pedocs
Open Access Erziehungswissenschaften

DIPF

Flütsch Keravec, Silvana

Speech-to-text technology for students with dyslexia. Benefits and challenges of the tool in the writing classroom

Münster ; New York : Waxmann 2026, 252 S. - (Sprachliche Bildung / Studien; 18)

Quellenangabe/ Reference:
Flütsch Keravec, Silvana: Speech-to-text technology for students with dyslexia. Benefits and challenges of the tool in the writing classroom. Münster ; New York : Waxmann 2026, 252 S. - (Sprachliche Bildung / Studien; 18) - URN: urn:nbn:de:0111-pedocs-366877 - DOI: 10.25656/01:36687; 10.31244/9783818851293

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-366877
https://doi.org/10.25656/01:36687

in Kooperation mit / in cooperation with:

WAXMANN
www.waxmann.com

http://www.waxmann.com

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: http://creativecommons.org/licenses/by-nd/4.0/deed.de - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: http://creativecommons.org/licenses/by-nd/4.0/deed.en - You may copy, distribute and transmit, adapt or exhibit the work in the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to alter or transform this work or its contents at all.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der Leibniz-Gemeinschaft

18 Studien

Sprachliche Bildung

MERCATOR INSTITUT
für Sprachförderung
und Deutsch
als Zweitsprache

Silvana Flütsch Keravec

Speech-to-Text Technology for Students with Dyslexia

Benefits and Challenges of the Tool in the Writing Classroom

WAXMANN

Sprachliche Bildung – Studien

herausgegeben vom
Mercator-Institut für Sprachförderung und
Deutsch als Zweitsprache der Universität zu Köln

Michael Becker-Mrotzek & Hans-Joachim Roth

Band 18

Sprachliche Kompetenzen sind eine wesentliche Voraussetzung dafür, dass Kinder und Jugendliche ihr Potenzial ausschöpfen können – in der Schule, im Studium, in der Berufsausbildung und in der Gesellschaft. Dennoch verfügt ein zu großer Teil der Schülerinnen und Schüler nicht über ausreichende bildungssprachliche Qualifikationen, um die Schule erfolgreich abzuschließen und ein selbstbestimmtes Leben zu führen. Die Schriftenreihe „Sprachliche Bildung" des Mercator-Instituts für Sprachförderung und Deutsch als Zweitsprache bündelt Modelle, Konzepte und Forschungsergebnisse aus Sprachdidaktik, Bildungswissenschaften, Linguistik und Psychologie und liefert neue Impulse für die Diskussion und Forschung im Feld sprachliche Bildung. In der Unterreihe „Sprachliche Bildung – Studien" erscheinen Monographien, in erster Linie exzellente und innovative Dissertations- und Habilitationsschriften. Im Zentrum stehen Studien zu Prozessen, Methoden und Formaten der sprachlichen Bildung und Förderung in einer durch Vielfalt geprägten Gesellschaft.

Herausgeber der Reihe ist das Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. Das Institut ist ein durch die Stiftung Mercator initiiertes und von 2012 bis 2024 gefördertes Institut der Universität zu Köln. Mit seiner Forschung und seinen wissenschaftlichen Serviceleistungen zu sprachlicher Bildung in einer mehrsprachigen Gesellschaft trägt das Mercator-Institut zu mehr Chancengleichheit im Bildungssystem bei.

Silvana Flütsch Keravec

Speech-to-Text Technology for Students with Dyslexia

Benefits and Challenges of the Tool in the Writing Classroom



Waxmann 2026
Münster • New York

Published with the support of the Swiss National Science Foundation

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the
Deutsche Nationalbibliografie; detailed bibliographic data
are available in the Internet at http://dnb.dnb.de

Sprachliche Bildung – Studien, vol. 18

ISSN 2569–5045
Print-ISBN 978-3-8188-0129-8
E-Book-ISBN 978-3-8188-5129-3
https://doi.org/10.31244/9783818851293
The e-book is licensed under Creative Commons CC BY-ND 4.0
(https://creativecommons.org/licenses/by-nd/4.0/legalcode.de).



