but is likely to be an issue in many other situations as well. Once again, although this is likely to be recognized by the teacher from previous experiences with text editors, it should be pointed out in the guidelines and acted on manually.

The “whitespace” pattern being triggered by sub-headings and indices in certain word templates falls into a similar category of issue where manual interaction is needed from the teacher. Ideally, this issue could be handled proactively by supplying students with clear instructions and a preferred template, or by the resources used for the tool being modified to either correct or ignore whitespace errors.

The issue arising from the piece of poetry being analyzed as containing a high frequency of JLT hits brings to the forefront the importance of understanding that a heavily prescriptive tool is not a catch-all solution. The tool designed for this dissertation is drawing on resources created for formal register and academic writing. Applying it to artistic and creative texts is likely to result in more manual labor for the teacher, and could ultimately result in a demotivating experience for the student. However, for formal to semi-formal texts in an academic setting the initial tests are encouraging in terms of the tool’s functionality.

The user testing pilot and main set returned interesting insights into how users experienced the tool, with the main result being possible to summarize as “ugly, but functional”. The SUS indicated that users had a decent experience using the tool, with the SF UES further indicating that the usefulness and reward factors were considered beneficial. The tool scored lower for aesthetic appeal.
and maintaining user focus, which can both be connected to the GUI design and the short time spent with the tool.

Furthermore, distribution issues are the likely cause of the low response numbers during the pilot, as indicated by correspondence with respondents. A solution to this was indicated via respondent comment, “I would prefer it to be a web-based program”, which would sidestep the distribution issues. However, such a solution would also require GDPR-compliant ways of hosting the tool and students’ texts, and introduce costs for hosting and maintaining an online version of the tool. While an interesting venue for future developments, it is not a solution that can be introduced within the scope of the current project. Finally, when comparing across the respondent groups, the SUS and SF UES scores indicate that the user experience is tied to previous experiences with using corpus linguistic methods, in the pilot, or expected reward in the current workplace, amongst the teachers involved in the user testing main set.
6. Teacher Experiences

This chapter moves the experimental tool created in the current project into classroom settings and is intended to explore how the application is experienced by in-service teaching staff in connection to their practices. The intervention phase of the dissertation contains two separate studies, with one concerning the use of the tool during one semester in upper-secondary English classrooms, and the second concerning reflective use in a tertiary education context.

Data is collected using a modified version of the previous interview guide (see ch. 2.4). Respondent drawings were also collected, and are used to compare the routines after the implementation of the tool. In addition, intervention participants filled out the SUS and SF UES questionnaires, so that changes in user experience could be explored.

The chapter starts with a short summary of the interview framework in the new context, including the rationale behind the new interview guide. This is followed by a description of how the tool was introduced to the intervention participants. Once the stage is set, the chapter describes the interventions, and summarizes the results using content analysis, as in the previous interview study (see ch. 4).

In order to see how the use of the tool has, or has not, affected the routines of the participants, the same instruments used for the survey are used for the intervention. However, the interview guide was changed in order to create a focus on the influence of the tool, as well as to create an evaluation of the usability. The interview guide makes use of the same four question categories as the previous guide (The Teacher and Variables, The Students, Routines and Feedback, and Digital Experiences and Tools), and is found in Appendix E. As with the survey, informed consent was collected in written form, with the consent form found in Appendix F.

While the first two question categories remain the same, as they were intended to elicit background data, the third and fourth were adapted to the evaluative purpose. The intention was for the questions to generate comparable answers to those produced in the initial survey, but also for them to contain explicit evaluations of the tool. As the tool is evaluated both concerning functionality and fit within routines, the thematization remains the same for the two latter question categories.

The interview materials are then presented in a narrative fashion in order to provide a holistic perspective on the participants’ experiences. The meaning units used in the narrative description are used to identify substantive codes or topics. The topics are then compiled into categories, or theoretical codes, that summarize the results of the interview study.

The use of the respondent drawing follows the same rationale of being able to view the intervention interviews in comparison to the initial survey
interviews. As the respondent drawings from the survey highlighted a cyclical pattern in the routines surrounding text submissions with an emphasis on the workload limiting the use of a desired multi-draft routine, the influence of the tool is especially interesting. The influence of the drawings, both as a focal point for discussion and an opportunity for expression, was experienced as benefitting the interaction by me, further motivating their inclusion in the evaluation. The benefits were mainly felt through access to an artifact that both the interviewer and interviewee could interact with and highlight specific key points on.

Both the SUS and SF UES user testing forms were administered to the intervention participants. This was partially done to obtain a quantitative evaluation of the user experience, through which the pattern of factors to do with functionality being scored higher than factors to do with the ability to maintain focus afforded by the tool could be further understood. The testing forms also allowed participants to submit additional comments anonymously after the debriefing.

Sampling for the intervention study was convenience-based and made use of my personal and professional network for recruitment. The participants of the intervention study were four in-service language teachers. One of the participants in the upper-secondary group (Jamie) had seen an earlier build of the tool before participating due to being involved in my VFU.

6.1 Interventions at Upper-Secondary School

As previously mentioned, two in-service upper-secondary EAL teachers were recruited through my personal and professional network. A third respondent was initially involved, but elected to end participation due to the workload caused by the pandemic. The participants were asked to use the tool for the spring semester of 2022 (from January to June).

Due to the low number of participants, the tool was introduced to the intervention participants individually. This was done through meetings in person for the upper-secondary participants. The introduction included working through the quick start document, starting the tool, and generating a report based on one of the Mimer EAL corpus samples used for the controlled environment testing. The report was then read together and discussed, in combination with explanations of the different visualizations provided by the tool’s output.

After the introduction to the tool, the participants were given free rein on how to make use of it in their practice. Both participants expressed an intention to use the tool as a support in their preparation for the national exams. Contact was maintained throughout the semester, and technical support was provided by me when needed.

At the end of the semester, interviews were carried out to debrief the participants regarding their experiences with the tool, and respondent drawings were collected in order for comparisons with the targeted routines to be
possible. As previously described, the same interview instrument used in the initial survey was adapted for the debriefing interviews, and the same respondent drawing template was used. The participants were given the pseudonyms Jamie and Charlie. An overview of their background information is seen below in Table 12.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Experience in years</th>
<th>Eng. 5</th>
<th>Eng. 6</th>
<th>Eng. 7</th>
<th>Primary/Lower-secondary experience</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamie</td>
<td>6</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Charlie</td>
<td>13</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

*Table 12: Base variables for upper-secondary intervention participants*

### 6.1.1 Jamie

Jamie works in a small upper-secondary school with a vocational profile in Sweden. They teach both the English 5 and English 6 courses, in addition to History and Religion. Jamie has worked as an upper-secondary teacher since 2016, and they are licensed for English and History. In addition, they also teach religion. They got their teaching degree through the teacher training program and have worked at the upper-secondary.

Jamie accessed the tool through the Google Drive repository and was able to install it on their work computer without issues. The tool functioned as intended during the intervention period, but Jamie did run into some format issues as the software used for exams (Inspera) outputs PDF. This was solved through document conversion outside of the tool.

When describing the student groups, Jamie emphasized the diverse background and language levels found:

(7) "Vi är en skola som har mycket introduktionselever, och man saknar ofta många ämnen när man börjar hos oss. Tyvärr ser vi ju också elever som fått betyg i grundskolans Engelska trots att de inte borde ha fått det. De hamnar då i Engelska 5 och får kämpa där trots att de inte kommer att kunna klara det." (Jamie)

My translation:

(7) "We are a school with a lot of students in the introduction program, so students are often missing many subjects when they start at the school. Sadly, we also see students who arrive with a passing grade from secondary school even though they should not
It is important to note the vocational profile of the school here, as the student group as a whole is not described as being especially interested in the non-practical core subjects, such as English. As Jamie describes, there are also students in the group who do not have the expected prior knowledge in the subject but have still received a passing grade. This places them in the standard track for their grade level, which makes the subject more difficult for them as they need to both keep up with the new content and make up for the missing content from lower-secondary school. Jamie elaborated on this with a further description of the background factors:

(8) "Det är ju självklart väldigt beroende på individ [...] Man måste nästan lära dem gå i skolan först, innan man kan ta tag i grammatiken [...] Vi har elever som inte tidigare har gått i skolan mer än några år, så man behöver jobba med väldigt mycket samtidigt. Ett vanligt problem är ju att om man ska prata om 'En' och 'Ett' i engelskan, vilket de ska ha med sig från grundskolan, så måste man börja med vokaler och konsonanter. Man behöver alltså backa bandet längre än man kanske egentligen ska behöva göra." (Jamie)

My translation:

(8) "It, of course, depends a lot on the individual [...] You almost have to teach them how to go to school first, before one can start working on grammar [...] We have pupils who have not gone to school for more than a few years before, so you need to work on a lot of things at the same time. A common problem is that if you are going to talk about ‘A’ or ‘An’ in English, which they should know from secondary, you have to start with vowels and consonants. You need to reverse the tape further than you probably should need to.” (Jamie)

As can be seen in the quote above, the classroom context described is extremely diverse in terms of students’ prior knowledge and experiences. Jamie describes that the group needs support not only with the subjects at hand but also with the general context of going to school. This has to be done in parallel with attempting to bring everyone up to the level needed to start processing the central content of the intended course. When describing the situation surrounding the indefinite article, Jamie refers to having to make up for content intended to be taught quite early on in the pupils’ schooling, but which is
necessary at the upper-secondary level. "Rolling back the tape" this far is quite drastically shrinking the number of hours available for the group to cover the central content of the course.

Jamie also describes a noticeable difference between the pupils’ written and spoken proficiency:

(9) "Muntligt brukar vara lättare för eleverna eftersom många är väldigt sociala. Man lär sig den sociala kontexten snabbare än den akademiska, som man behöver för att skriva." (Jamie)

My translation:

(9) "Speaking is easier for many of the students, as many of them are very social. One learns the social context faster than the academic, which they need for writing." (Jamie)

This specification of a social contra an academic context being enforced in the different forms of production stands out as interesting and could indicate that the group is vulnerable to avoidance issues when working with writing as they feel less comfortable with it and do not enjoy English. Turning to student writing, Jamie indicates spelling as a big issue due to the prevalence of spell-checking in most of the pupils’ other written communication:

(10) "Stavningen är ett stort problem, särskilt när de skriver i program utan rättstavning. Man förstår ju vad de menar, men det går för fort och de slarvar. De har alltid rättstavning annars." (Jamie)

My translation:

(10) "Spelling is a big problem, especially when they are using software without spell-checking. One understands what they mean, but they write too fast and they get sloppy. They always use spell-checking otherwise." (Jamie)

The issue described by Jamie can be broken down into two: reliance on spell-checking and lack of proofreading. While spell-checking software in general is not an issue, it becomes an issue when it is not always available for high-stakes situations. Jamie highlights one such situation in connection to the national test:

(11) "Problemät är att under NP får man ingen hjälp, så jag vill inte ge för mycket hjälp under lektionerna när de ska skriva själva. Då blir de chockade när de sätter sig med NP och inte kan fä
någon hjälp. Jag låter dem ibland skriva för hand för att de ska lära sig att skriva utan hjälp.” (Jamie)

My translation:

(11) "The issue is that during the national exams no help is allowed, so I do not want to provide too much help during the lessons when they are writing by themselves. If I do, they become shocked when they sit down with the exams and cannot get any help. I sometimes have them write by hand so that they will learn to write without supports.” (Jamie)

At this point the interview turned more explicitly to the use of the tool, and how it could be used to view the texts created without supports:

(12) "Texterna vi har här [pekar på verktyget] är ju gamla nationella prov, som de producerat själva utan rättstavning. Här ser man ju hur ‘TYPOS’, felstavningararna, skjuter iväg.” (Jamie)

My translation:

(12) "The texts we have here [points to the tool] are old national exams, that they produced by themselves without spell-checking. Here one can see how the ‘TYPOS’, misspellings, shoot up.” (Jamie)

However, not all misspellings are created equal, and the software does not provide any distinction between what is likely the result of a sloppy writer or what could be due to the student not knowing how to spell the word. Jamie indicates the importance of discussing these things with the class, and emphasizes conscious proofreading as an important factor in success:

(13) "Man får ju en väldigt tydlig sammanfattning. Annars har man ju en individuell bild av eleverna, men här får man en översyn av gruppen. [...] Jag har ju lyft flera gånger att de måste läsa igenom texten innan de lämnar in så att man inte får onödiga stavfel. Skillnaden mellan ‘mistakes’ och ‘errors’ har vi pratat mycket om. Det är svårt för dem, och de är stressade. De tycker inte att det är roligt och vill precis klara sig.” (Jamie)

My translation:
“You do get a very clear summary. Before one would have an individual picture of the students, but here one gets an overview of the group. [...] I have brought up multiple times that they have to read through the text before submitting it so that they do not get unnecessary spelling errors. We have talked a lot about the difference between ‘mistakes’ and ‘errors’. It is difficult for them, and they are stressed. They do not enjoy it, and they want to just barely pass.” (Jamie)

Taking the previously mentioned problem of avoidance issues into account, what Jamie describes here contributes to a very delicate situation surrounding feedback. When the students are stressed, not motivated to engage with the subject and would prefer to just ”barely pass”, feedback can easily push them to disengage with the teaching if done incorrectly, i.e., overwhelming the students or in other ways feeding into the development of avoidance issues. Jamie describes the feedback situation with an emphasis on the relational aspect:

“Jag jobbar på en liten skola, så man har en ganska personlig relation till eleverna och vet hur hårt man kan gå åt dem. Många texter är ju inte godkända texter, och då får man hitta vad vi ska fokusera på just den här gången. Man får ju också tänka på vad det var vi pratade om förra gången och lyfta att de gjort det bra. Om jag rättar en text sitter ju inte jag och rödmarkerar allt, utan väljer vad vi fokuserar på.” (Jamie)

My translation:

“I work at a small school, so I have a fairly personal relationship with the pupils and I know how hard I can push them. Many texts cannot pass, and then you have to find what to focus on at that specific time. One also has to think about what we talked about last time and highlight that they have done those parts well. If I am correcting a text, I do not just mark everything red, but I pick what we are going to focus on.” (Jamie)

The criterion for beneficial corrective feedback presented by Ferris (2010), such as error selection and self-revision, are highlighted throughout Jamie’s approach to feedback. By drawing on the relational approach Jamie has towards their pupils, Jamie is able to make decisions regarding the focus of the corrective feedback, and the amount of feedback possible without creating avoidance issues in the student group. As seen in the previous teacher interviews, there is also an interaction pattern similar to “two stars and a wish”, i.e., mixing
corrective feedback with positive feedback in order to highlight areas of progress for the pupil, as seen in the previous survey. This is an important part of maintaining engagement and creating a positive learning environment.

When discussing the use of the tool in such practices, Jamie points out that the tool does not take the level of the student into account:

(15) "Man skulle behöva välja någon inställning baserat på hur ens elever ligger till, och få välja vad man ska fokusera på. 'Detta vill jag titta på idag', liksom." (Jamie)

My translation:

(15) "You would need some sort of setting based on the level of your pupils, and select what to focus on. ‘This is what I want to focus on today’, kind of.” (Jamie)

While the tool offers the ability to select pattern categories to focus on in the analysis, what Jamie is talking about would be a level of selection based on pattern profiles tied to grade or proficiency levels. In addition, Jamie would like to see the tool being able to more clearly provide suggested focal areas based on the input texts:

(16) "Jag skickar till programmet och det säger till mig att nästa gång måste vi prata ordföljd." (Jamie)

My translation:

(16) "I send it to the software and it tells me that next time we have to talk about word order.” (Jamie)

In addition, Jamie would like to see a way of incorporating the pupils’ goals in the analysis, especially in relation to grading:

(17) "Man måste kunna välja vilken nivå man vill nå, så att feedbacken är relevant för det man siktar på. Betygen är ofta motorn för eleven, de ser ofta betyget som betalningen i slutet av kursen. Om man inte har betygen med på något vis så tar nog inte eleverna det seriöst.” (Jamie)

My translation:

(17) "You have to be able to decide on which level you want to reach, so that the feedback is relevant to what you are aiming for."
The grades are often the driving force for the pupils, they often see the grade as the payment at the end of the course. If you do not include the grading in some way the students are unlikely to take it seriously.” (Jamie)

The statement above highlights one of the pitfalls discussed in the tool development: scoring the texts. Since the tool was designed as support for data-driven teaching and language acquisition, it fails to navigate the realities of the school context. Scoring and grading are central components to the learning environment for Jamie’s pupils, and, since that component is explicitly missing from the tool design, it creates a mismatch that needs to be bridged by the teacher. However, it would also influence the teacher to make grading and scoring decisions based on the output, which has not been incorporated in the tool’s design. In essence, the tool does not score texts, but the output could be interpreted in such a way.

On the topic of the tool fitting with the teaching practices, Jamie provided a description of routines before and after having used the tool. The school uses a 1-to-1 setup with Chromebooks for the pupils. Texts are written in Google Docs and submitted via Google Classroom. Before using the tool, Jamie’s pupils would submit their texts to Jamie, who would then quickly skim through them and run controls for plagiarism.

The main points that Jamie looks for when reading the drafts are whether or not they have followed the task description and how they have done so far. The texts are then commented on and sent back to the pupils for revision. After revision, the final version is submitted via Google Classroom. This is similar to the routines described by teachers in the previous interview study. It is important to note that examining assignments fall outside of this routine, and instead makes use of a tool called Inspera, which limits support for the pupils and makes a direct submission to the teacher in PDF format.

In the routine with the tool, it is used at the initial and final submission. The routine remains unaltered in terms of the steps taken, but the tool is applied whenever the teacher interacts with the texts. Jamie created two different routines when filling out the respondent drawing (Fig. 35). The first one is as described above, but the second one describes the “dream scenario” for the tool.
In the second scenario, Jamie would like the tool to take on the role of an extra teacher who could provide a round of feedback for the pupils before submissions are delivered to Jamie. This would essentially create an additional round of automated feedback provided before the teacher is actively involved. The feedback should be tied to the content of the course and make reference to the grading requirements:

(18) "Eleverna kunde skicka in direkt till programmet och få feedback kopplad till kunskapskrav och styrdokument. Eleverna kunde få feedback innan de skickar in texterna till mig." (Jamie)

My translation:

(18) "The pupils could submit directly to the software and get feedback connected to knowledge requirements and regulatory documents. The pupils could get feedback before submitting the texts to me." (Jamie)

Jamie traced this scenario on the respondent drawing (Fig. 35), and provided a description of the steps as being:
1. Pupils submit text to the tool via Google Classroom
2. The teacher gets a notification and indicates focus areas
3. The pupil receives feedback from the tool
4. The text is revised according to the feedback
5. The teacher reads the final text

The purpose of this suggestion is elaborated on by Jamie with a focus on the role of the teacher, the possibilities showcased by the tool, and the nature of written assignments:

(19) "Datorn och eleven interagerar och att lärare sedan blir något mer slutgiltigt. Gärna feedback på uppgiften och rättstavning, alltså så som datorn redan kan hjälpa dem. Som lärare är det kanske mer att man hjälper dem utveckla uppgiften utifrån sig själva eftersom man känner dem och vet vad som är nästa steg. Då kan ju datorn hjälpa dem med grundgivarna istället. Exempelvis stavfel, ordföljd och sådana saker.” (Jamie)

My translation:

(19) "The computer and the pupil interact and then the teacher becomes something more final. Things like feedback on the assignment and spelling, the kind of things the computer can already help them with. As a teacher it is more like you were helping them with developing the assignment based on themselves since you know them and since you know what the next step is. Then the computer can help them with the basic stuff. For example, things like spelling, word order and such.” (Jamie)

The extension of the use of the tool suggested by Jamie indicates a desire for a more fully automated system, rather than the support tool designed for the current dissertation. Such a task would put much higher demands on the software used, as automated feedback that is delivered directly to the pupil sidesteps the use of the teacher as a facilitator and filter for information. Jamie also mentions the limitations considered when suggesting the extension of the tool:

(20) "Ur en social kontext bedömer man ju också textens innehåll, och datorn förstår ju inte.” (Jamie)

My translation:

(20) "In a social context, one also evaluates the text's content, and the computer doesn’t understand that.” (Jamie)
(20) "From a social context one also assesses the content of the texts, and the computer cannot understand it." (Jamie)

While not experiencing any technical issues while using the tool, Jamie does have considerations regarding the report provided, stating that the amount of information provided is overwhelming. In some ways Jamie considers this to be a similar experience to reading the text manually, in that the experience has to be scaled down in order to make sense:

(21) "När jag får en text måste jag skala ned på det jag ger feedback på, alltså vad vi ska fokusera på. För att hjälpa mig i vardagen måste ju verktyst get också hjälpa mig med det och visa att detta behöver vi fokusera på nu." (Jamie)

My translation:

(21) "When I get a text I have to scale down what I intend to give feedback on, what we are going to focus on. In order to help me in everyday work the tool will also have to help me with that and show me what we need to focus on at a given time." (Jamie)

This returns to the idea of somehow narrowing the output of the tool, along the same lines as discussed when level profiles were suggested. Jamie also goes on to suggest that the level profiles could be useful for the narrowing of the report:

(22) "Om man bara tittar på overviewen här [pekar på verktyst get], när man har haft det ett tag så känner man ju som lärare igen kategorierna. Som lärare hade jag ju velat rangordna dessa. Först måste de stava de individuella orden rätt, därefter tittar vi på ordföljd och grammatik, därefter tittar vi på punktuation. […] Då hade ju eleverna oavsett nivå kunnat titta och säga ’Aha, då måste jag in och jobba på nivå 3’." (Jamie)

My translation:

(22) "If you just look at the overview here [points to tool], when you have used it for a while you will recognize the categories as a teacher. As a teacher, I would have liked to tier these. First, they have to spell the individual words correctly, then we look at word order and grammar, then at punctuation. […] That way the pupils could also have a look and say ‘Aha, then I have to work at level 3’." (Jamie)
To summarize the interview with Jamie, the tool is working as intended in terms of parsing the documents and creating the output. However, having used it as designed, Jamie has provided a perspective on future development as well as an evaluation of the current state. Jamie works with pupils who often have low proficiency in English, and who are not entirely comfortable within the upper-secondary system. Jamie puts a lot of emphasis on the relational work with the pupils, as this is an important aspect of making them more comfortable in the school environment. This has shaped the form of the feedback provided and puts high demands on the teacher’s role as filter and facilitator.

The tool was mainly used during a period aimed at preparing the pupils for their national exams, which aligned well with the more formal requirements put on the text by the tool. According to Jamie, the tool highlighted issues to do with the prevalence of spell-checking software and worked well to create overviews of group-level patterns that would otherwise be harder to notice. The reports created by the tool were initially quite difficult to read, and the information provided was experienced as overwhelming for the teacher. Jamie suggested either a tier system or a grade level-based system as ways of narrowing the information down to a more accessible level. Future developments suggested by Jamie were, beyond the tier or grade level profiles, to include some sort of positive measurements in the report and to work towards a version of the tool that could be more pupil-facing.

6.1.2 Charlie

Charlie has been working as a teacher for 13 years and has taught at levels from primary to adult education. Charlie is currently employed as an upper-secondary teacher, and is licensed to teach English and History. In addition to the subjects for which they are licensed to teach, they have also taught upper-secondary Spanish and worked as a multi-subject substitute teacher.

Charlie initially downloaded the tool setup from Google Drive after receiving the necessary permits from the school’s IT department. However, after installation, it became apparent that the tool would not run due to the external stylesheet used to parse the .docx format into the .txt needed for the analysis. This was solved by branching the versions of the tool and supplying Charlie with a version that made use of an all-included process for transforming the text files. This process made use of the doc2text resource²⁸, rather than the procedure outlined in section 5.1.

The school at which Charlie works had a difficult time surrounding the emergence of the Omicron variant, and the planning of the semester was affected in such a way that the tool could not be used as extensively as Charlie

---
²⁸ John L. Sutherland through https://pypi.org/project/doc2text/ (7/8 -22)
had initially planned as other things had to take priority. In addition, attendance during the pandemic put further strain on classroom planning.

Charlie mainly teaches within the natural sciences and technology programs, both of which are preparatory programs for university studies. When asked about challenges in the English subject for their groups, Charlie describes the situation as follows:


My translation:

(23) "In most of the groups idiomatic correctness is not one hundred percent there. They struggle with it even at the higher levels. They do make certain minor grammatical mistakes. They have gone through the rules previously, but they make mistakes. Subject-verb agreement remains throughout English 5. There are many different things you would want to highlight for different pupil groups.” (Charlie)

Charlie also indicates that there are differences between the two programs that they teach. They also suggest that the use of digital interfaces prevalent within the technology program might influence some of the common errors found among the students within that group:

(24) "Jag undervisar i huvudsak på natur och teknikprogrammet. Tittar man på teknikprogrammet är det ju många som bommar exempelvis stor eller liten bokstav."

[...]

"Ta en sådan sak som stor bokstav i förstaperson pronomen. Man tänker sig att det beror på IT vanan, att det bara flyter på och de bryr sig inte om sådana saker. Det är olika för olika grupper.” (Charlie)

My translation:
"I mainly teach the natural science and technology programs. If you look at the technology program there are many who miss, for instance, capitalization."

"Take a thing like capitalizing the first-person pronoun. One could think it depends on their IT habits, that everything just flows and that they do not care about those kinds of things. It is different for different groups." (Charlie)

While there are differences between the groups in terms of error patterns, there are also differences between the mediums of production, such as written and spoken. Charlie describes the differences between the groups’ speaking and writing as being difficult to define, but including self-revision:

"In speaking I think there is both more and less. It is difficult to put your finger on something at a group level, because they will often correct themselves when they make a mistake. For example, they will start with ‘I are’, but immediately correct themselves to ‘I am’." (Charlie)

It is interesting to note the difficulty of defining recurring patterns at the group level here, as it could indicate a diversity within the proficiency levels within the classroom environment with errors being individualized to a high level amongst the pupils. Moving on to the diversity of their groups, Charlie describes having pupils at both extremes:

"Det är extremt stora skillnader. I en av mina Eng 5 klasser har jag en som är modersmålstalare och har bott större delen av sitt liv i U.S.A., och sen har jag elever där man kan fundera kring hur de passerade årskurs 6. Vissa kan man ju, när de börjar en kurs, säga att de kommer landa på ett A efter att ha träffat dem i 30 minuter, andra kan man se att de inte kommer att kunna klara
There are extremely large differences. In one of my Eng 5 classes, I have one who is a native speaker and has lived most of his life in the U.S.A., and then I have students where you can wonder how they passed grade 6. For some you can, when they start the course, say that they will land on an A after meeting them for 30 minutes, and for others you can see that they will not be able to pass the course. Everything depends on a correct assessment being made from below."

Charlie offers an explanation of the diversity of proficiency found as having to do with the previous schooling to some extent:

"One of the schools we often receive students from has consistently had the same teacher for several years, but at another they have had six teachers during lower secondary. They have not had any continuity there. It becomes very scattered in different ways." (Charlie)

Moving on to discussing the tool, Charlie found that the results differed quite a lot between the different courses. It is important to note that Charlie had mainly used the tool for the English 6 and English 7 courses due to experiencing that the tool functioned better at the higher proficiency levels. This was due to the high number of patterns found amongst the English 5 students. The differences experienced between the courses are described by Charlie as:

"Jag har främst kört det här i Eng 6 och 7, och tittar man på Eng 7 så är det väldigt mycket spelling som den markerar. Det är väldigt mycket namn och brittisk stavning, men det har varit det mest frekventa."
"På Eng 5 har det varit väldigt mycket punktuation och typografi. De har stuckit ut absolut mest." (Charlie)

My translation:

(28) "I have mainly run this in Eng 6 and 7, and if you look at Eng 7, there is a lot of spelling being marked. There are a lot of names and British spelling, but it has been the most frequent." (Charlie)

[...]

"In English 5 there has been a lot of punctuation and typography. They have stood out the most." (Charlie)

According to Charlie, the first-year course mainly returns error patterns that have to do with punctuation and typography, while the spelling error category is the most frequent one for the later courses. Discussing how the tool has interfaced with the systems already in use by the school, Charlie explains that their workplace makes use of two different systems for handling text submissions and feedback:

(29) "Vi har två olika system. Vi har en lärplattform som heter V Klass, men det är omöjligt att kommentera texter på ett smidigt sätt i den, så vi har även inlämningar via Microsoft Teams. Nackdelen där är att man inte får någon notifikation. Man väljer beroende på hur man vill jobba." (Charlie)

My translation:

(29) "We have two different systems. We have a learning platform called V Klass, but it is impossible to comment on texts in a comfortable way in it, so we also have submissions via Microsoft Teams. The disadvantage there is that you do not receive any notifications. You choose depending on how you want to work." (Charlie)

The use cases for the two systems can be generally divided so that V Klass is used for final submissions and summative feedback, while Microsoft Teams is used for drafts and summative feedback. The reasoning is that the in-text comment functionality available in Teams allows for the teacher to indicate areas in the text and provide more detailed feedback on specific language structures and content sections, while V Klass inhabits an administrative role.
V Klass is thus used for situations where a final grading or submission needs to be logged, communicated and documented. In addition, V Klass offers notifications for the teacher, while Teams makes it necessary for the teacher to manually check in and see who has submitted which assignment. Charlie’s workplace also makes use of programs specifically designed for examinations:

(30) "Vi använder ibland testprogram där de inte har rättstavning så att man kan få en mer genuin bild av deras språk utan att datorn korrigerar misstag. " (Charlie)

My translation:

(30) "We sometimes use test programs where they do not have spell-checking so that you can get a more genuine picture of their language without the computer correcting mistakes." (Charlie)

The school makes use of the Exam.net service, which offers a secured browser-based environment for the pupil to take the exam in. The pupils have access to supporting tools according to the settings provided by the teacher at the creation of the exam session. The output is then available for download by the teacher.

On the topic of submission routines, Charlie’s respondent drawing (Fig. 36 on the next page) describes a numbered exchange between the teacher, computer, and pupil.
The routine remains largely unchanged with the introduction of the tool, as the steps where the teacher is uploading to the tool and receiving feedback would previously instead have contained a skimming of the text and a deeper reading with a focus on the language. When discussing how Charlie had intended for the tool to fit within their routines, they describe the planned workflow as taking the form of a monthly check-in:

(31) "Jag hade nog först planen att de skulle skriva någon text i månaden, och sedan köra dem härigenom för att urskilja vad de behöver förbättra." (Charlie)

My translation:
"I initially had the plan that they would write a text or so a month, and then run them through this to discern what they need to improve." (Charlie)

Due to the pandemic, the amount of time available for experimenting with the tool was limited. In addition, the amount of data generated was experienced as overwhelming. The function of the tool in Charlie’s practice is described as providing a second opinion, which Charlie accessed through a quick reading of the results after conducting their skimming:

"De tillfällena då jag kört detta har jag snabbkollat. Det blir ofantliga listor när man kör en hel klass, så jag har snabbt scrollat igenom för att kolla om det fanns något specifikt jag missat. Då har jag gått tillbaka och poängterat detta. Så det har varit ett visst stöd i att gå igenom texterna en extra gång, som en second opinion nästan.” (Charlie)

My translation:

"I have skimmed on the occasions when I have run this. There are huge lists when you run a whole class, so I quickly scrolled through to check if there was anything specific I missed. If so, I went back and pointed this out. So, there has been some support in going through the texts an extra time, like a second opinion almost” (Charlie)

The use of the tool as generating a second opinion on the text as a complement to the manual reading could initially seem like a redundant practice, but Charlie indicated that it had been useful as a way of seeing new things in the texts or highlighting new features for the teacher:

"Varje gång vi testade det här programmet så kom det ju någon odd ball, där man kanske tidigare hade bedömt som ett F men analysprogrammet menade på att exempelvis komplexiteten var högre än man trodde. Just måttet komplexitet och fel per mening funkade ju väldigt mycket bättre i Eng 7.” (Charlie)

My translation:

"Every time we tested this program there was an oddball, where you might have previously rated it as an F, but the analysis program meant that, for example, the complexity was higher than
you thought. The measure of complexity and errors per sentence actually worked much better in Eng 7.” (Charlie)

The manual reading and overview produced by the tool were then used to formulate written feedback on Teams or V Klass, depending on whether summative or formative in nature. Charlie also discusses the use of oral feedback 1-on-1, and emphasizes the importance of it in the relationship between teacher and pupils:

(34) ”Jag försöker ge muntlig feedback minst en gång per termin, så att man kan sätta sig ned och prata med eleven. Då kan de ställa frågor om de har något de undrar över, och det är viktigt att de känner sig hörda. Det är viktigt att ha en dynamisk lärarroll och inte den auktoritär, så att det känns lite mänskligt också.” (Charlie)

My translation:

(34) "I try to give oral feedback at least once per semester, so that you can sit down and talk to the student. Then they can ask questions if they have something they are wondering about, and it is important that they feel heard. It is important to have a dynamic teaching role and not the authoritarian one, so that it feels a bit human too.” (Charlie)

Having used the tool in the different English language courses taught at the school, Charlie recommended two areas where the tool had to be improved in order to function well as a teacher support.

(35) ”Det är två saker som jag hade velat ha smidigare med programmet. Dels är det tiden för att analysera texten. När jag har matat in en elevskara på 30 elever där alla skickat in 2 sider är det mycket att gå igenom. Om min dator ska processa alla texterna tar det nästan en halvtimme, och under den tiden kan jag inte använda min dator. Då måste man planera in tid för att köra detta, och där jag inte kan använda min dator. […] För min del har det varit just tiden som gör att jag dragit mig lite för att köra programmet.” (Charlie)

My translation:

(35) "There are two things that I would have liked to have done more smoothly with the program. First is the time needed to
analyze the texts. When I have loaded a student group of 30 students where everyone submitted 2 pages, there is a lot to go through. If my computer has to process all the texts, it takes almost half an hour, and during that time I cannot use my computer. Then you have to schedule time for running this, and where I cannot use my computer. [...] For my part, it has been the time needed that has caused me to refrain a little from running the program."
(Charlie)

The speed of text processing mentioned here by Charlie has been considered an issue from early on in the development of the tool. The reason for this is that the tool runs locally, and as such is heavily dependent on the hardware available to the teacher. In order to mitigate this issue completely the processing would have to be taken off-site, meaning that a server would have to be set up and texts would have to be sent from the teacher’s computer over the internet. As discussed during the user testing pilot, this would bring a new set of issues to do with GDPR and the ethical aspects of handling student materials into play.

GDPR is also brought up as a concern by Charlie and is briefly discussed in connection to how pupils behave during the national exams. Before the exams the pupils were explicitly instructed not to use their names or sign the documents, as the assessment was to be done anonymously. However, Charlie still found that a fairly large number of pupils signed their names on the texts or included their identity in some other way. Due to this, the suggestion of running an off-site server environment to provide faster and more stable text processing was not considered viable by Charlie, unless a GDPR-compliant environment could be created.

The second issue brought up by Charlie has to do with the amount of data included in the reports provided by the tool:

(36) "Den andra saken som har gjort det tidskrävande är att det är väldigt mycket som räknas som fel. [...] I en Eng 5 grupp på 30 elever som har gjort 2-3 fel per mening, och har skrivit 50-60 meningar per elev så blir det väldigt mycket information. När man då måste in och sälla bland datan blir det mycket arbete."
(Charlie)

My translation:

(36) "The other thing that has made it time-consuming is that there are a lot of things that count as errors. [...] In an Eng 5 group of 30 students who have made 2-3 mistakes per sentence, and have written 50-60 sentences per student, there is a lot of information."
When you then have to sift through the data, it becomes a lot of work.” (Charlie)

The sifting mentioned by Charlie has to do with the tool’s inability to properly process proper names and British spelling conventions. This leads to an extensive amount of manual work if the produced reports are to be used in a pupil-facing manner. While sifting is still possible, the issues with these categories introduce a lot of noise into the report. The solution suggested by Charlie is to create a tier system that would help focus the teacher’s attention towards the errors that are actually relevant:


My translation:

(37) “The errors should be divided into 1st class and 2nd class errors. That way you can get a list of what is guaranteed to be wrong, and then check what needs to be sorted out. Then it would be a lot quicker.” (Charlie)

Deciding on what exactly should be considered a 1st or 2nd class error is a difficult topic, but initially the tier system could be focused on separating error-prone categories of JLT from better functioning categories, and separating patterns tied to British spelling conventions from actual spelling errors.

Charlie also shared a situation where a tool like this would have been useful and saved a lot of time:


My translation:

(38) "At my school, we had a joint assessment of texts after the national exams. [...] There will be around a hundred texts that you would have to read and argue over the results. This process took a long time - probably 30-40 hours. If we had had a more user-friendly analysis program that made it possible to sort out certain types of errors, we would have saved a lot of time. Then we could have skimmed instead. There is a golden potential here.” (Charlie)
have to read and argue for the assessment of. That process took many hours. Surely 30-40 hours. If you had a more user-friendly analysis program that did not require you to sift out certain types of errors it would have saved us a lot of time. Then you could have gone in and skimmed afterward. There is a golden potential here.” (Charlie)

While quite different from the intended purpose of the tool, Charlie provides a solid argument for a situation where the longer processing times would pose less of an issue. Granted, this would also necessitate the implementation of an error tiering system, as the output produced from such a large amount of input texts would likely exacerbate the already mentioned issues with the readability of the reports. This readability issue is further elaborated on by Charlie:

(39) "Det är ju inte så att det inte är tidsbesparande, det tror jag att det är. [...] Men på något vis är det ju psykiskt betungande att läsa listorna med fel och det känns som om det delvis är formulerat i kod när man möter en Excelfil. Det är ju en väldigt annorlunda upplevelse än att läsa skönlitteratur.” (Charlie)

[...]


My translation:

(39) "It is not that it is not time-saving, I think it is. [...] But somehow it is mentally taxing to read the lists of errors and it feels as if it is partly formulated in code when you read the Excel file. It is a very different experience than reading fiction.” (Charlie)

[...]

"It is possible that it takes less time than if you were to go through a text completely by yourself, but it is a big mental step to take. It is as if it is the last day of vacation, and you know you have to clean out the archives. It becomes quite a big mental step to take
once you say 'Oh, now we're doing it. It is a bit like that. I would like to think it is time-saving, but it is more mechanical’"

The mental strain described by Charlie is important to note, as it indicates a functional issue with the aesthetics of the tool that is severe enough to possibly counteract any benefits given by its use. As pointed out by Charlie, the experience of sifting through the data is very different from other types of reading that the teacher might often engage in, which makes the information more difficult to process. In some ways, this could be discussed from the perspective of subject literacy, or the previous experience the reader might have had with acquiring information from graphs and tables, which could transfer into how taxing the report provided by the tool would be to read.

Moving on to possible benefits gained by using the tool, Charlie brings attention to the possibility of using it as a way of equalizing the level of feedback provided to pupils:

(40) "En fördel är ju att alla lärare får förutsättningarna att ge liknande feedback. Att det blir mer rättvist. Ibland har man ju haft lärare som är briljanta, och ibland sådana som inte är så bra i sin feedback. Detta kan ju då vara ett stöd för att feedbacken ska bli mer likvärdig. [...] Det vet man ju själv att på de första texterna går man all-in, men på den 20e blir det inte lika mycket. [...] Att man oavsett innehållet kan ge en likvärdig feedback.” (Charlie)

My translation:

(40) "An advantage is that all teachers get the conditions to give similar feedback. That it becomes fairer. Sometimes you have had teachers who are brilliant, and sometimes ones who are not so good at feedback. This can then be a support for the feedback to be more equal. [...] You know that on the first texts you go all-in, but on the 20th there will not be as much feedback given. [...] That regardless of the content, you can give equivalent feedback.” (Charlie)

The description given by Charlie partially highlights the differences in feedback between different teachers, but also the variation in levels of feedback provided based on available energy.