Silvana Flütsch Keravec
Waxmann Verlag GmbH, 2026
Münster, Germany

www.waxmann.com
info@waxmann.com

Cover Design: Anne Breitenbach, Münster
Typesetting: MTS. Satz & Layout, Münster

This licence applies only to the original material. All marked third-party content (e.g., illustrations, photos, quotations, etc.) is excluded from the CC licence, and it may be necessary to obtain further permission from the copyright holder for its reuse.

Acknowledgements

First and foremost, I wish to thank my pillars, Erwan, Anaïs, Emilie and Maël, who supported and accompanied me through the PhD journey and to the other end of the world. Thank you for your patience and for forcing me to disconnect from time to time. Thanks to my dear parents, family and friends for your support and taking care of my kids when I was hidden behind a screen – and especially Florian and Anne for countless meals and fruit breaks.

I would like to thank Prof. Dr. Hansjakob Schneider for the support and advice for the grant application and to both supervisors, Prof. Dr. Erich Hartmann and Prof. Dr. Hansjakob Schneider, for the supervision, guidance and insightful feedback during the whole PhD journey.

Thanks to Prof. Dr. Michael Geiss, Dr. Ursula Stohler and the Centre for Education and Digital Transformation for including me in this inspiring and encouraging centre; the PHZH for the financial support during the grant application and for providing a great working environment. Thanks to the Swiss National Science Foundation and Doc.CH for accepting and financing this thesis research. Thanks to Prof. Dr. Anne Castles and Ass. Prof. Dr. Lisi Beyersmann for accepting me at the Macquarie University. Being part of the lab and getting a taste of morphology research means a lot to me.

Special thanks to PD Dr. Verena Hofmann, Ass. Prof. Dr. Peter Petocz and Prof. Dr. Roger Keller for your precious methodological advice and to Dr. Bradley A. Smith for patiently proofreading and teaching me scientific writing in English.

Many thanks to all the students, teachers, and specialists who have contributed to the studies included in this thesis. Thank you for your curiosity and courage to participate in a study and for your efforts during assessments, writing lessons and interviews.

Last but not least, I wish to thank my dear fellow PhD students and writing mates Dr. Yan Feng, Esra Ataman, Ozan Caglar, Simone Berner-Nayer, Mirjam Nievergelt, Dr. Carmen Flury, Angela Bonetti, Liana Pirovino, Judit Martínez Moreno, and Dr. Gunilla Almgren-Bäck. Your wise advice, our precious exchange, invaluable coffee breaks and so much more remain in my best memories.

Contents

1. Introduction 13
1.1 Relevance of research on speech-to-text technology for students with dyslexia 13
1.2 Research goal and questions 17
1.3 Research outline 18
1.4 Structure of the thesis 19

2. Writing despite Dyslexia 21
2.1 Writing competence 22
2.1.1 Models of writing 23
2.1.1.1 Written products, writers and the community: The Writer(s)-within-Community Model 23
2.1.1.2 Writing processes in models including development of processes 25
2.1.1.3 Toward a parallel and cascading model of the writing system – Olive (2014) 29
2.1.2 Writing instruction and writing task 31
2.1.2.1 Writing instruction 32
2.1.2.2 Task environment and writing task 34
2.1.3 Assessment of writing skills 35
2.1.3.1 Assessment of lower-order writing processes and products 36
2.1.3.2 Assessment of higher-order writing products 37
2.2 Dyslexia 40
2.2.1 Classification and definition(s) 41
2.2.2 Prevalence and aetiology 43
2.2.3 Symptomatology, secondary symptoms, and comorbidities 45
2.2.3.1 Symptoms of dyslexia 45
2.2.3.2 Secondary symptoms, consequences on school and professional career 47
2.2.3.3 Comorbidities 47
2.2.4 Assessment, intervention, adaptation, and compensation 49
2.2.4.1 Assessment of dyslexia 49
2.2.4.2 Interventions 51
2.2.4.3 Adaptational measures 52
2.2.4.4 Compensational measures and compensation of disadvantages 'Nachteilsausgleich' 53

3. Speech-to-text for Students with Dyslexia – Theoretical Approach 55
3.1 Overview of technology-supported writing tools for students with dyslexia 55
3.2 Speech recognition technology 57
3.2.1 Development of the technology 57
3.2.2 Considerations about speech recognition for different user groups 62