When asked what they would need to see changed in order for the tool to be a good fit for their teaching practice, Charlie highlights three separate points of necessary improvements:

My translation:

"The time it takes to process the texts, and that you would have some kind of different levels to sort the problems by. That this thing with proper names does not appear in the same category as more serious linguistic mistakes. Something that makes it easier to sift through everything. And then the aesthetics, that it is not as hard to go through the list." (Charlie)

In summary, the interview with Charlie indicates that the tool works well with higher-proficiency language learners but creates too much noise in the output for the lower proficiency levels. The main areas of functional issues are the time needed to complete the text analysis process and the lack of a tiering system for the error categories. The solutions suggested in the interview align well with issues and solutions seen in the previous interviews and user tests.

The problems created by the aesthetics and readability of the report are important to note as the mental strain caused by them is likely to make prolonged use problematic. The analogy used by Charlie, comparing the reading of the report to the last chore of the vacation, highlights an issue that might result in the tool not being used at all. This is especially poignant when considered in combination with the long processing time experienced with Charlie’s work laptop.

Charlie also highlighted possible areas where they considered the tool to be beneficial. The situation described surrounding the joint assessment of exams stands out as an interesting use case. Furthermore, the benefits in terms of equality of feedback discussed should also be considered a benefit of the tool, or rather automated diagnostics in general. The tool was also indicated as fitting well within Charlie’s routines surrounding text interactions and feedback, as it fulfilled a function already being performed.

6.1.3 SUS and SF UES at Upper-Secondary School

As previously mentioned, the SUS and SF UES forms were administered anonymously, and will not be tied to the pseudonyms used for the intervention participants. This section will instead present them under the titles respondent 1 and respondent 2 and analyze them with a general background in the interview results.
Respondent 1 returned a SUS score of 57.5, which corresponds to a letter grade of “D” and an adjective score of “Poor”. Looking at the factors presented by the SF UES form, a similar total score was recorded (Table 13). The reason for this score is visible in the individual factor scores, which indicate that the tool was experienced as having high usability and providing an above average reward, but the aesthetics were scored neutrally and the focused attention afforded by the tool was scored as poor.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Resp. 1</th>
<th>Resp. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused Attention</td>
<td>2.33</td>
<td>2.66</td>
</tr>
<tr>
<td>Perceived Usability</td>
<td>4</td>
<td>3.33</td>
</tr>
<tr>
<td>Aesthetic Appeal</td>
<td>3.33</td>
<td>1.66</td>
</tr>
<tr>
<td>Reward</td>
<td>3.66</td>
<td>3.33</td>
</tr>
<tr>
<td>SF UES Score:</td>
<td>3.33</td>
<td>2.75</td>
</tr>
</tbody>
</table>

*Table 13: SF UES score per respondent and factor for upper-secondary respondents*

Moving on to respondent 2, a SUS score of 50 was reported, resulting in a letter grade of “F” and an adjective rating of “Unsatisfactory”. As with respondent 1, the lowest scores are “Focused Attention” and “Aesthetic Appeal” (table 13). Both the usability and reward were scored as neutral. The total score provided was 2.75, which is below average and indicates a less-than-satisfactory user experience.

Neither respondent provided a free-form answer, but both the factor scores and the overall SUS score match well with the interview results. In summary, the evaluation forms indicate that the functionality and purpose of the tool fit well with the participants’ needs and that the results are experienced as rewarding to a certain extent, but that the current form of the tool makes meaningful interactions difficult due to poor aesthetics and the low readability of the final report. It is important to note that the types of issues concerning processing time described by Charlie are not explicitly represented in the factors of the SF UES but are likely to contribute to a poor user experience.

### 6.1.4 Themes at Upper-Secondary School

The substantive codes from the interviews with the upper-secondary school participants will now be presented and condensed into themes. Each theme contains several of the topics mentioned in the interviews, but has been grouped according to the theme of the topics as they relate to the use of the tool. The themes are each given their own heading. The headings for the themes reflect the nature of the tool design process, and the categories have been thematized
with a background in how the results could come to inform future work. The themes found in the upper-secondary interviews were challenges, affordances, issues, ecological incompatibility, and future development.

6.1.4.1 Challenges

Several challenges were highlighted by the participants when discussing their experiences of using the tool. A general challenge was the diversity of knowledge levels and proficiency found in their groups. This became challenging due to the approach that the tool takes to language analysis, which does not in any way take the levels of the students into account. While a group of students with diverse language proficiency is not uncommon, incorporating that diversity into one’s reading of the tool’s output is experienced as a fairly difficult task that demands a lot from the users. This was especially highlighted by Jamie, in the context of facilitating appropriate feedback for the students.

Challenges that were more directly related to the use of the tool were the readability of the output and the aesthetic appeal of the software. Both participants indicated that the output was overwhelming and that they found it difficult to properly engage with. Partially, this had to do with the amount of data, but also the form of the tables included. The results of the SF UES also indicate that the aesthetic appeal of the tool was a big challenge for one of the participants, which further supports the interface being a challenging aspect.

6.1.4.2 Affordances

Affordances indicated by Jamie and Charlie were mainly to do with the group-level overview and the tool drawing their attention to patterns that they had not noticed before. The use of the group-level overview was conceptualized as a second opinion or support opinion, with which the participants could consult regarding their own reading of the texts. This was indicated as benefitting both patterns in the entire student group, but also individual patterns among the students.

The discovery of a new pattern, or getting thrown an “oddball”, was indicated as being a beneficial experience. The example lifted by Charlie had to do with text complexity, but could be understood in the context of new patterns emerging in other metrics as well. This is an area where the role of the tool as a second opinion becomes clearer, as consulting it was shown to highlight patterns that would otherwise have gone unnoticed.

The equalizing effect of the tool was also brought up during the interviews and could be considered a major benefit. The tool was experienced as a way of possibly offering a standardized library of focal points for teachers, which could lead to a more equal level of corrective feedback amongst students with different teachers.
6.1.4.3 Issues and Limitations

The theme of issues and limitations was initially considered as two separate headings, but were combined as it was difficult to create a clear distinction between the two. The reason for these difficulties was that limitations often ended up being the operationalization of issues, i.e., a limitation is put in place due to an issue with the task the user wants to perform.

The main issue brought up was the overwhelming nature of the output of the tool when interacted with in table form. This issue was due to both the mental strain put on the reader when reading the table and the number of false positives provided by the tool. As such, the overwhelming output issue could be further understood as being due to readability and precision.

This leads to limitations regarding the depth of engagement possible for the participants. Simply put, when the tool is being used in a working environment, the effort needed to engage with the output adds to the experienced workload in a problematic way. Furthermore, the lacking precision makes this deeper engagement necessary as the participants needed to manually inspect the data before it could be confidently used. This can be recognized from the testing done during my VFU period, and it can be concluded that the implementation of pattern selection in the tool did not fully solve the problem. It rather made it possible to negate the issue somewhat by de-selecting patterns that regularly caused false positives.

A second barrier of use brought up by the participants was the processing time needed for the tool to analyze the texts of a full classroom group. Charlie, especially, highlighted that this would deter use and lead to a higher experienced workload when using the tool. This issue limits the perceived usability of the tool depending on the hardware available. It must also be acknowledged that this is likely a limitation brought on by my lacking experience in software development, and could be alleviated by a more experienced programmer.

Finally, text scoring was mentioned by both participants, which highlighted a shortcoming of the tool in relation to the desired usage. Situations brought up by the participants could broadly be defined as high-stress periods in their practice, such as the assessment of national tests. Here there is a strong limitation of the current tool, as it was explicitly designed not to operate within those conditions. The limitations are to do with the previously mentioned precision of the analysis, as well as the output provided not being aligned with knowledge requirements in any way.

6.1.4.4 Ecological Incompatibility

Ecological incompatibility as a theme is intended to highlight points of failure for the tool’s integration into existing routines and contexts. The main point of failure was experienced by Charlie, and had to do with the downloading and running of executables on the work computer. This was caused by the security limitations placed on the user accounts by the school’s IT staff. This was also
experienced by one of the testers involved in the first trial run of the distribution model.

There were also compatibility issues stemming from the choice to make use of the .docx file format. While the initial assumption about the format’s accessibility held true, it led to issues when participants made use of alternative word processors for specific activities, such as test preparation. This was mainly experienced by Jamie, who ran into complications while using Inspera. As Inspera would only export PDFs, Jamie had to go through an additional step of converting files from PDF to .docx when using the tool.

A previously mentioned issue, the overwhelming output, could also be considered an ecological incompatibility. As alluded to under the issues and limitations heading, the output was experienced as overwhelming due to two factors: readability and precision. While the latter is a mechanical error within the analysis, the reading being experienced as overwhelming can also be seen as an incompatibility with the workload experienced by the participants. The failure to properly weigh the information presented with the time and energy available for the participants in their work environment could also be understood as a failure to integrate the tool into the existing context.

6.1.4.5 Future Development
The upper-secondary participants offered many suggestions for the future development of the tool. These could roughly be placed into three categories: output sorting, assessment usage, and student-facing features. Output sorting would be suggested developments aimed at supporting the users in engaging with the tool’s output. Suggestions included the tiering of error patterns according to severity and/or the tool’s precision and the use of focal areas. Both of these suggestions were intended to limit the amount of data included in the output, and thus leading to a smaller amount of output needing to be inspected. The use of focal areas could either be dependent on frequency or categorization of which patterns should be considered relevant in what course.

Using the tool for assessment and text scoring was brought up as interesting for high-stakes assessments, such as the national tests, but also as a way to perform co-assessment. The tool was envisioned as having the role of providing a baseline from which the teachers could then focus their work on texts that stand out in any way. Co-assessment of the national tests was brought up as a situation where the tool was seen as having a lot of potential.

Further automating the dissemination was also suggested as a desired development. The main desire was indicated as being able to use the tool in a more student-facing manner, i.e., having students interact directly with the analysis and the output. This would bring with it higher demands on precision, transparency, and the way in which error patterns are communicated. As the current version of the tool was developed as teacher-facing support, it lacks didactic considerations regarding the output.
Finally, the tool was indicated as needing a visual overhaul in order to make it more aesthetically appealing. This is important, as the visual experience influences the perceived effort of working with the tool, as well as engaging with the output. The aesthetics were indicated as an issue both in the SF UES and the interview data.

6.2 Interventions at University

Recruitment for the university participants as with the upper-secondary respondents, was done using a convenience-based sampling method. The respondents were both teaching writing instruction at the level of B2 in the CEFR, corresponding to the English 6 course in the upper-secondary context. The course is mainly geared towards exchange students, who are taking it in preparation for further English medium studies in different fields. The focus of the writing course was on academic writing and the form of the target texts could be described as formal essays.

As previously mentioned, the tool was introduced to the tertiary level respondent remotely through the use of Zoom. The respondents sat down individually with me and ran the tool together using one of the Mimer samples provided with the setup package.

The initial design for the interventions was for the respondents to make use of the tool for a semester and then participate in a debriefing interview, as was done with Jamie and Charlie, the upper-secondary-school participants. However, due to the increase in workload and general uncertainty brought on by COVID-19 during the semester, both respondents were unable to make use of the tool during the actual teaching of the course. As an alternative to this, the respondents instead made use of the tool after the course had finished, and were asked to evaluate it reflectively based on their teaching experiences from the semester.

The evaluations and interviews were carried out in person. The respondents made use of a computer with the tool already installed in order to alleviate the issues around set up and had the option of using either texts produced during their courses or one of the samples provided from the Mimer EAL corpus. The computer used was more powerful than the work laptops used by Jamie and Charlie, which means that the users’ experiences of the processing times of the text analysis are not comparable.

As with the previous group, the respondents have been given pseudonyms. The base variables for respondents in the tertiary instructor group are found in Table 14 on the next page.
Table 14: Base variables for upper-secondary intervention participants

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Experience in years</th>
<th>Upper-secondary</th>
<th>Primary/Lower-secondary</th>
<th>University</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quincy</td>
<td>9</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Brita</td>
<td>15</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

6.2.1 Quincy

Quincy has worked as a language instructor for nine years. They have only worked at a university level, but they are a licensed teacher for the English subject. The evaluation of the tool was done using texts from Quincy’s students, which had already undergone formative assessment. The proficiency levels found in Quincy’s group of students are described as being diverse, with the learners arriving at differing levels of capability.

When asked about recurring errors in the student group, Quincy indicated sentence structure, verb forms, and word order as being the most prevalent:

(41) “Meningsbyggnad, verbformer och ordföljd, skulle jag nog säga är de vanligaste i den här gruppen.” (Quincy)

My translation:

(41) "Sentence structure, verb forms and word order, I would probably say are the most common in this group." (Quincy)

Quincy describes that these error patterns occur in both oral and written production, although sentence structure is much less visible in the students’ oral production. While these are the most common error patterns in the student group, Quincy makes use of an extensive list of prepared annotations for use in commenting on student production, covering around 30 different types of errors.

When looking at the output of the tool, Quincy states that it has provided more annotations on punctuation than they had previously considered:

(42) "Det var mer kommatering och den typen av kommentarer än vad jag brukar kommentera på, var första tanken.” (Quincy)

My translation:

(42) "There was more punctuation and that kind of comment than I usually comment on, was the first thought." (Quincy)
Apart from punctuation, Quincy recognizes the report provided by the tool from their own assessment of the texts, although the categories are not entirely transparent:

(43) "Jag känner ju igen antalet fel i verktyget, det stämde överens med min bedömning. Men typerna kan jag inte riktigt säga något om eftersom jag använder andra termer. Skulle man använda detta måste man se över vilka ord man använder." (Quincy)

My translation:

(43) "I recognize the number of errors in the tool, it agreed with my assessment. But I cannot really say anything about the types because I use other terms. If you use this, you have to look at the words you use." (Quincy)

Regardless of the functionality of the tool, Quincy expresses some concern regarding the influence the tool might have on the reading of the texts.

(44) "Jag är lite kluvig i hur mycket insyn jag vill ha. [...] Det är lite på gott och ont att få att 'den här texten är ungefär här och den här ungefär här' utan att själv gå in helt blank." (Quincy)

My translation:

(44) "I am a bit torn about how much transparency I want. [...] It is a little bit good and bad to get that 'this text is roughly here and this one is roughly here' without going in completely blank yourself." (Quincy)

The idea of not being able to go into a text “completely blank” after having made use of the tool is an important thing to note. As Quincy states, it can be good or bad, but having a tool influencing assessment beyond the formative aspect creates much higher demands on the precision of the tool as it might influence grading to a further extent. As the tool was not designed for the purpose of summative assessment, the influence of it on the grading teacher would need to be explored further in a more controlled way.

On the topic of the functionality of the tool, Quincy has noticed some issues with the parsing of the student texts that they brought for the tool evaluation:

(45) "Den verkar ha svårt med radbrytningar och sådant, vilket stör väldigt mycket. Det ser jag som lite avskräckande. Det blir mycket att gå igenom och se om det är relevant." (Quincy)
(45) "It seems to have difficulty with line breaks and such, which is very annoying. I see that as a bit discouraging. There will be a lot to go through to see if it is relevant." (Quincy)

This is likely due to the same issue with parsing Word templates noticed in the controlled environment testing. This issue ties into the previously described problem of the amount of manual inspection needed in order for the analysis to be usable. While this specific case could be handled through the use of adapted templates, the problem at large will still remain due to the number of other errors in the tool’s analysis mentioned in the previous interviews.

In terms of the possible benefits, Quincy highlights the previous example concerning punctuation as a gain from using the tool:

(46) "On the subject of punctuation, it could have opened my eyes to things I do not usually focus on, and it could have affected the assessment. And that could have been good. You do not see everything, you are a human being doing the reading, so we could have complemented each other, me and the tool." (Quincy)

It is important to note the conceptualization of the tool as a second opinion reappearing here. The influence of the tool on Quincy’s workload is also discussed, and the answer indicates that it would have to be a long-term investment:

(47) "Det skulle öka arbetsbördan initialt eftersom man skulle behöva mycket tid att sätta sig in i hur det funkar. I det långa loppet, om man utvecklar kategorierna och eftersom man får förslag på kommentarer, så absolut. På lång sikt, ja, men det är nog en ganska lång startsträcka.” (Quincy)
"It would increase the workload initially because you would need a lot of time to familiarize yourself with how it works. In the long run, if you develop the categories and because you get suggested comments, then absolutely. In the long term, yes, but there is probably quite a long starting stretch.” (Quincy)

Quincy mainly makes use of Moodle and Word. Moodle is used for submissions and returning the texts to the students, while Word is used to interact with the actual texts:

"Det enda jag använder är Moodle, där laddar jag ned dokumentet och kommenterar på det i Word. Så, jag använder mig av kommentarsfunktionen i Word.” (Quincy)

"The only thing I use is Moodle, where I download the document and comment on it in Word. So, I use the comment feature in Word.” (Quincy)

While not explicitly instructing students on how to use spell-checking software or other digital supports, Quincy does instruct them to engage with the red underlines provided by, for instance, Word. Quincy describes the reading process as being based on a selection of focal points:

"När jag läser en text gör ju jag ett urval. Om någon gör 50 fel med verbformer så väljer jag kanske att inte kommentera på kommentering eftersom de redan har tillräckligt att arbeta med.” (Quincy)

"When I read a text, I make a selection. If someone makes 50 mistakes with verb forms, I might choose not to comment on punctuation because they already have enough to work with.” (Quincy)

When talking about the work routines with student texts, Quincy describes a multi-text routine that takes on the cyclical nature of previously described interaction routines:
"Oftast är det någon typ av formativ bedömning på första inlämningen, och sedan en inlämning till som bedöms summativt. Det är nog standardupplägget." (Quincy)

My translation:

"Usually there is some type of formative assessment on the first submission, and then another submission that is summatively assessed. That is probably the standard arrangement." (Quincy)

The students submit a text, get formative feedback, write a new text, and then get summative feedback on the final text (visualized in Fig. 37):

![Diagram](image)

**Figure 37: Quincy's respondent drawing (S = student, L29 = teacher)**

Quincy believes the tool would fit best for use within the first cycle, where the text is only given a formative round of assessment. This means that the tool

---

29 From "Lärare", which is teacher in Swedish
would be used to support the teacher in providing the student with formative feedback after selecting the relevant focal points:

(51) "Sedan skulle jag nog använda det som en utgångspunkt för formativ bedömning, och ha det i bakhuvudet när jag gör min slutbedömning. Så ser jag spontant att jag skulle kunna göra." (Quincy)

My translation:

(51) "Then I would probably use it as a starting point for formative assessment, and keep it in the back of my mind when I make my final assessment. That is how I spontaneously see myself doing it." (Quincy)

During the debriefing portion of the interview, Quincy summarizes their evaluation of the tool as demanding a lot of time and resources in order to implement it as a functional part of their practice:

(52) "Det skulle krävas att jag hade gott om tid och resurser för att sätta mig in i ett sånt här system, men jag kan ändå se någon vinning med det på lång sikt." (Quincy)

My translation:

(52) "It would require me to have plenty of time and resources to get into a system like this, but I can still see some benefit in the long term." (Quincy)

In summary, the interview with Quincy indicates that the tool has been useful in identifying new features of student writing that the instructor had previously not been aware of. It is important to note that the texts used for the evaluation had already been revised according to Quincy’s feedback, which made it difficult to answer questions regarding whether or not the tool was able to identify patterns that were commonly commented on by the teacher.

Quincy noted the issues with Word templates as being discouraging when considering the use of the tool due to the amount of manual labor needed to go through and sort the results. This, in combination with the unfamiliar terminology used in the error categorization, indicated that a lot of effort and time would be needed in order for Quincy to implement the tool within their practice in a good way.

However, the tool seemed to be a good fit for the routines that Quincy’s interactions with student texts. Due to the issues mentioned above, Quincy
evaluated the tool as being useful in the early cycle of their routine, where the focus is on selecting focal points for formative feedback.