3.2.3 Definition and characterisation of assistive technology ... 64
3.3 Reasons for STT as assistive and/or enabling technology ... 66
3.3.1 Cognitive Load Theory ... 67
3.3.1.1 Theoretical framework ... 67
3.3.1.2 Working memory and long-term memory in CLT ... 68
3.3.1.3 Categories of cognitive load ... 68
3.3.1.4 Writing processes from a CLT perspective ... 70
3.3.1.5 Deduction from CLT perspective regarding the application of STT for students with dyslexia ... 71
3.3.2 Writing motivation ... 73
3.3.2.1 Framing the construct of writing motivation ... 74
3.3.2.2 Development of writing motivation ... 79
3.3.2.3 Writing motivation and achievement ... 81
3.3.3 Situating STT for students with dyslexia within further theoretical perspectives ... 84
3.3.3.1 Rationale for text production with STT from the perspective of New Literacies Theory ... 84
3.3.3.2 STT in the writing classroom from the perspective of media education ... 85
3.4 Summary ... 89

4. State of the Art on Speech Recognition Technology for Learners with Dyslexia ... 91
4.1 Overview of the state of research regarding the focus user group ... 91
4.2 Benefits and challenges of STT for students with dyslexia in school ... 94
4.2.1 Research findings on STT condition and higher-order writing skills ... 94
4.2.1.1 Text quality ... 94
4.2.1.2 Vocabulary diversity, text length, and word length as predictor variables of text quality ... 95
4.2.1.3 Writing strategies and revision ... 96
4.2.2 Research findings on STT condition and lower-order writing performance ... 97
4.2.3 Research findings on STT condition and writing-related skills and demands ... 99
4.2.3.1 Reading fluency ... 99
4.2.3.2 Writing motivation, general school motivation ... 99
4.2.4 Participation, social inclusion, and well-being ... 99
4.2.5 Recognition issues, technical issues, and time consumption ... 101
4.3 Conditions for successful STT use ... 102
4.3.1 Contextual factors ... 102
4.3.2 Software introduction and support ... 104
4.3.3 Combination with effective writing instruction ... 106

5. Intervention Study – Methodological Approach ... 108
5.1 Preliminary notes on the present methodological approach ... 108
5.1.1 Three study groups and their functions ... 108
5.1.2 Investigating the effect of a writing tool by writing assessment ... 109

5.2 Research questions and hypotheses ... 109
5.2.1 Research question 1 ... 109
5.2.1.1 Effects of STT on text quality – RQ 1A ... 109
5.2.1.2 Effects of provided writing instruction by videos on text quality – RQ 1B ... 110
5.2.1.3 Effects of STT on text length – RQ 1C ... 111
5.2.1.4 Effects of provided writing instruction by videos on text length – RQ 1D ... 112
5.2.2 Research question 2 ... 112
5.2.2.1 Effects of STT on handwriting fluency – RQ 2A ... 112
5.2.2.2 Effects of STT on reading fluency – RQ 2B ... 113
5.2.3 Research question 3 ... 114
5.2.3.1 Dictating fluency versus handwriting fluency – RQ 3A ... 114
5.2.3.2 Development of dictating fluency – RQ 3B ... 115
5.2.4 Research question 4 ... 115
5.2.4.1 Effects of STT on writing motivation –RQ 4A ... 115
5.2.4.2 Effects of provided writing instruction on writing motivation – RQ 4B ... 116
5.2.5 Research question 5 ... 116
5.3 Research design ... 117
5.3.1 Overall research design – quantitatively-driven, mixed methods design ... 117
5.3.2 Integration strategy of quantitative and qualitative data ... 119
5.3.3 Quantitative core – quasi-experimental study ... 119
5.3.3.1 Quantitative design ... 120
5.3.3.2 Statistical analyses – GLM ... 126
5.3.4 Sample – Quasi-experimental study ... 132
5.3.4.1 Participant recruitment ... 132
5.3.4.2 Inclusion and exclusion criteria ... 133
5.3.4.3 Sample description – Quasi-experimental study ... 134
5.3.4.4 Participant characteristics and comparability of the study groups ... 134
5.3.5 Qualitative supplemental component – Semi-structured interviews ... 141
5.3.5.1 Design of qualitative supplemental component ... 141
5.3.6 Sample – Interview study ... 144
5.4 Selection, adaptation, and fit of assessment instruments ... 145
5.4.1 Assessment of text quality – TOWL-4 ... 147
5.4.2 Assessment of writing fluency – CBM-W ... 151
5.4.3 Assessment of reading fluency – SLS 2–9 ... 155
5.4.3.1 Salzburger Lesescreening 2–9 (SLS 2–9) ... 156
5.4.3.2 Standardised assessment by instructional video ... 157
5.4.4 Assessment of writing motivation – Translation & adaptation of SWAS ... 157
5.4.4.1 Considerations for the selection of an assessment tool ... 158
5.4.4.2 Fit and selection of SWAS to assess writing motivation ... 158
5.4.4.3 Process of translation and adaptation of the survey ... 159
5.4.4.4 Limitations of writing motivation assessment in this thesis research ... 162
5.4.4.5 Additional questions on tool predilections for the EG 1 ... 163

5.4.5 Preparation of qualitative data collection – interview guide and code system ... 164
5.4.5.1 Construction of interview guide ... 164
5.4.5.2 Development of code system ... 165
5.4.6 Summarised overview of the survey instruments ... 167
5.5 Intervention – Equipment, instructional, and teaching materials ... 167
5.5.1 Selection of hardware and software ... 169
5.5.2 STT-instruction in individual setting for EG 1 ... 170
5.5.3 Writing intervention for EG 1 and EG 2 ... 171

6. Study Results ... 175
6.1 Results of quantitative data analysis ... 175
6.1.1 Selection of students' characteristics from theory and qualitative component ... 176
6.1.2 Research Question 1 – Text quality and text length ... 177
6.1.2.1 RQs 1 A and 1 B – Text quality ... 177
6.1.2.2 Exploration of effects of participants' characteristics on text quality ... 178
6.1.2.3 RQs 1 C and 1D – Text length ... 180
6.1.2.4 Exploration of effects of participants' characteristics on text length ... 181
6.1.3 Research Question 2 – writing fluency and reading fluency ... 182
6.1.3.1 RQ 2A – Writing fluency ... 182
6.1.3.2 Exploration of effects of participants' characteristics on writing fluency ... 183
6.1.3.3 RQ 2B – Reading fluency ... 184
6.1.3.4 Exploration of effects of participants' characteristics on reading fluency ... 185
6.1.4 Research Question 3 – dictating fluency ... 186
6.1.4.1 RQ 3A1 and 3B1 – Writing productivity – TWW and TWD ... 186
6.1.4.2 RQ 3A2 and 3B2 – Writing accuracy in dictated texts – %CWS ... 187
6.1.4.3 Exploration of effects of participants' characteristics on dictating fluency ... 188
6.1.4.4 Exploration of effects of participants' device preference on dictating fluency ... 190
6.1.5 Research Question 4 – Writing motivation ... 191
6.1.5.1 Exploration of effects of participants' characteristics on overall writing motivation ... 192
6.2 Results of qualitative research strand ... 193
6.2.1 Research Question 5 – Challenges of STT in school ... 193
6.2.1.1 Reported challenges of STT for students with dyslexia in school ... 195
6.2.1.2 Supportive features for success of STT in school ... 198
6.2.1.3 Motives and attitudes of teachers/specialists ... 200
6.2.1.4 Benefits observed and reported by teachers and specialists ... 200
6.2.2 Survey of students' device preferences ... 201
6.2.2.1 Students' tool predilection ... 202
6.2.2.2 Students' own expectations for dictation in the future ... 202
6.2.2.3 Students' feelings while dictating ... 202

7. Discussion and limitations 204
7.1 Discussion of research results 204
7.1.1 Research Question 1 – Effect of STT on text quality and text length 204
7.1.1.1 Text quality 205
7.1.1.2 Text length 208
7.1.2 Research Question 2 – Effect of STT on writing and reading fluency 210
7.1.2.1 Writing fluency 210
7.1.2.2 Reading fluency 212
7.1.3 Research Question 3 – Handwriting vs. dictating 212
7.1.3.1 Comparison of modalities 213
7.1.3.2 Development of dictating fluency 214
7.1.4 Research Question 4 – Writing motivation 215
7.1.5 Research Question 5 – Challenges of implementation 216
7.2 Limitations 218