6.2.2 Brita

Brita has worked as a language instructor at the secondary, upper-secondary, and tertiary levels for 15 years, and is a licensed teacher for the English subject. Brita is evaluating the tool against the background of a writing course within the same preparatory program that Quincy taught. When describing common issues amongst the students, Brita indicates grammar and text structure as recurring areas:

(53) “Jag skulle nog vilja säga att det är framför allt grammatik de har problem med. Men också rent generellt hur man strukturerar en text och hur man får en text att hålla ihop.” (Brita)

My translation:

(53) "I would probably say that it is mainly grammar that they have problems with. But also how to structure a text and how to make a text hold together in general." (Brita)

Brita explains that they provide feedback on a wide range of things, including congruency, sentence fragments, and run-on sentences. They also describe that the differences between the oral and spoken production are minor, and that they find similar patterns in both forms:

(54) “Det man inte behärskar i skrift behärskar man sällan heller i talet.” (Brita)

My translation:

(54) "What you do not master in writing you rarely master in speech either." (Brita)

When asked about the diversity of proficiency found in the student group, Brita explains that the groups in the course used to be more diverse, but that the introduction of an IELTS requirement has leveled things out to some extent in the current one:

(55) "Det har varit en större spridning tidigare än nu. Sedan vi införde IELTS test som krav för kursen är det mindre, men det är fortfarande en stor spridning.” (Brita)
My translation:

(55) "There has been a greater spread before than now. Since we introduced the IELTS test as a requirement for the course, it is less, but it is still a large spread." (Brita)

Brita sees the tool as being an interesting way of performing diagnostics on a group level, and suggests an implementation during the early stages of a course so that the information can be used for planning purposes. The information could then also be shared with the other instructors involved in other parts of the program:

(56) "Det hade varit spännande att använda verktyget som diagnos i början av kursen och sedan följa utvecklingen för att planera lite mer strategiskt. Det skulle man också kunna kommunicera till den som undervisar i grammatik." (Brita)

My translation:

(56) "It would have been exciting to use the tool as a diagnosis at the beginning of the course and then follow the development to plan a little more strategically. You could also communicate that to the person who teaches grammar." (Brita)

The idea of following the language development through check-ins with the tool is an interesting take on a use case, especially in combination with communicating the results amongst the other teachers involved with the same group of students. This could take the form of having several reports made using the tool at specific points in time during the run of the course:

(57) "Man skulle kunna tänka sig att man ger, eller skapar, ett diagnostiskt prov i början av terminen och sedan använder verktyget för att få fram en individuell profil eller en profil på gruppnivå. Sedan kan man dra slutsatser för undervisningen. [...] Mot mitten av terminen kan man ju göra en ny analys och kolla på hur texterna har utvecklats på individnivå och se vad som har hänt. Sedan på slutet kan det ju också vara kul att se vad som har hänt." (Brita)

My translation:

(57) "One could imagine giving, or creating, a diagnostic test at the beginning of the semester and then using the tool to produce
an individual profile or a group-level profile. Then conclusions can be drawn for teaching. [...] Towards the middle of the semester, you can do a new analysis and check how the texts have developed on an individual level and see what has happened. Then at the end it could also be fun to see what has happened." (Brita)

In terms of functionality, Brita thinks that the tool has indicated relevant categories in the output, and that it seems to be functional in that regard. However, Brita also emphasizes the importance of manually controlling the results and not trusting the analysis too much:

(58) "Jag tycker att verktyget ringar in de delarna här som är relevanta. Men man måste ju också dubbelkolla eftersom verktyget är grovkornigt. Man kan inte tro analyserna fullt ut, inte lita på dem." (Brita)

My translation:

(58) "I think the tool circles the parts here that are relevant. But you also have to double check because the tool is coarse-grained. You cannot fully believe the analyses, not trust them." (Brita)

The main issue Brita sees in the results provided by the tool are to do with the use of American English as the target variety. This has to do with the validity of other varieties, especially when working with students who come from areas with their own second language varieties of English:

(59) "Sen är det ju en brist att det är en amerikansk engelska. Utifrån ett andraspråksperspektiv är det ju beklagligt att man bara har amerikansk engelska." (Brita)

My translation:

(59) "Then it is a shortcoming that it is American English. From a second language perspective, it is regrettable that you only have American English." (Brita)

While the reason for using American English as the target variety in the current version of the tool was motivated by the specifications of the TextStat resource, the point highlighted by Brita is incredibly important if the tool is to see widespread usage.
In general, Brita believes the tool could become useful for their teaching practice. They also point out that the tool could fulfill a function in terms of equalizing assessment:

(60) "Det har ju att göra med testvaliditet. Det kan ju hända att man är trött när man läser en text och då kan tycka att texten är lite bättre än den egentligen är. Det här är ju ett mer objektivt verktyg."

[...]

"Jag tänker på interbedömmarreiabilitet, och då kan ju det här vara ett sätt att få upp ögonen för det man gör." (Brita)

My translation:

(60) "It has to do with test validity, after all. It can happen that you are tired when you read a text and then you can think that the text is a little better than it really is. This is a more objective tool."

[...]

"I am thinking about inter-rater reliability, and then this can be a way to open your eyes to what you are doing." (Brita)

Brita describes the routine regarding student text interactions as students submitting their texts and then receiving feedback. While the students normally do not revise and re-submit their texts, they write 3 texts over the course period and are expected to consider the feedback of the previous texts when submitting a new one. During the semester against which the tool is being evaluated here, the students were allowed to revise their first text as Brita wanted them to get into the writing properly.

(61) "Det är ju flera uppgifter, så de får kontinuerlig feedback. De gör 3 uppgifter och får sedan en slutbedömning." (Brita)

My translation:

(61) "After all, there are several tasks, so they get continuous feedback. They do 3 tasks and then get a final assessment." (Brita)
The routine surrounding the 3 submissions described by Brita, and visualized in Figure 38, includes different routes of communication regarding the submitted text:

![Diagram showing communication routes](image)

**Figure 38: Brita’s respondent drawing**

As indicated by Brita’s respondent drawing above (Fig. 38), there is sometimes also communication directly between the teacher and the students regarding the feedback given on the texts. These exchanges take place on Moodle, which is sometimes an issue for the students:

(62) "Jag använder MyMoodle. Jag tycker att den fungerar väldigt bra eftersom jag behärskar den, men jag vet att studenterna har lite problem och kan ha svårt att hitta saker.”  
(Brita)

My translation:

(62) "I use MyMoodle. I think it works very well because I have mastered it, but I know the students have a bit of a problem and can have a hard time finding things.” (Brita)
Brita does not explicitly instruct the students to make use of digital support when writing, but some of them do. Brita does remind them to make use of the supporting features included in Word:

(63) "I viss utsträckning gör de det. Jag får påminna dem om det. Det tar ett tag innan de förstår hur verktygen fungerar." (Brita)

My translation:

(63) “To a certain extent they do. I have to remind them of that. It takes a while for them to understand how the tools work.” (Brita)

When asked if they believed the tool would influence their workload, Brita indicated that the initial period of getting acquainted with the tool would likely contain a higher workload than usual, but that the practices would adapt to its inclusion over time.

(64) "Det är svårt att säga. Jag kan tänka mig att det i början blir mer jobb och man kanske inte vet hur man ska jobba med det. Då kan det ta lite längre tid. Men sen kan jag tänka mig att min bedömningspraktik eller sätt att jobba förändras och man lägger lite mindre tid på något annat. Värden är nog mer att man får ett annat perspektiv på texterna." (Brita)

My translation:

(64) "It is hard to say. I can imagine that in the beginning it will be more work and you may not know how to work with it. Then it may take a little longer. But then I can imagine that my assessment practice or way of working changes and you spend a little less time on something else. The value is probably more that you get a different perspective on the texts." (Brita)

The main value of the tool identified by Brita is that it provides the teacher with an additional perspective on the student texts, rather than it saving time. As indicated previously, the main time sink of the tool in its current form is the need for manual inspection of the results. However, Brita still thinks there are advantages in using the tool, despite the identified shortcomings:

(65) "Man måste ju dubbelkolla. Man kan inte lita på analyserna, och där är det ju ett problem. Sedan detta med hur man ställer sig
till att amerikansk engelska är normen, det blir ju en begränsning. Utöver det ser jag fördelar med det.” (Brita)

My translation:

(65) "You have to double-check. You cannot trust the analyses, and that is a problem. Then this with how you feel about that American English is the norm, it becomes a limitation. Beyond that, I see advantages in it.” (Brita)

These issues also influence how Brita evaluates the fitting use cases for the tool. The main points made are that the tool could be used beneficially for formative assessment, but that the limitations in terms of precision and language variety make it inappropriate for summative usage:

(66) "Man kan ju inte använda det här verktyget i summativt syfte, då hamnar man nog fel. Jag tänker mig att det fungerar i ett formativt syfte, att man kan använda det för att se var studenterna befinner sig.” (Brita)

My translation:

(66) “You cannot use this tool for a summative purpose, then you will probably end up wrong. I imagine that it works for a formative purpose, that you can use it to see where the students are.” (Brita)

In summary, the interview with Brita indicated that the tool would be a better fit for formative assessment than for any summative efforts. The main utility seen by Brita is as a diagnostics tool at the beginning of a course, with additional usage at checkpoints throughout the semester.

As in previous interviews, Brita finds it likely that implementing the tool would result in a higher workload at the start of usage, but that benefits could be had long term. This would necessitate further development of the tool, and the inclusion of a wider spectrum of language varieties. However, the tool seems to fit well with the current routines of Brita’s teaching, although they believe that implementation would require some changes to the amount of time spent on different aspects of their practice.

6.2.3 SUS and SF UES at University

The SUS and SF UES forms were administered anonymously for tertiary respondents, as they were for upper-secondary respondents, and will thus also be presented under the respondent 1 and respondent 2 labels.
Respondent 1 returned a SUS score of 47.5, which corresponds to a letter grade of “F” and an adjective score of “Unsatisfactory”. Respondent 2 scored the tool at 65 in the SUS form, giving it an adjective score of “Poor” and a letter grade of “F”. As such, the general user experience for the tertiary respondents, who evaluated the tool reflectively without extended usage, is lower than the scores provided by the upper-secondary respondents.

Moving on to the SF UES scores, the total scores were 2.08 for respondent 1 and 4.08 for respondent 2. This would indicate neutral to negative for respondent 1 and neutral to positive for respondent 2. Looking at the breakdown of factors seen in Table 15, a similar distribution of the scores amongst the factors as seen in the upper-secondary group is found.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Resp. 1</th>
<th>Resp. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused Attention</td>
<td>1</td>
<td>3.33</td>
</tr>
<tr>
<td>Perceived Usability</td>
<td>3.33</td>
<td>5</td>
</tr>
<tr>
<td>Aesthetic Appeal</td>
<td>1.33</td>
<td>3</td>
</tr>
<tr>
<td>Reward</td>
<td>2.66</td>
<td>5</td>
</tr>
<tr>
<td>SF UES Score:</td>
<td>2.08</td>
<td>4.08</td>
</tr>
</tbody>
</table>

*Table 15: SF UES score per respondent and factor for tertiary respondents*

Respondent 1 has provided a neutral evaluation of the “Perceived Usability”, with negative scores for all other factors. The “Aesthetic Appeal” and “Focused Attention” scores are very low, while the “Reward” factor is scored slightly higher. Respondent 2 has given the “Reward” and “Perceived Usability” the highest score possible, while the remaining factors are scored neutrally. Neither respondent provided a free-form answer.

### 6.2.4 Themes at University

As with the themes for upper-secondary, the substantive codes derived from the meaning units identified in the interviews with the university participants were condensed into themes. It is important to note that this categorization took place after the themes were created for the upper-secondary respondents, meaning that the process is not entirely inductive in nature. The substantive codes from the university participants were also found to match well thematically with the substantive codes from upper-secondary, and the previously mentioned themes were re-used in their entirety. As the testing scenario for the tool at the university level was reflective in nature, and since the tool was not used in actual teaching practice, the theme of ecological incompatibility was not used.
An argument could be had here regarding whether or not this level of fit between the themes from the upper-secondary and the university participants could be considered to indicate a saturation of knowledge, as was argued for the survey interviews. However, the low number of participants in the interventions makes me hesitant to argue for a saturation being achieved, as six participants are suggested as a minimum while the current study only contains four (Guest, Namey and Chen, 2020).

6.2.4.1 Challenges
A major challenge indicated by the university participants was the number of false positives generated by the tool. As with previous evaluations, the main culprit was found to be the whitespace error pattern. In addition, line breaks were shown to cause issues as well. In general, this could be added to the list of challenges and issues that have to do with the precision of the text analysis when applied to student texts.

The time investment needed in order for the participants to see themselves using the tool efficiently was also highlighted as a challenge. This was partially due to lacking terminological transparency (which is further discussed under issues), but also due to the changes needed in order for them to implement the tool in the desired way. Suggestions regarding the implementation from both participants indicated that it could be used formatively at early submission and revision cycles in the courses, but this would also mean that courses would have to contain several cycles.

A challenge that was shown to be compounded by the precision issues was the granularity of the output. As indicated by Brita, the coarse-grained nature of the analysis means that beyond the manual inspection needed to identify and remove false positives, the user would also have to manually go through the output and summarize the content to an appropriate level for their student group. This also contributes to the initial time investment, as the procedure is likely to be more challenging for a user who is not familiar with the tool.

6.2.4.2 Affordances
An affordance of the tool highlighted by Quincy was that the analysis, in broad strokes, matched with the manual assessment of the student texts. However, the analysis also highlighted patterns that both respondents would not, under normal conditions, have picked up on. Much like with the upper-secondary participants, the university participants considered the tool viable as an additional perspective that could complement their own reading of the texts.

Another highlighted affordance of the tool was its use for planning purposes. As indicated by both university participants, the tool was seen as being a good fit for formative assessment in cyclical teaching designs. This would suggest that the tool was understood as a good fit for the teaching routines described by the survey participants before the start of the design.
Finally, an additional affordance within the idea of the tool as a complement to the participants’ practices was the use of it as a validity check in assessment situations. While both participants were skeptical about using the tool for an outright summative purpose, the addition of the overview to the already existing practices could fulfill an important role. However, this would also put emphasis on the user’s role in considering the limitations and issues of the tool when taking the output into account.

6.2.4.3 Issues and Limitations

One of the limitations that would need to be accounted for in situations involving assessment is the single variant of English used as the target by the tool’s analysis, American English. As pointed out by Brita, this does not only cause issues when students might want to write in other inner-circle variants, such as British English, but will also lead to issues regarding other variants that could be desired as the target for a multitude of reasons. In addition to the prescriptive perspective inherent to the language analysis discussed previously, this issue also highlights inclusivity issues due to the narrow target language variant.

Another issue brought up was the lack of terminological transparency in the tool’s output categories. Quincy brought up that they would make use of a set of standard annotations for errors in their feedback comments, that this was not entirely matched by the output of the tool. The use of standard annotation would be a good use case for the tool’s output, but the current annotation used for the output has not been selected with any consideration for student-facing transparency, which is a limitation.

Finally, the aesthetics and the tool’s ability to hold the participants’ focused attention is an issue that creates a limitation in terms of length of engagement. This issue becomes more problematic due to the amount of time needed for manual inspection of output. This was indicated by both the interview materials and the low scores in the SF UES.

6.2.4.4 Future Development

Improving the precision of the tool and translating the annotation into a more transparent system that could also be used for annotation in feedback comments stand out as two developments needed for future versions. This should be considered in addition to the need for a more limited output in order to improve readability. As with the upper-secondary participants, the university participants indicated that the aesthetics need improvement in order for the tool to become usable.

The selection of English variants must also be widened in order to accommodate targets beyond the current single choice. The selection must also be developed towards becoming wide enough to support learners who are not only writing in inner-circle variants, but also writers who want to make use of
other Englishes. While this will not directly change the functionality of the tool, it is a central component needed for the tool to be viable for broader implementation.

6.3 Conclusion of Intervention Studies

The SUS scores for both groups indicated a neutral to negative user experience overall. Looking at the factor breakdown provided by the SF UES, the reason for the poor user experience seems to mainly have to do with the aesthetics of the tool and the ability of the report to focus the teachers’ attention. The upper-secondary interviews indicated that this was mainly an issue experienced while reading the reports, and is likely the result of both the “mechanical” feeling of parsing the data tables and the overwhelming amount of information provided.

The interview with Jamie showcased a desire for the tool to move away from the supporting role imagined in the current dissertation toward the tutor role. The intervention also highlighted the issues that would have to be navigated in such an implementation. The classroom situation described by Jamie showcased the importance of a relational approach to pupils who need a lot of support, and who are vulnerable to developing avoidance issues when given too much (and too harsh) feedback.

The interview also highlighted the report produced by the tool as a source of issues for the teacher, as it was experienced as overwhelming. Jamie indicated that the categories seemed to become more familiar with time, but maintained that some sort of narrowing function should be implemented in future versions. The two versions suggested by Jamie were grade and proficiency-based profiles that would narrow the output depending on selection. Jamie also suggested working towards incorporating grade aims in the profile suggestion and making the tool pupil-facing to a further degree. The tool seemed to fit with Jamie’s routines, and can be categorized as belonging in the roles envisioned in the initial interview study.

Charlie’s interview highlighted the different outcomes of using the tool at different proficiency levels, or indicated proficiency levels through the course designations. While finding the tool functional, Charlie also provided very concrete ideas on how it needs to change in order to make sense within their work environment. The two main suggestions were the introduction of a tiered error categorization and optimization for shorter processing times.

In addition, Charlie described the joint assessment as a use case that should be targeted by the tool, as it could help the teachers in coping with the peak in workload around exam periods. Feedback was also discussed, and the tool was indicated to fit well within Charlie’s pre-existing routines. Charlie also suggested that the tool could fulfill an equalizing function when it comes to
feedback, both over time within the practice of an individual teacher and between different teachers.

Quincy’s interview showcased how the tool could be used to highlight features of student writing that the teacher might not normally pay attention to. However, the issues experienced with line breaks and other components of the Word templates used by their students were experienced as an issue. The terminology used by the categorizations in the report produced by the tool was indicated as likely to contribute to a peak in workload when first implementing it.

Brita indicated issues similar to the other interviewees in terms of expecting a heightened workload early on in an implementation. In addition, Brita highlighted the importance of applying a critical perspective on when and where to make use of the tool, deeming it useful as a support for formative assessment.

The interview with Brita also highlighted the extended perspective on issues with limiting the language varieties processable by the tool. Not only does this result in a high number of false positives, as attested by all respondents, but it also creates an issue through supporting an American English language norm within the educational context.

Some practices specific to the upper-secondary and tertiary levels seem to be found in the debriefing interviews, for instance the importance of oral feedback in upper-secondary and the importance of a wide selection of varieties in tertiary. In general, however, there seems to be a consensus regarding the evaluation of the tool.

The weight given to manual inspection is a workload issue amongst all of the respondents. The same can be said for the high initial workload resulting from the implementation of the tool. A majority of the respondents would also like the tool to more explicitly support the choice of a focal area for individual texts as a way to fit in with their current approach to assessment.

Both of the upper-secondary respondents, who used the tool for a longer period, explicitly mentioned issues with the readability of the resulting report. This indicates that this is an issue that is likely to become more apparent with time, and might cause problems before arriving at the long-term benefits mentioned by the tertiary level respondents.
7. Results

This section will present a summary of the results obtained in the current dissertation. The results are presented thematically, with the main themes being Swedish EAL Teachers’ Diagnostic Needs, Tool Development, and the Intervention Study. The intention is for this section to summarize the main takeaway from the different sections in a concise manner, and to act as a summarizing step before the dissertation moves on into the discussion.

7.1 Swedish EAL Teachers’ Diagnostic Needs

The initial survey concerning the diagnostic needs of in-service EAL teachers in Sweden was done using an interview instrument and a respondent drawing. The interview questions were aimed at understanding which features of student texts teachers often provided comments and feedback on, the routines of their interactions with student texts, and their experiences of using digital supporting tools. The respondent drawing was included as a way for teachers to visualize their text interaction routines and would be used to inform the tool design about which routines the tool would have to fit within. The survey had two pilot sets, which were used to refine the interview instruments, and a main set consisting of four in-service EAL teachers in Sweden.

The results indicated that any automated attempt at supporting diagnostics within the school context would have to be able to identify a series of error patterns, such as sentence structure, punctuation, verb tenses, subject-verb agreement, article usage, and spelling. In addition, the tool would have to fit within an already established practice aimed at providing learners with the necessary supports for them to succeed, while also avoiding pitfalls such as creating avoidance issues or in other ways overwhelming the students. This emphasized the importance of the teacher as a facilitator of the tool, as the tool itself would not be able to produce these considerations regarding the students.