8. Conclusion and implications for future research and practice 220
8.1 General conclusion 220
8.2 Implications for future research 222
8.3 Implications for practice 224

9. Summary 226

References 229

Abbreviations 250

List of Figures 251

List of Tables 253

1. Introduction

1.1 Relevance of research on speech-to-text technology for students with dyslexia

Writing is a key competence and essential for academic and professional achievement. In formal schooling, writing remains the most common way of demonstrating knowledge and is an essential tool for acquiring knowledge. Moreover, people write to record information, regulate emotions, and persuade or entertain others. For everyday communication, writing has become more important than ever, as writing, using a variety of social networks and media, provides a powerful tool for keeping people connected (Arcon et al., 2017; Dockrell et al., 2015; S. Graham & Harris, 2020; S. Graham et al., 2016; Troia & Graham, 2017).

However, not all students acquire writing skills equally quickly and equally well. According to S. Graham (2019), many students with diagnosed educational needs in primary and secondary school do not meet even basic standards in writing. One subgroup of students with special educational needs are students with diagnosed dyslexia, a disorder in the acquisition of written language that affects a considerable part of the population (Scheerer-Neumann et al., 2010).

Over the course of a student's school career, the gap between pupils with poor writing skills and others widens, despite absolute learning progress, which limits these learners' ability to progress in school. Dyslexia is therefore, for many students, a substantial barrier to their future academic and professional achievement (Scheerer-Neumann et al., 2010; Svensson et al., 2019). So how can students with dyslexia overcome this barrier to acquire the key skill of writing and to have equal opportunities in school and in their professional careers?

To answer this question, it is necessary to differentiate between the writing processes and writing products; that is, depending on the particular field of writing research and writing instruction, the focus may be on either writing products or writing processes. In addition, an analytical separation regarding the hierarchical levels of both writing processes or products is important to investigate writing (Philipp, 2017). The above-mentioned educational and extracurricular writing activities use the term writing in the sense of text production, depending on the discipline, also known as higher-order processes. However, the term writing also applies to the transcription; or in other terms, the lower-order, to processes such as the mechanics of handwriting or typing to produce letters and words with a writing tool – on paper or on computer – while respecting the rules of the writing system one is using, i. e., spelling.

Motor skills and correct spelling are not ends in themselves but support the comprehensibility and readability of a text and affect text quality (Bachmann &

Becker-Mrotzek, 2017b; A. Sturm & Schneider, 2019). Non-automated mechanics of handwriting or typing and the orthographic processes of spelling place major demands on the working memory (Kellogg et al., 2013). Cognitive Load Theory by Sweller (1994) suggests that working memory resources are limited and that different processes compete for them. As a result, the writer's limited resources are not available for higher-order processes (Kellogg et al., 2013). Thus, if writers often have to solve orthographic problems during text production and frequently think about how an expression should be written, this impacts text quality in a negative way (A. Sturm & Schneider, 2019). This particularly applies to students with dyslexia. It is assumed that, due to non-automated lower-order writing processes, students with dyslexia lack cognitive resources for the development of the higher-order writing processes, in other words, text production (Haug & Klein, 2018).

According to Bachmann and Becker-Mrotzek (2017b), the term, text production, does not presuppose that someone producing text must have the ability to do the mechanics and spelling themself. Rather, it denotes the ability to create text in a form that can be re-used and transmitted, thus, in spoken or written form. In this sense, dictation is also a way of producing text. However, while dictating takes place in an oral modality, dictating needs, on a conceptual level, a concept of writing, meaning that the way of thinking and planning text corresponds to the written form (Koch & Oesterreicher, 1985).

Given the importance of writing in terms of producing text for multiple purposes in school, and the fact that lower- and higher-order writing processes can be dissociated, the question arises as to whether students with dyslexia would benefit from bypassing the lower-order writing processes in favour of the higher-order processes. Digital applications that enhance text production present a possible means of circumventing challenges such as those of orthography or handwriting, because they are assumed to free up cognitive resources for planning, structuring and revising a text. Technologies that are, firstly, aimed at improving functioning, and secondly, intended for people with disabilities, are called assistive technologies (ATs). Svensson et al. (2019) consider that ATs are now sufficiently effective to be a real alternative for students with writing difficulties. According to these authors, in the future, there will be at least two options for text processing and written communication: either people will read and write in the traditional way; or they will use technological tools to do so. Both speech-to-text technology and text-to-speech technology have, in recent years, transitioned into mainstream technology that can be easily accessed on computers, tablets or mobile phones (Almgren Bäck et al., 2023; D. L. Edyburn, 2015). The focus in this thesis research is on speech-to-text technology (STT), also known as dictating software, which is based on an automated speech-recognition system and assists in writing.