In terms of second language acquisition models, the choice was made to view this interaction between teacher and student from the perspective of the social constructivist model, using the concept of scaffolding as a blueprint for how the tool could be applied. The model was selected due to the interpersonal aspect of text interactions found in the in-service teacher survey conducted at the beginning of the study. Other models were also considered, as they related to the backgrounds of the later sections and since the tool would have to remain functional from the perspectives of other models due to the implementation depending on teacher preference.
7.2 Tool Development

The error patterns identified by the participants of the in-service teacher survey were used as a basis for resource selection at the beginning of the tool design process. The previous research on corrective feedback indicated that an important criterion for a beneficial implementation was that the targets of the feedback were rule-based phenomena, and that the learners were aware of the rule governing their use. The results of the in-service teacher survey matched well with this criterion, which indicated that the previous international research included in the background could be considered to have a level of ecological validity in the Swedish context.

The research on HCI indicated ease of use and perceived usability as predictors for teachers actually making use of the tool, meaning that they would have to be at the center of the design process. In addition, participants were explicitly informed about their responses being directly used in the design of the tool as a way of making use of participatory design, i.e., including the target user base in the design.

Based on the previous research on CALL, the resources chosen for the language analysis functionality of the tool would have to be precise to such a level that they could provide accurate data on student texts. The resource selected for the language analysis was the Java LanguageTool, which provides a lexicon-based approach to identifying error patterns in text. The lexicon provided for the English language was a good fit with the patterns indicated as relevant by the participants in the in-service teacher survey. The Java LanguageTool also stood out as a good option due to the possibility of using a local server, rather than sending texts over an API online. This is important due to the sensitive nature of student texts, and the possible breaches of GDPR resulting from distributing the text somewhere other than the teachers’ local machines.

The Java LanguageTool also provides elaborating comments on the identified patterns, making the output more easily understood. While the comments are usually not fit for direct distribution to the language learners, they do serve a transparency purpose for the teacher. In addition, as indicated by the previous research on HCI, the interaction between a human and a computer goes both ways, meaning that the existence of a comment field in the output might influence the user towards commenting on error patterns to a higher extent. The Java LanguageTool comment can serve as a template to work from and elaborate on in those situations.

TextStat was used to provide data regarding the complexity of the student texts. Complexity was chosen as a relevant metric due to complexity being explicitly mentioned amongst the goals of the higher courses in English at the upper-secondary level, and since it has been connected to the formality of the register. Register was explicitly mentioned in the in-service teacher survey as
something that would be of use. TextStat was chosen due to providing a consensus score of several different types of readability/complexity measurements, and thus being less dependent on single variables within the text.

Dash was chosen for the task of visualizing the reports for the teachers. The reasoning here was that Dash, like the Java LanguageTool, can be used from a local server on the teacher’s machine and does not necessitate sending student texts across the web. In addition, Dash displays its output in the web browser, which is likely to be a more familiar setting for the teacher than creating a new environment for the results. Dash also included the functionality needed for teachers to be able to export their results to Excel sheets or download visualizations to their computers. Depending on the browser being used, the entire report could also be exported as a PDF.

The dashboard visualizations created for the report produced from the analysis were chosen based on areas indicated as interesting by the in-service teacher survey. The report was split into two sections, focusing on group-level and individual-level overviews respectively. This was intended to support the teacher in tasks concerning lesson planning and individual feedback.

Once built, the tool was tested in a controlled environment created through web-scraping content from the online student text repository Mimers Brunn. The results indicated that the tool worked better when being used to analyze more formal materials.

At this point, the tool was used for a miniature pilot study during my practice placement at an upper-secondary school. The use of the tool in practice indicated that the analysis of student texts resulted in very noisy data, with many of the error patterns not being relevant to the teacher. As a result, the tool underwent a second design iteration, and the ability to filter error pattern categories before producing reports was implemented, and the most relevant categories as indicated by the survey were set as the default selection. In addition, a guiding document was included in the design and would be opened when the tool generated the report by default.

The tool was then subjected to user testing. The pilot test indicated that familiarity with corpus linguistics or computational linguistics was a predictor for benefitting from the tool’s output and rating the user experience as enjoyable, while English language subject teaching students rated the user experience as neutral to positive, with lower scores for “Aesthetic Appeal” and “Focused Attention”. When carrying out the user tests after conducting the pilot, three EAL teachers were recruited. The results indicated that the teachers could see benefits in using the tool, indicated by the “Reward” factor being scored high, while the scores that had to do with the interface of the tool were scored lower. One of the participants scored all categories the same, and gave high scores to all components of the user experience.

The major limitation of the chosen tool design is the use of American English as the target variant. This is due to the TextStat resource used specifying
American English as the target variant for its analysis. In addition, the Java LanguageTool was shown to be prone to false positives when analyzing Word templates containing line breaks or sub-headings.

7.3 Intervention Studies

The intervention studies were conducted with 2 respondents from the upper-secondary level EAL courses and 2 respondents from the tertiary level. The tertiary level respondents were evaluating the tool based on their experiences from teaching a course for exchange students at a level corresponding to the English 6 course at upper-secondary. The upper-secondary respondents made use of the tool for a full semester, while the tertiary level respondents evaluated the tool reflectively at the end of the semester without having applied it in their actual teaching. This was due to the increased workload and the general instability of the course brought on by COVID-19. The interview instrument used in the initial survey was adapted for the tool evaluation, and the same respondent drawing was used. After the interview, the user testing suite of questionnaires was administered digitally.

The user testing results indicate a stronger outline of the pattern seen in the first round of user tests, where the aesthetics and focus afforded by the tool score poorly, while the functionality and reward factors score higher. In general, the results indicate a neutral to positive experience of the functionality, with a neutral to negative experience of the visual interface.

The themes found in the post-intervention interviews were challenges, affordances, issues and limitations, ecological incompatibility, and future development. Ecological incompatibility as a theme was not supported in the university level interventions as the tool was only used reflectively and was not included in the actual teaching practice.

Challenges indicated by the upper-secondary respondents reported the output of the tool as being overwhelming and suggested tiering the output according to either user profiles or error significance. A majority of respondents also reported the readability of the output tables as an issue, stemming both from the experience of parsing a table, the number of false positives, and the terminology used to categorize the error patterns.

Amongst the affordances of the tool, the results indicated that it had achieved a good fit with the respondents’ routines, which were similar in nature to the routines found in the results of the first survey. The tool was characterized as a “second opinion” on the assessment of student texts, which was considered beneficial due to it supporting the teachers in noticing things they could otherwise have missed. Two participants also highlighted the equalizing of feedback provided by the tool, as their assessments would not be as influenced by things like tiredness or fatigue.
Issues and limitations found were that the majority of the respondents agreed that the tool would be a good fit for formative assessment efforts, but was too unstable to be used in a summative manner. The amount of manual sorting and pruning needed was mentioned as an issue by all respondents. Additionally, the lack of any other language variants than American English resulted in many error patterns that were not immediately relevant to the teachers, and respondents indicated that it could risk enforcing an American English norm in second language education long term. All respondents indicated that the implementation of the tool would likely result in a heightened workload for a time, but they were positive about the long-term effects. A majority of the respondents expressed a desire for the tool to further support them in clearly identifying focus areas for their planning and feedback.

Ecological incompatibility was mainly found related to the limitation on file format in the tool. Software or digital solutions used to limit the students’ access to the internet or other functionalities of their computer during testing situations could not be guaranteed to provide .docx files, as seen with Inspera. While the participant who encountered this incompatibility was able to solve it through third-party tools for format change, it should be noted as a shortcoming of the current design. The tool was also found to be incompatible with certain security measures, and required both contact with the school in question’s IT staff and alterations to the resources used for document parsing in order to function at one participant’s work laptop.

Future developments needed based on the intervention studies include a broadening of the target language, a rework of the aesthetics, and the implementation of an output limitation of some kind. Broadening the target language means including variants beyond American English, either through selection or as a broader scope of the error pattern analysis in general.

Reworking the aesthetics is indicated as being most needed concerning the interface with the tabular data. As this interface was highlighted as the most unpleasant to use while also being needed for the most time-intensive aspect of the use, i.e., manual inspection, an overhaul is needed. This should be done in combination with some kind of output limitation that would serve to lower the amount of information presented. Suggestions from the participants included user profiles, level profiles or automatically implemented focal area identification.

The issue of false positives was shown to exacerbate the overwhelming nature of the output in two ways. Firstly, the number of false positives contributes to the amount of data being communicated to the user. Secondly, it forces the user to spend more time manually inspecting and pruning the results, thus making longer contact with the tabular data necessary. As such, further development of the language analysis resources would benefit the tool not only functionally, but also through mitigating the user experience of the interface.
8. Discussion

This section will first discuss the results in the context of the previous research and the intentions behind the tool design. In addition, the ecological validity of the tool will be assessed based on the intervention study and the possible outcomes of the use of the tool will be assessed through the perspective afforded by Educational Linguistics. Following the section on the interpretation of results, the methodological considerations of the dissertation will be discussed. The discussion then moves on to the limitations of the current study. Finally, the lessons learned throughout the development and evaluation of the tool are summarized. The intention is for this final section to function as a list of concrete suggestions for future efforts at automating text diagnostics within similar contexts to consider.

8.1 Interpretation of Results

The previous research on corrective feedback indicated that in order for it to be successful it had to fulfill three criteria: (1) That students are focused on the importance of editing, (2) are trained to identify and correct patterns\(^{30}\) of frequent and serious errors and (3) have been given explicit teaching on the grammatical rules relevant for those errors (Ferris, 1999, p. 5). This should be seen as the baseline of necessary actions in which the tool was intended to provide teachers with support towards a best practice of corrective feedback.

The results of the initial in-service teacher survey indicated that the backbones of such a system were already in place amongst the respondents. There was a cyclical pattern of feedback where students were expected to perform self-revision and re-submit their revised texts for assessment. Self-revision is also mentioned among the predictors of successful corrective feedback, and as an important factor in acquiring new language features (Ferris, 2010, p.189).

The delivery of the corrective feedback was seen to be a deciding factor in the efficiency of it, which is where scaffolding would become an important consideration in the application of corrective feedback in the educational context. Scaffolding should take the form of the teacher facilitating the learning experience in such a way that the learner is supported towards achieving the furthest reach of development given prior knowledge (Wood, Bruner and Ross, 1976). This scaffolding does not concern only the acquisition of language, but the entire learning experience. This aligns with previous research regarding the

\(^{30}\) Bold for emphasis in original
need for broader considerations surrounding the use of automated writing assessment for the purposes of formative feedback (Ware, 2018).

Scaffolding, coaching, and feedback were already actively being applied by the respondents of the initial survey, although the results indicated that there was a desire for more time and resources going towards it than what was currently possible. Especially 1-on-1 oral feedback sessions with individual students were mentioned as being effective, but difficult to manage due to the time constraints imposed by the teachers’ other tasks. The workload of the teachers influencing them away from what they consider best practice was touched on in the research on corrective feedback as well, and is presented as one of the main reasons as to why teachers do not provide comprehensible feedback on grammatical error patterns in student writing (Ferris, 1999, pp. 6-7). My interpretation of these results is that the feedback routines desired by the teachers align with the best practice described by the research, but that their workplace environments do not provide the opportunity to apply those routines in practice.

Regarding the target features of corrective feedback, the previous research stated that the most beneficial features to provide such feedback on were rule-governed patterns. Examples of such patterns are subject-verb agreement, run-ons, comma splices, missing or incorrect articles, and verb forms (Ferris 1999, p. 6). These examples match exceptionally well with the target patterns suggested by the results of the in-service teacher survey, which were sentence structure, punctuation, verb tenses, subject-verb agreement, article usage, and spelling.

The survey also provided substantive codes relating to the application of the tool. These were to do with avoidance strategies, overwhelming the students, self-revision, and feedback. Seen from the perspective of the social constructivist model of SLA, these topics indicate a need for facilitation by the teacher within the writing instruction. Overwhelming the students would be an indication of the materials and interactions over-extending beyond the ZPD of the learner and rendering the exchange incomprehensible to them. The tool must thus allow for the design of interactions that lead the learners toward understanding the information they are given, as suggested by Long (1980, 1991). The interactions must also be operationalized through a socially supportive surrounding, enabling the learners to engage meaningfully with their texts to the extent they can, given their current ZPD. If successfully facilitated, such an interaction would open up for self-revision and feedback sessions that result in language development. If poorly facilitated, or not facilitated at all, the interaction and feedback would result in avoidance issues (Truscott, 1996, p. 333). The results of the survey seem to indicate that the respondents are balancing between beneficial and detrimental corrective feedback in their practice, and are making use of scaffolding to shift the balance towards the positive. The results also indicate that the respondents have a clear idea of how
they would like to shift the weight further towards a positive result for the learners, but are, as previously mentioned, held back from doing so by their workload and time constraints.

Considering here the placement of the tool within the routines described by the survey participants, they were all aimed at providing feedback on drafts for the purposes of language development. The respondents described their reading of materials as first skimming for content and then investing more time and effort into a second reading aimed at providing comments and feedback on specific error patterns. In order to maintain the scaffolding structure and the teachers’ interactions with their students, the tool was designed to support the second reading that focused on form. From an L2 Writing perspective, the tool falls in the category of “focus-on-text”, meaning that it can only focus on the text as a product, since it has no understanding of the human aspects of the student author (Hyland, 2016, p.48).

The routine described by the interview participants could be summarized as a cyclical interaction where texts were created by the learners and then received feedback from the teachers before being revised and then re-submitted. This feedback routine is centered on the idea of scaffolding the learner through the writing process, with incremental improvements in accordance with the ZPD or i+1, i.e., the appropriate level for the learner engaged with through a modified interaction (Long, 1991). The interaction, as far as that specific text would be concerned, would end on a final submission, which would then inform the teachers’ summative assessment of the students’ production.

However, the assessment of a student paper is not a strictly summative affair. As indicated in the background section on EAL education in the Swedish upper-secondary system, the students are also intended to develop abilities shown during the writing process. The central one to the topic of the current dissertation was the ability to revise their own text production (Skolverket, 2012b). Beyond being one of the results of the teacher survey, self-revision was one of Ferris’ requirements for beneficial corrective feedback (2010), and its inclusion in teacher routines could be considered a pre-requisite for a tool such as the one designed in the current dissertation to be able to function as intended. In theory, the results of the teacher survey indicated that automated focus-on-form corrective feedback could be successfully implemented, as the routines described by the teachers provided the scaffolding needed for students to benefit from such diagnostics. The teacher would, however, remain the facilitator ensuring that feedback was delivered in such a way that it would lead to students producing the language independently, as seen in Hammond and Gibbons (2001).

Looking at this placement of the tool in the teaching practice from the perspective of the SAMR model, the tool would fit within the augmentation category as an enhancement of existing practices (Puente Dura, 2010). It is important to note that the SAMR model was initially intended for categorizing
the integration of digital technologies in learning, but the designations offered by the model can also help in conceptualizing the integration of digital teaching tools. The category of augmentations indicates that a technology is intended to act as a direct substitute for an existing practice, while also providing some improvement to the outcome.

After the tool had been designed, initial tests were conducted in order to gain some understanding of how it would perform in a real scenario. This is where the connection to the ecological validity of an experimentally designed tool becomes the most apparent, and the most problematic. In essence, the initial tests conducted with the tool were carried out in an environment that differed from the one into which it would later be implemented, which goes against the idea of simulating reality within experimental design (Schmuckler, 2001).

The results of the initial test in the controlled environment indicated that the formality of the genre was a good predictor for the performance of the tool. This makes sense, as the strict ruleset applied by the Java LanguageTool results in a very prescriptive approach to language correctness. Genres like essays and reports produced usable results, while poetry and lyrics did not. The main takeaway from the tests in the Mimer EAL corpus is that teachers using the tool should target a formal language register in their student production in order to produce usable results. This would also mean that the tool would not be fit for use with student groups who are not at a proficiency level where the formal register is being applied in written assignments, as this could create a mismatch in regards to their ZPD. In addition to the issue of formality, the Java LanguageTool necessitates selecting a variety of English in order to provide the text analysis.

As I had been provided a practice placement as a part of my teacher education, a miniature field test was conducted in order to test the tool in a real-life context. The results indicated that the tool needed some kind of sorting mechanism when creating the report, as the noise found in the data made it difficult to identify relevant patterns. A pattern selection screen and an instruction document were added to the design in response to the experience.

The pilot test was used to ensure that the instrument used to distribute the SUS and SF UES forms functioned as intended, and the results returned the conclusion that the tool was more likely to be perceived as useful by users with experience within corpus linguistics. Furthermore, the pilot tests indicated that the functionality and benefits of using the tool were scored positively, while the appearance and the ability to focus on the results afforded by the tool were scored negatively.

In the free-form answers, the respondents of the pilot user test indicated that the main issue they experienced with the tool had to do with the setup procedure and the local installation. Since the use of a local installation was motivated by the GDPR regulations governing the spreading of students’ written materials, these results were difficult to do anything about given the limitations of the
current project. A setup capable of handling the distribution of student texts from the teacher’s computer to a GDPR-compliant server and then returning the information would require the involvement of IT and legal professionals, and is beyond the scope and resources of the current project.

Since the free-form answers provided interesting results, and since the setup was experienced as being difficult for the respondents, the user test was changed for the next set of respondents. The tests made use of the same questionnaires, but the respondents were given access to a completed install of the tool on a virtual machine, and were joined by the author during their time with the tool. This was done in order to provide immediate technical support. This also allowed for the use of a think-aloud protocol to be used during testing.

The second set of user tests returned results similar to the ones produced by the pilot set in terms of the user experiences of the different components of the tool. The aesthetics and ability to provide focus were ranked neutrally, while the functionality was ranked positively. The think-aloud protocol indicated that the respondents were considering the usage of the tool along the lines of the implementation it had been designed for, i.e., as a way of limiting the workload when performing readings of student texts with a focus on form. The respondents noted that the tool could enable them to notice issues that they would normally not have engaged with, and it was experienced as a way of extending the grammar teaching provided by the textbooks used. The respondents also stated concerns regarding the amount of time that would be needed to get acquainted with the tool to such an extent that it would benefit them.

The challenges highlighted by Jamie surrounding the use of feedback highlight the importance of the teacher as a facilitator and engineer of interactions where students have the ability to develop their new language (Brown, 2000, p.87). The situation described in the interviews mimics the balancing act seen in the initial interviews, as participants leveraged their interpersonal relationship with the students and their understanding of the social factors influencing the student’s experience to provide the correct amount of feedback in a comprehensible way. The emphasis on comprehension here can be taken as an operationalization of the student’s ZPD, and the teacher is actively engineering a situation based on what they have assessed as appropriate for the student through modified interaction, as suggested by Long (1980, 1991). The application of the tool in this task, albeit challenging, is indicated in the interviews as actively contributing to the teachers’ scaffolding of the students’ acquisition (Wood, Bruner and Ross, 1976).

Amongst the affordances, the tool was seen as a support for identifying the “basic” grammar issues, which would allow the teacher to focus more on the aspects of an assignment that a computer cannot comprehend, such as content. It also opens up for teachers to apply a full analysis despite the limited time available, answering to the needs described in Ferris (1999, p.6). This result is
interesting as it indicates the tool functioning well, and a shift being possible for the teacher due to the support provided by it. The previous research on HCI indicated that an influence on the teacher’s reading was likely to appear through the interaction between user and computer, but also described the computer “steering” the user's behavior towards what would feel intuitive in the situation (Nielsen, 1994). Jamie’s shift towards focusing more on content when provided with the tool is thus an interesting find. The routine in place for student text production is similar to the ones seen in the initial survey, and conforms to the criteria for beneficial corrective feedback as they included both training in error recognition and self-revision (Ferris, 2010; Oscarson, 2009), while also aligning with self-revision as an aim for the upper-secondary English subject courses (Skolverket, 2012b).

Extending these results further through automating the response given to the student brings the conceptualization of a diagnostics tool full circle, and returns to the “computer as tutor” described by Johns (1991). While the text analysis software needed for such an implementation exists today, the social aspects of the learner experience indicate that fully automating a tutoring program might not be the optimal implementation for a classroom context. Furthermore, such applications are available in the form of, for instance, Grammarly31, albeit with a broader intended user group.

Looking at the use of the tool from an SLA perspective, the output can be seen as useful for formal rule learning based on self-produced examples. The use of self-produced examples becomes important here, as Ellis (2003) indicated that a language learner moves from exemplar-based language production towards rule-based. The use of self-produced examples opens up for this to be combined with self-revision, indicated as beneficial for converting technical knowledge of grammar into procedural knowledge (Ferris, 1999). Roughly speaking, this could be understood as a way for the student’s monitor function to acquire and apply new patterns for rule-based language phenomena, as described by Gregg (1984). However, it would not align with Krashen’s (1981a) original concept, as it necessitates overlap between formal learning and acquisition.

Another affordance of the tool indicated by the participants was the equalizing effect the use of it could have on the feedback received by students. The standardization of both recognized patterns and, to some extent, feedback format could contribute to a more stable learning experience for the students. Brita also suggested that the tool could be beneficial for inter-rater reliability, which supports the equalizing function of the tool from a more summative perspective.

Taking a corpus linguistic perspective on how the tool was applied in practice, or reflectively on previous practice, the separation between corpus-
informed, corpus-based, and corpus-driven approaches provides a practice-oriented perspective on corpora in language assessment (Callies, Díez-Bedmar and Zaytseva, 2014). Corpus-based approaches test previous descriptions with the corpus acting as evidence, corpus-driven approaches make use of patterns appearing in the corpus in order to formulate descriptions, and corpus-informed approaches make use of the corpus to inform or validate human decisions (Callies, Díez-Bedmar and Zaytseva, 2014, pp. 74–75). Callies, Díez-Bedmar and Zaytseva (2014) also show that these approaches may overlap in some cases. The use of the tool, as described by the intervention respondents, could be seen as such a case of overlap. The current version of the tool could be seen as corpus-driven, as it simply summarizes the error patterns discovered. However, as empirical evidence connected to the CEFR proficiency levels starts to amass and the features of the different proficiency levels become more clearly defined, future versions of the tool could come to include corpus-informed tiering. In fact, Callies, Díez-Bedmar and Zaytseva’s summary of the corpus-informed approach discusses the use of such applications for validation of human ratings, which is the use-case described by the respondents of this study (2014, p. 76).

Moving on to the issues and limitations, the main issue experienced implementing the tool within existing practices has been the workload needed to manually sift through the results in order to find the relevant patterns. The report provided by the tool was experienced as overwhelming, and aesthetically unpleasant. This is concerning, as it could negatively impact both intention to use, which is a strong predictor for actual usage (Teo and van Schaik, 2012), and the experienced time investment needed. The time investment is important here as the time needed for a focus-on-form reading was indicated as a common reason for teachers not to engage in corrective feedback (Ferris, 1999). The overwhelming reports described by the participants offer a possible explanation for the score distribution seen in the user tests, where the ability to focus on the results and the aesthetics were scored lower than the functionality of the tool. An overwhelming experience might very well turn teachers away from using the tool for a long enough period to develop comfort with it.

The SUS and SF UES scores provided by the participants in both sets follow the general trend seen in the earlier user tests. The functionality of the tool was assessed as neutral to positive, and the aesthetics and ability to focus as poor. It is interesting to note that the tool was scored lower overall in the SUS score in the intervention study as compared to the earlier user tests. It seems that the experience of the aesthetics and focused attention become worse over time.

The tool was experienced as functioning better at the higher-level courses, which corresponds with the results of the initial testing indicating that the formality and register of the writing was an important factor for the tool’s performance. This could also be connected to the overwhelming reports, as the
higher proficiency level in the later courses would likely result in fewer data points in the final output.

It is interesting to note that the main complaints were to do with the output, rather than with the interaction with the functions of the tool. While Römer was mainly discussing the collaborative use of corpora for concordance analysis, the wish list item concerning concordance programs which could provide “...with just one or two mouse clicks, some kind of advanced structuring of concordances [...] for instance to ensure quick access to relevant examples that show typical collocations, or to extract common patterns and highlight the different senses of a polysemous item” (2006, p. 127) could be considered emphasized by the results of the current dissertation. Engagement with a fairly complex linguistic output, in this case error patterns, seems achievable in teaching practice, provided that the user interaction leading there is experienced as simple enough. Whether or not this tool would also aid in the “missionary work” of including corpora in language education would still need to be explored, but it could be argued that the teachers would build confidence in other types of DDL methods over time, and that the adoption of one tool would make them more likely to adopt others in the future.

One important point raised by Brita is the norm-enforcing function of the use of a single variety of English as the target for the tool, in this case American English. This creates a major problem with the implementation, as the tool could very well end up as a gatekeeper in terms of language variants, and give rise to issues similar to the ones described by Spolsky, i.e., gatekeeping and a narrowed perspective on ‘correct’ language variants (2008, p. 3).

The issue of norms in learner corpus research was recently explored by Gilquin (2022), who discussed three possible factors according to which norms could be described: regional variety, level of literacy, and text type. While Gilquin (2022) is focused on learner corpora used for comparisons, the context of use in SLA makes the discussion relevant for the current dissertation as well.

Concerning regional varieties, Gilquin (2022) suggests focusing on the normative use relevant for the target variety indicated by the students themselves. In the comparative approach discussed in her article this makes sense as it would be considered the most relevant comparison, and I would argue that the same is true when using normative use to identify possible error patterns. This further supports Brita’s position of adding more varieties being an important feature for the future development of the tool.

Text type and level of literacy are both interesting perspectives on the issue of the overwhelming output. As was seen in the Mimer corpus testing of the tool, text type is an important predictor of the current tool’s precision. Following the discussion carried out regarding comparative learner corpus use by Gilquin (2022), the variety of text types is likely to become more of an issue as the tool increases in precision, i.e., as a granularity of the detection of relevant patterns increases the importance of considering text type beyond the formal/informal
increases as well. However, following this line of design choices to their full extent is likely to result in a tool similar to Versant (Streeter et al., 2011), which is dependent on the specific instructions and prompt used for a repeated assignment. The level of literacy expected in the student group is central to the idea of tiering the output, but it might not be as simple as identifying the CEFR level, IELTS level or position on the processability path due to literacy being tied to text type and register.

Ecological incompatibilities were found both in terms of the tool interacting with other software employed at the schools, as well as with the digital ecosystem in general. As has been previously described, this had to do with security limitations on one of the participants’ work laptops and the output format from a program used to limit the students’ access to their computers during testing situations.

The obvious solution for the issues of processing time, setup process, and compatibility with IT security limitations would be to move the processing to a server and make the tool entirely accessible through a web browser. However, this creates issues with GDPR compliance and ethics. The magnitude of work needed in order to ethically manage incoming student texts and create a system that is in compliance with GDPR regulations was beyond the scope of the current thesis, but should be considered for future development or future iterations on the concept as a whole.

Beyond the previous suggestion to create a web-based tool in order to mitigate setup issues and improve processing time, other future developments indicated as desirable by the participants were the broadening of the target language and improvements to the precision of the language analysis. The broadening of the target language could either be approached through adding a selector menu from which the teacher can choose which variant to use or through a parallel implementation of several variants where the tool would use the variant-specific patterns or spellings to predict which one would be the best fit for each document. The precision problems had mainly to do with false positives, which could be improved by further limiting the patterns searched for. Certain patterns, such as the double whitespace pattern, could be disregarded completely, while others could be modified. Future development could also focus on building analysis resources specifically for language learners, and implementing the error pattern tiering suggested by Jamie.

Tiering could, in future versions, be conducted through the use of Processability theory (Pienemann, 1998). Processability theory suggests that the human language processing ability follows certain developmental routes, and that the order in which language processes are learned and procedural skills attained can be, at least in part, predicted (Pienemann, 2005, pp. 3–4). By first identifying the current processing level of the student group, the error tiering could then be focused on the next procedural skill in the suggested hierarchy. This would necessitate an initial assessment step conducted by the teacher.
where the current process level of the student group is identified. This could either be done manually, which would assume familiarity with the Processability theory amongst the teachers, or through automating a scoring based on the error pattern categories’ placements within the developmental route and their frequencies. In order to accommodate both group-level and individual-level perspectives, the automated option would have to identify error frequencies in the individual documents, corresponding to individual students, and produce an average for the group.

Moving away from the content analysis themes to discuss the ecological validity based on the interventions, the tool seems to function as predicted in a real-world scenario and the international research seems to have provided a description of practices and routines that is also applicable within the Swedish context. The former could be seen through the match between routines described in the initial survey and in the intervention study, as well as in the descriptions of routines before and after implementing the tool in the upper-secondary interventions.

However, the ecological validity of the initial tests of the tool in the controlled environment can be called into question, as the issues experienced there were found more frequently amongst the intervention participants. This could have something to do with the format of the texts in the Mimers Brunn repository, or possibly with the extra attention given to a text that the author knows that they are going to upload for public display. This is also likely to have contributed to the overwhelming nature of the reports experienced by the participants, as the texts produced in actual school settings might have received less revision by their authors.

The tool’s fit within the routines of the respondents corresponds to the role of an augmentation, according to the SAMR model (Puentedura, 2010), as it was reported as fulfilling an existing function while adding some benefit. This is important, as perceived usability and added benefit are central components of enticing users towards adapting new technologies. This is likely to be especially true in workplace environments with a high workload, such as upper-secondary schools.

Turning to the perspective afforded by Educational Linguistics, the intervention study provides some red flags in terms of the possible outcome of a wide implementation of the tool. While the functionality was mainly perceived as beneficial from a formative standpoint, examples were also given of situations where the tool could be seen as beneficial in summative assessment or grading. Summative assessment creates an issue for the tool as it is, by its nature, a heavily prescriptive approach to language assessment. As seen in the intervention studies, many of the error patterns identified were not considered relevant or major issues by the respondents. The tool, however, does not differentiate, it simply looks at the frequency of occurrence. The tool could, of course, be developed toward these types of use cases.
Using the tool for grading could be accomplished by connecting the error categorization explicitly to the grading policies of the Swedish upper-secondary system. The connection to policy would implement the requirements therein without consideration, and in a very mechanical manner. One issue rising from this is that policies are rarely perfect, as highlighted by the studies on the topic referenced in the background materials on Educational Linguistics (for instance Hornberger, 2001). Add to this the fact that grading policies are political documents, and that any connection to actual second language acquisition research or writing research cannot be guaranteed, especially long-term.

As has been seen throughout this dissertation, student motivation and the students’ learning experience are central components to achieving a functioning language acquisition. The results have also indicated that a majority of respondents use their personal knowledge of the individual students to sort out relevant patterns to provide feedback on, as providing the full assessment at the wrong time would cause avoidance issues. As previously mentioned, an automated tool makes no such considerations.

The use of the tool to support teachers in the assessment of the national exams as suggested by Charlie could be seen as providing a middle-ground, where teachers can view trends in the patterns identified by the tool and use it as a supporting metric for their assessment. Having seen the workload reported by Charlie in relation to the joint assessment of national exams, the use case makes perfect sense. However, as Nielsen (1994) pointed out, the influence in an interaction between human and computer goes both ways. A tool built with the purpose of influencing the grading of exams would have to be constructed quite differently from the tool built for this thesis, with a much higher emphasis on precision. One necessary addition would be some feature that could help the teacher differentiate between errors and mistakes, for instance by looking at the frequency of error patterns compared to the frequency of correct pattern uses within individual documents. There is certainly good reason to consider such implementations. The tool’s potential to improve equality of feedback and assessment was mentioned by the intervention participants, and it does provide a strong argument for the development of tools with designs more explicitly targeting for those purposes. The idea of a tool that could provide some measure of inter-rater reliability in the summative assessment of student materials is an enticing one.

However, as Brita pointed out, the automated approach is also prone to enforce unwanted aspects of our teaching, such as enforcing specific varieties over others. The issues of enforcement of language norms and gatekeeping based on mastery of those norms were at the core of the formulation of Educational Linguistics (Spolsky, 2008). The reason for this is that they were also at the core of the educational system for a long time, and the work of getting institutions to move away from them in their formulation of, for instance, grading criteria and governing policies for language education is far from done.
The risks of introducing a prescriptive assessment solution into an environment like current-day second language education, at least to me, far outweigh the benefits.

These risks can only be alleviated through the facilitation of a skilled and trained teacher, providing formative feedback for the purpose of language development after having inspected the output of the computer. The intervention studies show that such an implementation is possible, and all of the participants have shown that it can be done without massively changing their current routines. However, in order for such an implementation to succeed, the issues brought up by the respondents in terms of initial workload, report readability, error pattern tiering, processing time, and English varieties must first be addressed through further development of the tool.

In general, the results of the intervention study indicate that the respondents find it likely that the implementation of the tool would result in a higher workload initially, but would be beneficial long-term. However, since the SUS and SF UES scores indicate that the issues with the user interaction with the tool become worse over time, the question of whether the intention to use will remain long-term becomes a pressing one. This is especially worrying when considering implementations in the high-workload environments for which the tool was designed.

8.2 Methodological Considerations

There are several methodological considerations that need to be kept in mind when reading the results of the current study. The main issue of consideration is to do with the validity of the results throughout. This refers both to the validity of the interview data and the ecological validity of the tool.

For the interview data, the validity of the obtained results must be considered from two perspectives: collection and transcription. As mentioned in the method section for the interview instrument, the current dissertation has attempted to verify the collected data by allowing respondents to read and review the materials before submission (Kvale, 2013. p. 121). However, these reviews are not carried out in a vacuum, and the relationship between the interviewer and interviewee must be considered. As the recruitment for the dissertation was conducted through my personal and professional network, 71% of the participants (10 out of 14) had a previous relationship with the person conducting the interviews and transcribing the data. This is likely to impact the validity of the data collection. While this could have had positive effects on the willingness to share during the interview sessions, it does create a bias in the sample that is important to be aware of. Things like intention to use and willingness to engage with the tool were likely biased positively in the sample used for the dissertation, as the project as a whole was understood as being of importance to someone who the respondents were already acquainted with. This
would also contribute to the bias of the results regarding the ecological validity of the tool, which is already apparent through the fact that recruitment would likely attract teachers already open to the idea of technological implementations in their practice and who have an interest in the use of digital tools in the language teaching context.

An additional consideration regarding the validity of materials collected during the intervention study is the collection of data from the upper-secondary participants at the end of the semester. This means that the data collected offers only the final snapshot of the user experience, and does not provide any real insight into how the experience developed over time. While the quantitative collection instruments, the SUS and the SF UES, do indicate that there were developments over time when compared with the initial user tests, such insights have a low validity without measurements detailing the development diachronically amongst the long-term users.

The tool design, as a method of deploying the understanding of the language teaching routines and focal patterns gained from the survey and as an operationalization of the previous research, should also be considered in terms of its validity. As was detailed in the tool development section and the discussion of the results, there were aspects of the design that were not motivated by the survey or previous research, but by practicality and resource limitations. Prime amongst these was the single variant selection, American English, offered. The choice was partially made based on the limited selection of variants offered by the Java LanguageTool, but it was mainly informed by the specifications of applicability found in the TextStat documentation. As the use of several varieties had impacted the validity of the measurements provided by TextStat, a choice was made to limit the tool to American English. However, after receiving the results of the intervention study the better choice seems to have been to lower the validity of the complexity scoring to benefit the diversity of English variants accessible in the tool.

8.3 Limitations

The most visible limitation of the current thesis is the applicability being explored only in the Swedish context. While the research used in the operationalization was international in nature, although mainly from natively English-speaking countries, the implementation of the tool took place within the Swedish system and the results obtained should not be interpreted too strongly beyond that context. This does not mean that the results should be considered generalizable for the Swedish context either.

As stated by Kvale (2013, p. 126), generalizability is often indicated as a flaw in research conducted using interviews as the method. This is due to the often-limited number of participants when compared to other methods. The generalizability is further limited in the current study due to the low number of
participants in both the initial survey and intervention study. Partially, this has
to do with the attrition of respondents over time, but also with the difficulty of
recruiting teachers due to the workload and uncertainty experienced as a result
of COVID-19. This situation also made it difficult to include redundant
participants to ward off issues to do with attrition over time, as suggested by
Källkvist and Juvonen (2021).

Another limitation is the choice of SLA model used to design the intended
use of the tool. While the social constructivist model of SLA and scaffolding
are widely applied in Sweden, as indicated by the results of the initial survey,
the design might not fit equally well in contexts relying on other models. Due
to this, the results should be interpreted as being quite specific to the Swedish
context using that specific model of SLA, in combination with not being
generalizable to the larger Swedish context due to the small number of
participants and biases in the sample. Furthermore, due to the low number of
participants in the intervention, the dissertation is unable to provide any idea of
whether a saturation of knowledge regarding the implementation could be
reached. To once again quote Ferris: “…it is virtually impossible to support any
generalization other than the cliché ‘Further research is necessary’” (1999, p. 5).

8.4 Lessons Learned

As previously mentioned, this short section is intended to spell out the lessons
learned while conducting this dissertation project in order to support future
projects on the same or similar topics. It will repeat some of the content covered
in the section on limitations (8.3) but from a more practice-oriented perspective.
The section has taken inspiration from the template of end-of-project Lessons
Learned documentation available on Monday.com.32 The sub-headings are
Mistakes, Strengths, and Weaknesses.

8.4.1 Mistakes

In retrospect, the reliance on the end-of-semester debriefing of participants for
the main data was a mistake as it positioned the discussion of how the user
experience and usability developed over time as nearly speculative.
Systematically collecting data throughout the entire period of user testing would
have allowed for much more robust insights regarding how the flaws and
affordances of the tool were experienced by the teachers over time, and would
therefore have made it possible to more fully understand the final evaluations.
While a balance between the need for research data and the demands placed on
the participants in terms of time and energy investment must be found, I believe

32 https://monday.com/blog/project-management/lessons-learned-template/ (17/11 -22)
that it would have been a worthwhile endeavor for the current thesis to explore continuous data collection more fully.

Running the tool locally was motivated by the complex issues surrounding sending student texts from the teachers’ work computers for processing somewhere else. However, the evaluations and user tests continually indicated that the setup process and installation were experienced as big barriers to entry, and the intervention studies provided a case where the tool had to be partially redesigned due to the specific conditions of one participant’s work computer. While the background work of navigating the ethical and legislative frameworks needed for running the tool on a server would have taken up a lot of time needed for other aspects of the current thesis, the use of local software should be seen as a mistake from the user perspective and as something that needs to be solved in future implementations of similar tools.

8.4.2 Strengths

A strength of the current project has been the validation of the viability of the concept. Automated diagnostics have been shown to fit within the routines of a small sample of upper-secondary EAL teachers in Sweden, and the features included have also been experienced by those teachers as useful for their practices. The current dissertation has also been able to indicate some aspects of the tool development and design as important for this viability. This ties into the second strength of the project: the participatory design framework.

Employing a teacher-driven design process allowed for the focal patterns of the tool to fit well with the needs of in-service teachers, and it is an aspect that should be further expanded in future projects. The results produced in this project indicate that there are changes in how the teachers experience and make use of the tool over time, meaning that a participatory design process used in combination with the previously mentioned continuous data collection throughout the user evaluation period could lead to meaningful changes in the design of the tool and its features.

The implementation of the tool in teacher practice is also a strength of the approach shown in this project, even though it should be expanded to a larger sample in future studies. The staggered approach of involving the actual intended user base not only in the final evaluation but also in the user testing and formulation stages has allowed the tool developed for the current project to remain connected to the practicalities of the situation for which it was designed.

8.4.3 Weaknesses

Information overload experienced by the teachers using the tool could be pointed to as the main weakness of the approach used in the current project. This is a result of the general approach taken, where the analysis of the text data is, at its first stage, concerned with creating an overview of the documents being
analyzed. The amount of data produced was simply experienced as too much to comprehend in a practical way.

In addition to the amount of data, the opaqueness of the pattern categories inherited by JLT is a weakness of the current approach. Not only were users provided with too much data, but they were also given data categorized in a way that they did not recognize. This issue opens up for a new discussion regarding what the intuitive taxonomy for this type of error pattern categorization would be.

The lack of tiering of the results also contributed to both the overwhelming nature of the tool’s output and the opacity of the resulting categorizations. There are different tiering suggestions discussed in this dissertation, but as these have not been included in the tests of the tool their viability will have to be a future question.
9. Conclusion

This dissertation investigated the possible implementation of automated text diagnostics in Swedish upper-secondary EAL courses through the design and deployment of a tool based on previous research and a survey regarding the routines and focal points of focus-on-form readings among in-service upper-secondary EAL teachers in Sweden. The tool was shown to fulfill the diagnostic needs of upper-secondary EAL teachers, fit well with the pre-existing routines amongst upper-secondary teachers, and fit within the categorization of an augmentation within the SAMR model of technological implementations in the educational context.

The results of the evaluative part of the study indicated that the workload of teachers implementing the tool was likely to initially peak, but that the participants in the intervention study believed the tool could provide long-term benefits. The main obstacles were indicated as being the workload involved in manually inspecting and selecting patterns for relevance and correctness. In addition, the evaluation highlighted the dangers of enforcing unwanted language norms through the prescriptive nature of the tool.

This adds to concerns regarding the outcome of implementing an automated language diagnostics technology in the school context, as there were indicators that the technology might be used for summative assessment and scoring rather than formative assessment and diagnostics. The results indicated that the benefit of such an extension beyond the intentions of the design used for the current project would be the equalizing effect objective analytics could have on summative assessments, as well as negating the heightened workload experienced by teachers during examination periods. However, such an implementation would need other considerations at the core of its design than the ones explored in this dissertation.

From the perspective of the research questions of the dissertation, the results produced are able to provide some answers. Regarding how previous research on feedback in language teaching could influence the design of a digital tool for Swedish EAL teachers, this dissertation chose to focus on fit with current routines and affordances aimed at supporting the teachers in providing scaffolding for their students through corrective feedback. Furthermore, the initial survey indicated that a focus on scaffolding and formative assessment would fit well with the already established practices amongst Swedish upper-secondary EAL teachers. The research on corrective feedback indicated that the tool design would have to be focused on rule-governed error patterns in order for the feedback provided by the teachers to be beneficial to the students. The research on both SLA and corrective feedback showed that the tool design would have to emphasize the role of the teacher as an engineer of the learning experience and as a facilitator of the feedback in order to not create avoidance.
issues and overwhelm the student group. This was also seen among the concerns indicated by the initial survey participants.

Moving on to the actions the tool would have to be able to carry out in order to provide beneficial aid to Swedish EAL teachers, the initial survey provided a list of error patterns that would have to be reliably identified in order for the tool to function as an augmentation of their practices. These error patterns aligned well with the example patterns provided by Ferris (1999). In terms of fitting into the routine, the tool needed to be able to parse student-produced texts, identify error patterns and provide the teacher with a report indicating frequencies and distributions of the patterns amongst the students’ texts.

This brings the focus to the practices and routines that the tool would have to fit into in order to be implemented. As shown by the initial survey, Swedish EAL teachers employed a cyclical, multi-draft routine in their interactions with student texts. The aim of this routine was the incremental scaffolding towards the learning goals, with a focus on formative rather than summative assessment. The tool was designed as an augmentation of the focus-on-form readings performed as a part of the cyclical formative assessment routine.

Finally, the intervention study served to assess the choices made in the design through results indicating how the tool was experienced by the participants. The interventions indicated that upper-secondary participants experienced the tool as the design intended throughout one semester of usage. The issues with noisy data reports were experienced as worsening with extended use, and participants desired a more pupil-facing implementation of text analysis. In terms of teacher-mediated formative assessment as the main goal of the development, the design is indicated as being successful. The main functional issue indicated by the respondents was the amount of work needed to manually inspect material due to the amount of data produced, and the tiering of error patterns was suggested as a solution by the participants.

The university respondents made use of a multi-text routine rather than a multi-draft one, but the tool was still evaluated as fitting well within its categorization as an augmentation of reading student materials with a focus on form. The issues with the overwhelming nature of the report were experienced at the university level as well, in combination with concerns regarding the limited target language variant. In general, the results of the intervention study are interpreted as indicating that a technology based on automated text analysis can fit within the role of an augmented teacher support tool if designed to support formative assessment and feedback, but that it becomes less viable the more assessment approaches summative practices. This was also found to be supported in the previous research, as incorrectly delivered summative assessment is tied to avoidance strategies being developed in the student group.
9.1 Contributions

The main contribution of the dissertation is the expansion of our understanding regarding how digital support tools could fit within the already existing routines of upper-secondary EAL teachers in Sweden. The use of automated text analysis, characterized by respondents as a second opinion, for the focus-on-form reading involved in the text interactions performed by the teachers was shown to provide functional support for the teachers and is thus indicated as being a fitting venue for future tool development. The results also highlighted that a supporting tool for text diagnostics in the position of Augmentation within the SAMR model is a functional one.

The results also highlighted the importance of the teacher as an engaged facilitator of the results from digital analysis tools. Furthermore, the results from the intervention study described the selection of focal points from the teachers’ reading as an integral part of engineering the learning experience for the students. The contribution here is the indication of the continued importance of emphasizing the role of the teacher as central to the success of the students, regardless of the use of automated diagnostics.

9.2 Future Research

Future research based on the results of this dissertation could further explore the second opinion position of automated text analysis as an augmentation of the formative assessments done by teachers in multi-draft designs. This could take the form of the implementation of new tools or continued iterative development of the current tool. The new designs could focus on the aspects of the current tool indicated as problematic by the respondents, such as the limited language varieties available, improving the readability and sorting of the output data, or the onboarding of new users.

An interesting continuation of the current dissertation would be an exploration of how the use of automated text analysis tools influences the teachers’ assessment of student texts. As suggested by the HCI research, the interaction goes both ways and in order to fully understand how these types of tools would influence outcomes in the educational context there must be an understanding of how the tools influence the teachers. This could, for instance, be operationalized through studies using parallel assessment of a collection of learner texts where the inter-rater reliability can be measured to see a difference between assessments with or without the support of a tool.

9.3 Future Development

Future development should focus on the creation of specialized resources for the types of tools explored in this dissertation. This could include the
development of text analytics resources specific to the analysis of English as a second language through the use of learner corpora or manual error annotation, thus improving the precision of the analysis, or the development of distribution pipelines in compliance with GDPR that would allow for centralized operation of the tool to a larger extent. The latter development would also make it possible to collect data regarding the use of the tool and the analysis done on the texts in order to further understand the use and outcomes of implementing the tool in actual school environments.

The use of a server-based solution would also allow for centralized storage of individual student development profiles. Having access to this information could allow for profiling of the analysis in accordance with the students’ learning trajectories, and could also serve to provide an overview of an extended period of time, for instance a semester, much more easily. While this could also be done on the teacher’s computer, there would be benefits of having the data available for development purposes. The access to such data in the continuous development of the tool could be used to incrementally develop the language analysis resources through logging which patterns were removed from the materials by the teacher either due to being irrelevant or due to being erroneous. In addition, this collection of preferences could also be used to create tiering solutions in direct connection to the teachers’ practices.

This dissertation being defended in 2023, the year of the Large Language Model AI, of course also makes one curious about how such technologies could be used for focus-on-form diagnostics in the future. While this project has highlighted the benefits and dangers of applying the static ruleset provided by JLT, future development could engage with the, possibly grander, benefits and dangers of implementing modern AI. This could also bring the discussion back to the two positions from Davies’s (2016) summary: “those who favor the use of AI to develop CALL programs (Matthews 1994) and, at the other extreme, those who perceive this approach as a threat to humanity (Last 1989)”.

203
10. Sammanfattning på svenska

Här följer en kort sammanfattning av avhandlingens innehåll på svenska. Håll i åtanke att detta endast är en sammanfattande del med fokus på resultaten, och att många av de val och överväganden som varit en del av projektet inte finns med.

10.1 Inledning


Avhandlingens huvudbidrag är att skapa en startpunkt för automatiserade stödverktyg för användning i gymnasieskolor inom det svenska skolsystemet. Digitaliseringen av skolmiljön är en självklarhet vid det här laget (Nationell digitaliseringsstrategi för skolväsendet, 2017), men det har främst handlat om att utveckla digitala hjälpmedel för eleverna och allmänna krav på "digitala kompetenser" hos lärare (Eurydice, 2019). Även om dessa saker är viktiga bör stöd till lärare genom teknisk innovation också betraktas som en möjlig väg för förbättring, i enlighet med Skollagens perspektiv på forskning som något som bör informera och påverka lärarnas praktik (SFS, 2010).

Området för denna avhandling är fokus-på-form-diagnostik av skrivna texter på engelska som andraspråk (EAL) 33, men introduktionen av digitala, automatiserade stödverktyg för språklärarer är ett nytt område på många sätt. Således handlar insikterna från detta projekt inte bara om ämnet engelska, utan...

---

33 Taget från engelskans "English as an Additional Language", förkortat EAL.
Frågorna som besvaras i projektet är:

1. Vad påvisar tidigare forskning om korrigerande feedback i språkundervisning som viktigt för designen av ett digitalt verktyg för EAL-lärare i Sverige?
2. Vilka funktioner måste verktyget ha för att ge stöd till EAL-lärare i Sverige?
3. Vilka praktiker och rutiner måste verktyget passa in i för att kunna implementeras?
4. Hur upplever verksamma lärare ett verktyg som är designat utifrån dessa frågor?

Automatisering av fokus-på-form-analys för EAL-klassrumet kan konceptuellt verka som ett enkelt sätt att minska lärarens arbetsbelastning. Ett sådant verktyg skulle också kunna tillhandahålla data som kan vara till nytta för lektionsplanering, individuell utvecklingsplanering och identifiering av lämpliga nivåer av stadiematerial och uppgifter. Dessutom skulle feedback baserad på automatiskt genererade data kunna fungera som ett sätt att betona framsteg som en motivator för eleverna. Denna typ av diagnostisk data kan också användas som ett sätt att identifiera små, konkreta områden för elever som har svårt att engagera sig och behöver mindre områden att fokusera på. Även om det finns tidigare automatiserade lösningar för bedömning, såsom Versant, tenderar dessa att fokusera på uppgiftsspecifik poängsättning och betygssättning snarare än att tillhandahålla information som kan användas för feedback (Streeter et al., 2011). Denna förändring av fokusområde för det automatiserade stödet är både den huvudsakliga funktionella luckan och den huvudsakliga forskningsluckan som adresseras i denna avhandling. Tidigare insatser har utforskat liknande tillvägagångssätt i andra språk utanför den svenska kontexten, till exempel galiciska (Gamallo Otero et al., 2015), men utan utvärdering tillsammans med yrkesaktiva lärare.

10.2 Metod

Eftersom detta projekt kombinerar kvalitativa och kvantitativa data har det placerats inom det pragmatiska paradigmet, där valet av metod till slut bestäms av de praktiska behoven som ingår i att lösa ett specifikt problem (Feilzer, 2009, s. 2). Morgans forskningsdesign (2014, s. 8) beskriver det pragmatiska paradigmet som "the paradigm of choices" på grund av behovet av konstant medvetenhet om de val som ingår i forskningens utformning. Uttrycket kommer ursprungligen från Michael Quinn Patton (2015), som utforskade tillämpningen av mixed methods inom samhällsvetenskapen. Projektet skulle kunna klassificeras som en undersöknings, ett "explicit försök att producera ny kunskap
genom att vidta åtgärder och uppleva deras resultat34" (Morgan, 2014, s. 8). En undersökning inom det pragmatiska paradigmet beskrivs ytterligare som "att möta situationer som faller utanför din befintliga kunskap...35" och lägga till ny kunskap genom resultaten för att navigera den nya situationen i framtiden (Morgan 2014, s. 7). I grunden bygger det pragmatiska paradigmet på pragmatism som en social filosofi, där handlingar leder till erfarenheter som samlas in till trosföreställningar som leder till nya handlingar som leder till nya erfarenheter och så vidare (Morgan, 2014, s. 43).

Ekologisk validitet spelar också en huvudroll i avhandlingen, och används för att ge ett perspektiv på hur verktyget kan anses fungera i den specifika kontext för vilken det är tänkt, dvs undervisningen i engelska på gymnasieskolan. Eftersom den initiala användningen av verktyget sker i kontrollerade miljöer och under begränsade omständigheter spelar de avslutande interventionsstudierna på gymnasiet en stor roll för slutsatsen om hur väl verktyget egentligen passar in.

Designen av verktyget är, som tidigare nämnt, inspirerat av erfarenheter insamlade från lärare i gymnasieskolan, och fokus har legat på hur ett digitalt diagnostiskt verktyg skulle kunna passa in i deras rutiner. Detta undersöcktes genom semi-strukturerade intervjuer i kombination med en respondentteckning. Undersökningen följes av en litteraturöversikt som syftade till att sätta resultaten i sammanhang med den nuvarande litteraturen om förvärv av engelska som ett andra språk (EAL) i det svenska skolsystemet och korrigerande feedback på elevtexter.


10.3 Engelsklärarens diagnostiska behov i svensk gymnasieskola

Den inledande undersökningen om diagnostiska behov hos yrkesverksamma EAL-lärare i Sverige gjordes med hjälp av ett intervjuinstrument och en respondentteckning. Intervjufrågorna syftade till att förstå vilka delar av

34 Min översättning från originalets engelska
35 Som ovan
elevtexter lärare ofta gav kommentarer och feedback på, rutinerna för deras interaktion med elevtexter och deras erfarenheter av att använda digitala stödjande verktyg. Respondentteckningen ingick som ett sätt för lärare att visualisera sina textinteraktionsrutiner, och användes för att informera verktygsdesignen om vilka rutiner som verktyget skulle behöva passa in i. Undersökningen inleddes med två pilotstudier för att förfina intervjuguidernas, som därefter användes för att intervjuja fyra gymnasielärare i engelska.

Resultaten indikerade att digitala resurser ämnade för form-dagnostik i skolmiljö skulle behöva kunna identifiera en rad felmönster, såsom meningsstruktur, interpunktion, tempus, subjekt-verbkongruens, artikelanvändning och stavning. Dessutom skulle verktyget behöva passa in i en redan etablerad praxis som syftar till att ge eleverna det stöd som krävs för att de ska lyckas, samtidigt som man undviker fallgropar som att skapa undvikandeproblem eller genom att överväldiga eleverna.

När det gäller andraspråksinlärningsmodeller gjordes valet att se denna interaktionen mellan lärare och elev utifrån den sociokulturella modellen, med begreppet stödning, från engelskans scaffolding, som en plan för hur verktyget skulle kunna tillämpas. Modellen valdes på grund av den interpersonala aspekten av textinteraktioner som återfinns i den lärarundersökning som genomfördes i början av studien. Andra modeller övervägdes också, eftersom de relaterade till bakgrunden till de senare avsnitten och eftersom verktyget måste förbli funktionellt ur andra modellers perspektiv på grund av att implementeringen i praktiken är beroende av lärarens preferenser.

10.4 Verktygsdesign

De mönster som identifierats av gymnasielärarna användes som underlag för resursval i början av designprocessen. Den tidigare forskningen om korrigerande feedback indikerade att ett viktigt kriterium för en framgångsrik implementering var att målen för feedbacken var regelbaserade fenomen och att eleverna var medvetna om regeln som styrde deras användning. Resultaten från undersökningen stämde väl överens med detta kriterium, vilket tydde på att den tidigare internationella forskningen som ingick i bakgrunden kunde anses ha en viss nivå av ekologisk validitet i den svenska kontexten.

Forskningen om människa-datorinteraktion (HCI) visade på användarvänlighet och upplevd användbarhet som prediktorer för lärare faktiska användning av verktyg, vilket innebar att dessa aspekter måste stå i centrum för designprocessen. Dessutom informerades deltagarna uttryckligen om att deras svar användes direkt i utformningen av verktyget som ett sätt att använda sig av deltagande design, det vill säga att inkludera den tänkta gruppen användare i designen.

36 Taget från engelskans ”Human-Computer Interaction”, förkortat HCI
Baserat på den tidigare forskningen om dator-assisterad språkundervisning (CALL\textsuperscript{37}), skulle de resurser som valts för verktystets språkanalysfunktioner behöva vara exakta i en sådan grad att de skulle kunna ge funktionellt korrekt data om elevtexter. Resursen som valdes för språkanalysen var Java LanguageTool, som tillhandahåller ett lexikonbaserat tillvägagångssätt för att identifiera felmönster i text. Reglerna tillgängliga för det engelska språket inom Java LanguageTool stämde väl överens med de mönster som angivits som relevanta av deltagarna i lärarundersöknings. Java LanguageTool utmärkte sig också som ett bra alternativ på grund av möjligheten att använda en lokal server, snarare än att skicka texter till en annan server över nätet. Detta är viktigt på grund av den känsliga karaktären hos elevtexter, och de möjliga brotten mot GDPR\textsuperscript{38} som kunde skett ifall texten skickats någon annanstans än till lärarnas lokala datorer.

Java LanguageTool ger också utförliga kommentarer om de identifierade mönstren, vilket gör resultaten lättare att förstå. Även om kommentarerna vanligtvis inte lämpar sig för direkt spridning till språkinlärare bidrar de till öppenhet om analysen gentemot läraren. Dessutom, som indikerat av tidigare forskning om HCI, går interaktionen mellan en människa och en dator åt båda hållen, vilket innebär att förekomsten av ett kommentarsfält i resultaten kan påverka användaren att kommentera felmönster i högre utsträckning. Java LanguageTool-kommentaren kan också fungera som en mall att arbeta utifrån och utveckla i formuleringen av feedback.


Dashboardvisualiseringarna som skapades för rapporten valdes baserat på områden som angetts som intressanta i lärarundersöknings. Rapporten delades

\textsuperscript{37} Taget från engelskans "Computer-Assisted Language Learning", förkortat CALL

\textsuperscript{38} Genera Data Protection Regulation, ett ramverk som reglerar behandlingen av personliga data.
upp i två avsnitt, med fokus på översikter på gruppnivå respektive individnivå. Detta var tänkt att stödja läraren i uppgifter som rör antingen lektionsplanering eller individuell feedback.

När verktyget väl byggts testades det i en kontrollerad miljö skapad genom insamling av innehåll från textarkivet Mimers Brunn. Resultaten visade att verktyget fungerade bättre när det användes för att analysera mer formella texter.

Användningen av verktyget i praktiken, under en mindre pilotsstudie, påvisade att analysen av elevtexter resulterade i otydliga data, där många av felmönstren inte var relevanta för läraren. Som ett resultat av detta utvecklades verktyget ytterligare och möjligheten att filtrera felmönsterkategorier innan rapporter producerades implementerades, med de mest relevanta kategorierna enligt undersökningen angivna som standardval.

Efter genomförd pilotstudie av användartester för instrumenten rekryterades tre engelsklärare för ett mer omfattande användartest. Resultaten visade att lärarna kunde se fördelar med att använda verktyget, bland annat fick "Belöning"-faktorn höga poäng, medan poängen som hade att göra med verktygets gränssnitt fick lägre poäng.

### 10.5 Interventioner


Användartestresultaten visar en starkare kontur av mönstret som sägs i den första omgången av användartester, där estetiken och det fokus som verktyget tillåter fick dåliga resultat, medan funktionalitet och belöningsfaktorer fick högre poäng. Generellt sett indikerar resultaten en neutral till positiv upplevelse av funktionaliteten, med en neutral till negativ upplevelse av det visuella gränssnittet.

Intervjusdata analyserades med hjälp av innehållsanalys och framträdande teman var Utmaningar, Möjligheter, Problem och begränsningar, Ekologisk inkompatibilitet, och Framtida utveckling. Ekologisk inkompatibilitet som tema stöddes inte i intervjuerna på universitetsnivå eftersom verktyget endast användes reflekterande och inte ingick i undervisningspraktiken.
Utmaningar som angetts av lärarna tydde på att verktyget upplevdes som överväldigande, och det föreslogs av deltagarna att analyseresultatet av elevtexterna skulle anpassas efter antingen användarprofiler eller felsignifikans. En majoritet av de tillfrågade rapporterade också läsbarheten av datatabellerna som ett problem, bland annat på grund av upplevelsen av att analysera data i tabellformat, antalet falska positiva utslag och diskrepans mellan den terminologi som används för att kategorisera felmönstren i verktyget och den lärarna vanligtvis använder.

Bland fördelarna med verktyget tydde resultaten på att det hade uppnått en god passning med respondenternas rutiner, vilka liknade de rutiner som fanns i resultaten från lärarundersöknings. Verktyget karakteriseras som en ”second opinion” på bedömningen av elevtexter, vilket ansågs vara fördelaktigt eftersom det stöttade dem i att identifiera sådant som de annars kunde ha missat. Två deltagare lyfte också fram likvärdigheten i feedbacken som möjliggjordes av verktyget, eftersom deras bedömningar inte skulle vara lika påverkade av saker som trötthet.


Frågan om falska positiva utslag visade sig förvärra rapportens överväldigande karaktär på två sätt. För det första bidrog antalet falska positiva utslag till mängden data som kommunicerades till användaren. För det andra tvingade det användaren att lägga mer tid på att manuellt inspекtera och filtrera resultaten. Således skulle vidareutveckling av språkanalysresurserna gynna verktyget inte bara funktionellt, utan också genom att gynna användarupplevelsen av gränssnittet.

10.6 Slutsatser

Denna avhandling har undersökt implementeringen av automatiserad textdiagnostik i engelskundervisning på gymnasiet genom utformning och användartesting av ett verktyg baserat på tidigare forskning och en kartläggning av rutinerna och fokuspunkterna bland engelsklärares på gymnasienivå. Verktyget visade sig uppfylla de diagnostiska behoven hos lärarna, passa väl med de redan existerande rutinerna och passa in i
kategoriseringen av en augmentering inom SAMR-modellen, vilken erbjuder en taxonomi för verktygs roll och inverkan i lärmiljöer.

Resultaten av den utvärderande delen av studien indikerade att arbetsbelastningen för lärare som implementerade verktyget sannolikt skulle bli högre till en början, men att deltagarna i interventionsstudien ansåg att verktyget kunde ge långsiktiga fördelar. De huvudsakliga hindren angavs vara den arbetsbelastning som är involverad i att manuellt inspektera och välja skrivfelsmönster baserat på relevans och korrekt identifiering. Dessutom lyfte utvärderingen fram farorna med att upprätthålla oönskade språknormer gällande målspråksvariant (exempelvis amerikansk eller brittisk engelska) genom verktygets preskriptiva karaktär.

Vidare visade utvärderingen att tekniken skulle kunna komma att användas för summativ bedömning och poängsättning snarare än för syftet formativ bedömning och diagnostik. Detta vittnar om att automatiserad språkdiagnostik riskerar att användas för andra syften än de ursprungliga i skolsammanhang. Resultaten indikerade att fördelen med en sådan utvidgad tillämpning av den nuvarande designen projektet skulle vara den utjämnande effekten objektiva analyser skulle kunna ha på summativa bedömningar. Verktyget sågs också som en möjlighet att mildra den ökadearbetsbelastningen som lärarna upplevde under provperioder. En sådan summativ implementering skulle dock kräva andra överväganden i designen än de som utförkors i denna avhandling.

När det gäller hur tidigare forskning om feedback i språkundervisningen skulle kunna påverka utformningen av ett digitalt verktyg för svenska gymnasielärare i engelska, valde denna avhandling att fokusera på anpassning till existerande rutiner och funktioner som syftar till att stödja lärarna i att tillhandahålla stöttning till sina elever genom korrigeraende feedback. Forskningen om korrigeraende feedback indikerade att verktygsdesignen måste fokuseras på regelstyrda felmönster för att feedbacken från lärarna ska vara till nytta för eleverna. Forskningen om både andraspråksiinlärning och korrigeraende feedback visade att verktygsdesignen skulle behöva betona lärarens roll som föremålare av feedback. Detta sågs också bland de farhågor som de första undersökningsdeltagarna lyfte.

Slutligen tjänade interventionsstudien till att bedöma de val som gjorts i designen av verktyget genom resultat om hur verktyget upplevdes av deltagarna. Interventionerna visade att gymnasielärarna upplevde verktyget som designen avsåg under en termins användning. Problemen med överväldigande datarapporter upplevdes förvärras vid långvarig användning, och deltagarna önskade en mer elevinriktad implementering av verktyget. När det gäller lärarföremålade formativ bedömning som huvudmål för utvecklingen av dessa typer av verktyg indikeras designen generellt som fungerande. Det huvudsakliga problemet som respondenterna tog upp var mängden arbete som behövdes för att manuellt granska den stora mängden material. Som tidigare
nämnt föreslogs en nivåindelningen av felmönster som en möjlig lösning av deltagarna.

Universitetslärarna använde sig av en rutin med flera separata textinlämningar snarare än en med flera utkast, men verktyget bedömdes ändå passa väl in. Problemen med rapportens överväldigande karaktär upplevdes även på universitetsnivå, i kombination med oro för det begränsade antalet målspråksvarianter. Generellt sett tolkas resultaten av interventionsstudien som att en teknik baserad på automatiserad textanalys kan passa in i rollen som ett stödverktyg.
Works Cited


Davies, G. (2016) *CALL (computer assisted language learning)’. Southampton. Subject Centre for Languages, Linguistics and Area Studies Guide to Good Practice*. Available at: http://www.llas.ac.uk/resources/gpg/2241.


Lundh, F. (1999) *An introduction to tkinter*. Available at: URL:


Nordin (red), A. and Uljens (red), M. (2022) *Didaktikens språk: om skolundervisningens mål, innehåll och form.*


Skolverket (2018) Curriculum for the compulsory school, preschool class and school-age educare.


Appendices:

Appendix A: Interview guide 1

Category 1: The Teacher and Variables

1. Hur länge har du arbetat som lärare?
2. Vilka skolformer har du arbetat i, och vilka kurser undervisar du i?
3. Hur ser din utbildning ut?
4. Har du lärarlegitimation för ämnet engelska?
5. Är du legitimerad och/eller undervisar i andra ämnen?
6. Hur har ditt arbetsliv sett ut? (Narrative)

Category 2: The Students

7. Vilka språkliga svårigheter har dina elever när de skriver på engelska?
8. Finner du samma svårigheter i skriftlig och förberedd muntlig produktion?
9. Finner du samma svårigheter i skriftlig och icke-förberedd muntlig produktion?
10. Vilka är de vanligaste språkliga svårigheterna du möter i dina elevgrupper?
11. Ser du en koppling mellan elevgruppers tidigare språkundervisning och de svårigheter som är återkommande i deras texter?
12. Är det stora skillnader mellan dina elevers förkunskaper?

Category 3: Routines and Feedback

13. Har de språkliga svårigheter du hittar i elevers texter någon inverkan på din grammatikundervisning?
14. Vilken form brukar den återkoppling dina elever får på sina skriftliga uppgifter ha?
15. Hur skulle du beskriva ditt arbetssätt med elevtexter från inlämning till återkoppling?
16. Finns det rutiner för att hjälpa elever som har återkommande svårigheter i engelska på din arbetsplats, eller på några av dina tidigare arbetsplatser?
17. Vad brukar du ge feedback på vad gäller språket?
18. Om du inte hade några begränsningar vad gäller tid och resurser, hur hade du velat ge återkoppling till dina elever?

Category 4: Digital Experiences and Tools

19. Använder ni någon form av digital plattform, och hur funkar den?
20. Hur fungerar elevinlämningar på den plattformen?
21. Känner du dig bekväm med det/de digitala stöd du har tillgängligt?
22. Använder du någon form av digitala verktyg eller stöd i din undervisning eller feedback?
23. Har du använt någon form av digital språk-korrigerings? (Word, Grammarly, Hemingway etc)
24. Vilken information skulle du vilja ha tillgång till vad gäller dina elevers skriftliga språk?
25. Märker du några nya problem pga elevers tillgång till digitala verktyg?

Debrief:

Repetera huvudpunkterna för eventuella tillägg.

Har du några frågor eller skulle du vilja tillägga något?
Behov inför utformning av diagnostiska verktyg i engelskundervisning

Du inbjuds härmed att delta i en intervjustudie som är utforskar vilka typer av språkproblem som lärare i gymnasieskolans engelskkurser upplever som återkommande i sina elevgrupper, samt hur man i praktiken väljer att arbeta med dessa. Projektet genomförs av mig under ledning av min handledare professor Jukka Tyrkkö vid Linnéuniversitetet i Växjö och pågår mellan 2018 och 2023. Intervjun tar 40 minuter.

Vad handlar projektet om?

Intervjustudien är en del av ett doktorandprojekt med målsättningen att utveckla resurser för att hjälpa lärare att snabbare få en översikt över elevgruppens skriftliga produktion på engelska. Syftet med det digitala verktyget är att öka lärarens förutsättningar att på ett tidseffektivt och fokuserat sätt undervisa i språkriktighet och
språklig korrekthet i engelskan. Planen är att anpassa digitala verktyg som används i språkvetenskaplig forskning för syftet.

Intervjuerna handlar om vilka språkproblem som du ser som återkommande i de elever du möter, eller har mött, i ditt arbete samt vilka metoder, övningar och tillvägagångssätt som du anser vara effektiva. Jag är också intresserad av hur ditt arbetsflöde ser ut efter att elever har lämnat in arbeten, eftersom projektet syftar till att utveckla ett verktyg som passar ditt arbetssätt

Hur kommer materialet att användas?

Det är möjligt att all forskning på det insamlade materialet inte ryms inom ramen för det här projektet; det blir då publicerat efter projektets slut istället. Om du svarar ”ja” betyder det helt enkelt att jag får använda materialet i min forskning och publicera forskning baserad på materialet. Ditt samtycke innebär också att materialet kan användas i samarbeten med andra universitet och andra forskare, men även där i anonymiserad form. Även om du kryssar i ”Ja” nu kan du när som helst välja att dra tillbaka ditt samtycke och utgå ur studien.

Projektet följer lagstiftning och regelverk om forskningsetik. I reglerna ingår att den som är med i ett forskningsprojekt ska veta vad forskningens syfte är, hur den går till och vad det innebär att vara med i detta projekt. Deltagande i projektet är helt frivilligt. I ett formulär på nästa sida kryssar du i om du vill vara med. Alla som

---

39 Finns att läsa på Codex.vr.se
deltar i studien kommer att vara anonyma. Ljudinspelningar och annat material kommer att förvaras på ett säkert sätt så att det inte sprids. All data som samlas in inom ramen för studien kommer att behandlas konfidentiellt och enbart i forskningssyfte. Transkriptioner kommer att avidentifieras. Efter projektslut lagras anonymiserat material enligt riktlinjer från Vetenskapsrådet.

Du har enligt lag rätt att en gång per år, gratis, få ta del av de personuppgifter som registreras om dig som en del av projektet. I sådana fall kontaktar du mig så berättar vi hur det går till.

Om du har några frågor om projektet eller vill tillägga något är du alltid välkommen att kontakta mig!

Med vänliga hälsningar
Daniel Sundberg
Tel. 0768 89 7939

Jukka Tyrkkö
Jukka.tyrkk@lnu.se
Tel. +46470767852

---
40 I got married during my PhD studies, which is why documents from before December of 2020 is under the name Daniel Sundberg rather than Daniel Ihrmark.
Vänligen fyll i talongen nedan och lämna in innan intervjun.

**Informerat samtycke**

- Genom att kryssa 'Ja' och skriva under ger jag mitt samtycke till deltagande i forskningsstudien. Jag har tagit del av informationen och känner till hur studien kommer att gå till. Jag är också medveten om att jag när som helst kan välja att dra tillbaka mitt samtycke och utgå ur studien.

☐ **JA**

__________________________________________

För- och efternamn

__________________________________________

Ort och datum
Appendix C: Quick start guide

Instructions for FirstContact v.0.1
Contact: daniel.o.sundberg@lnu.se
OS: Windows, 64-bit
Input type: .docx

Note that this version will open a terminal window when the program is started. If you encounter a crash, malfunction or other issues, please copy the text in the terminal and send it to the email above labelled “Malfunction notice”.

Before starting: Please make a note of your system specifications below\(^4\). This should be included when submitting a debrief or malfunction notice:
Processor:
RAM:

Guidelines:

- The tool is intended for use with formal to semi-formal text production.
- The intended input language variant is American English.

Always inspect output manually.

1. Installation
   a. Download the appropriate archive from https://drive.google.com/drive/folders/1SpeG4juXwCPvkoD6lDUcehA2aGZj8RvE?usp=sharing (.7z file only works with 7zip zip archive manager).
   b. Extract all files from the archive using either 7zip or other .zip archive manager.
   c. Install Java SDK through included installer (Skip if already installed).

2. Using the software
   a. Start the program by clicking the shortcut “FirstContact v.0.1”.
   b. Click “Browse input” and select the folder where your texts are located.
   c. Click “Update file list”. Inspect the output under “Files in directory” to make sure you have chosen the correct folder.
d. Click “Analyze texts” to process the documents. This may take several minutes.\textsuperscript{42}

e. The analysis is complete when “Analysis completed!” is shown in the “Processed files” box. Do not press any buttons while the program is processing the texts.

f. Click “Generate report” to close the launch client and launch the dashboard.

Glossary:

i. JLT = Java LanguageTool. Indicates frequencies of issues found during language analysis and categories.

ii. Consensus score = Combined readability and complexity scores supplied by TextStat. The consensus score is expressed as U.S grade level.

\textsuperscript{42} Program might crash here during first run. This is due to Java LanguageTool being installed, which should be visible in the terminal window. Program should run on second start.
Dashboard Overview:
Starting tab / Overview tab

Graph of JLT pattern counts and readability / complexity scoring of all texts

Overview of Textstat and descriptive statistics for JLT and texts

Grouped bar chart indicating JLT categories per text

Sunburst graph detailing categories and sub-categories per text

Table with JLT category and sub-category count for the texts
Individual Overview tab:

Sunburst of categories and sub-categories for selected text

Bar graph indicating number of JLT hits in the texts if no text is chosen

JLT hits for text selection
Appendix D: User testing invitation

Är du sugen på att testa ett nytt digitalt verktyg för lärare i engelska?

Jag heter Daniel Ihrmark (f.d. Sundberg) och är doktorand i engelska vid Linnéuniversitetet i Växjö. Mitt doktorandprojekt har handlat om att bygga ett verktyg som är tänkt att stötta lärare på gymnasienivå i deras läsning av elevtexter, framförallt genom att minska arbetsbörden kring närläsning med fokus på form. Det finns en länk längst ned i posten där man kan läsa mer om projektet. Utvecklingen av verktorget har nu kommit så långt att det är dags att börja testa hur lärare upplever användandet av det, och det är därför jag postar här.

Användartesterna går till på så vis att deltagare får tillgång till en virtuell dator och elevtexter hämtade från Mimers Brunn via programmet TeamViewer. Därefter får ni analysera elevtexterna med hjälp av verktorget medans ni diskuterar hur ni tänker och vad ni ser med mig på Zoom. Slutligen fylls en enkät kring användarupplevelsen i. Sammanlagt är det hela beräknat att ta ungefär 45 minuter.

Informationen som samlas in från användartesterna är anonym, och man är helt fri att dra tillbaka sitt deltagande när som helst. Om detta låter intressant så får ni gärna kontakta mig och boka en tid för att testa på verktorget. Min adress är daniel.o.sundberg@lnu.se. Jag svarar gärna på frågor kring verktorget och projektet både via mail och i kommentarerna.


Mvh
Daniel
Appendix E: Interview guide 2

Category 1: The Teacher and Variables

1. Hur länge har du arbetat som lärare?
2. Vilka skolformer har du arbetat i, och vilka kurser undervisar du i?
3. Hur ser din utbildning ut?
4. Har du lärar legitimation för ämnet engelska?
5. Är du legitimerad och/eller undervisar i andra ämnen?
6. Hur har ditt arbetsliv sett ut? (Narrative)

Category 2: The Students

7. Vilka språkliga svårigheter har dina elever när de skriver på engelska?
8. Finner du samma svårigheter i skriftlig och förberedd muntlig produktion?
9. Finner du samma svårigheter i skriftlig och icke-förberedd muntlig produktion?
10. Vilka är de vanligaste språkliga svårigheterna du möter i dina elevgrupper?
11. Ser du en koppling mellan elevgruppens tidigare språkundervisning och de svårigheter som är återkommande i deras texter?
12. Är det stora skillnader mellan dina elevers förkunskaper?

Category 3: Routines and Feedback

13. Har verktyget identifierat språkmönster bland dina elever, i så fall vilka?
14. Har de språkmönster verktyget identifierat i elevers texter haft någon inverkan på din undervisning?
15. Har de språkmönster verktyget identifierat i elevers texter haft någon inverkan på den feedback du gett dina elever?
16. Hur skulle du beskriva ditt arbetssätt med elevtexter från inlämning till återkoppling innan du börjat använda verktyget?
17. Hur skulle du beskriva ditt arbetssätt med elevtexter från inlämning till återkoppling efter att du börjat använda verktyget?
18. Vad brukar du ge feedback på vad gäller språket, och har verktyget haft någon inverkan?
19. Har verktyget påverkat hur du ger feedback till dina elever?
20. Upplever du att verktyget har påverkat din arbetsbördapositivt, negativt eller inte alls?

Category 4: Digital Experiences and Tools

21. Använder ni någon form av digital plattform, och hur funkar den?
22. Har du använt någon form av digital språk-korrigerings innan detta verktyget? (Word, Grammarly, Hemingway etc)
23. Hur har verktyget hanterat formatet på dina elevinlämnningar?
24. Har du haft någon nytta av informationen verktyget producerat?
25. (Om 22 == Ja) Vilken typ av information från verktyget har du funnit användbar?
26. Hur har informationen använts?
27. Märker du några nya problem pga verktyget?
28. Vad hade du velat ändra?

Debrief:

Repetera huvudpunkterna för eventuella tillägg.

Har du några frågor eller skulle du vilja tillägga något?
Appendix F: Consent form 2

Utvärdering av digitalt verktyg för språkanalyse
Du inbjuds härmed att delta i en intervjustudie som är utforskar och utvärderar användandet av verktyget FirstContact i engelskundervisning. Projektet genomförs av mig under ledning av min handledare professor Jukka Tyrkkö vid Linnéuniversitetet i Växjö och pågår mellan 2018 och 2023. Intervjun tar 40 minuter.

Vad handlar projektet om?
Intervjustudien är en del av ett doktorandprojekt med målsättningen att utveckla resurser för att hjälpa lärare att snabbare få en översikt över elevgruppers skriftliga produktion på engelska. Syftet med det digitala verktyget är att öka lärares förutsättningar att på ett tidseffektivt och fokuserat sätt undervisa i språkriktighet och språklig korrekthet i engelskan.

Intervjuerna handlar om hur du har upplevt användandet av verktyget i ditt arbete. Jag är också intresserad av hur ditt arbetsflöde ser ut efter att elever har lämnat in skriftliga arbeten, eftersom projektet syftar till att utveckla ett verktyg som passar ditt arbetssätt
Hur kommer materialet att användas?

Det är möjligt att all forskning på det insamlade materialet inte ryms inom ramen för det här projektet; det blir då publicerat efter projektets slut istället. Om du svarar ”ja” betyder det helt enkelt att jag får använda materialet i min forskning och publicera forskning baserad på materialet. Ditt samtycke innebär också att materialet kan användas i samarbeten med andra universitets och andra forskare, men även där i anonymiserad form. Även om du kryssar i ”Ja” nu kan du när som helst välja att dra tillbaka ditt samtycke och utgå ur studien.


43 Finns att läsa på Codex.vr.se
Du har enligt lag rätt att en gång per år, gratis, få ta del av de personuppgifter som registreras om dig som en del av projektet. I sådana fall kontaktar du mig så berättar vi hur det går till.

Om du har några frågor om projektet eller vill tillägga något är du alltid välkommen att kontakta mig!

Med vänliga hälsningar
Daniel Ihrmark
Daniel.o.sundberg@lnu.se
Tel. 0768 89 7939

Jukka Tyrkkö
Jukka.tyrkko@lnu.se
Tel. +46470767852
Vänligen fyll i talongen nedan och lämna in innan intervjun.

**Informert samtycke**

- Genom att kryssa ’Ja’ och skriva under ger jag mitt samtycke till deltagande i forskningsstudien. Jag har tagit del av informationen och känner till hur studien kommer att gå till. Jag är också medveten om att jag när som helst kan välja att dra tillbaka mitt samtycke och utgå ur studien.

☐ **JA**

________________________________________

För- och efternamn

________________________________________

Ort och datum