Legally, the use of STT, as one type of AT, falls under the right to compensation for disadvantage which, in Switzerland, is based on the Federal Act on the 'Elimination of Discrimination against Persons with Disabilities' (BehiG SR 151.3, 2002).

As dyslexia is classified as a language disability (Langenfeld, 2007), this law applies. The approval of a technological aid is, for example in the Swiss canton of Freiburg, listed as a possible measure to compensate for disadvantages (Staat Freiburg, 2023).

However, not all professionals in the educational system, e. g. teachers, speech pathologists or special needs teachers, perceive the use of a technology in the inclusive classroom in the same way. While some consider the use of a technology like STT as a compensational measure, in other words an AT only, others consider it as a common tool for learning or remediation. Researchers from a range of disciplines, including media education, writing research and special needs education, advocate a shift in this respect. As a result of rapid and significant changes in literacy and technology, they promote a recognition of these tools as common learning or educational tools (I. Bosse et al., 2018; Matre & Cameron, 2024). The emergence of new tools and the improvement of existing ones have impacts on both literacy and language modalities, especially writing processes and products. Therefore, research and schools, particularly the field of writing instruction, must adapt to these changes (Applebee & Langer, 2009; Leu et al., 2013). In the classroom of today and tomorrow, roles and expertise are changing, as teachers no longer necessarily know all the available technologies better than their learners. Instead of instructing on the use of tools, teachers in future, might have to orchestrate learning with such tools (Leu et al., 2013; Tate et al., 2016).

According to S. Graham (2019), today, students in the USA receive few specific instructions for writing on the computer or opportunities to practice, even though digital writing clearly predominates in writing inside and outside of school. In addition, school-related writing on computers strongly impacts school and career opportunities, which is particularly true for learners with writing difficulties, who tend to write less than others in school and in everyday life (S. Graham, 2019). Although the research literature does not yet provide a solid answer to the question of which media and methods to use, researchers see opportunities in digital media, specifically for struggling writers, to improve their writing (S. Graham, 2019; Regan et al., 2021; Schröder et al., 2022).

Arguments for compensating for inadequate writing skills can also be found in studies about writing motivation (e. g. Ekholm et al., 2018). People with dyslexia experience high levels of anxiety and stress (McArthur et al., 2000) and suffer from low self-esteem and social exclusion because of their difficulties. These feelings are likely to have a negative impact on motivation (Svensson et al., 2019). Meanwhile, it is well established that the motivational variables, writing attitudes and self-efficacy in writing, are important predictors for writing performance (S. Graham et al., 2017; Wright et al., 2021), which is even more the case for struggling writers (S. Graham et al., 2018). Thus, if negative feelings are reduced by an AT intervention and, in turn, writing motivation and self-efficacy are improved, this could, potentially, have a positive impact on writing performance. Svensson et al. (2019), investigating STT and TTS, consider the possibility of a positive impact of the application of apps

on writing motivation and general school motivation, especially for students with severe dyslexia. However, according to Matre and Cameron (2024), there is little evidence on the question. No study specifically investigating writing motivation with STT and on students with dyslexia is found in the research literature reviewed for this thesis.

Regarding the literature about STT as AT for students with poor writing skills[1], quantitative research studies report benefits in outcomes such as text quality, text length (MacArthur & Cavalier, 2004; Schüler, 2020), and number of ideas (Arcon et al., 2017), for students with learning disabilities and struggling writers, but not for students without difficulties. Haug and Klein (2018) report a transfer of writing strategies from one modality to another. Svensson et al. (2019) confirm transfer effects also for spelling, in a study on different types of AT for students with dyslexia in regular classrooms. However, some authors report mixed results, both on higher-order and on lower-order writing measures (e. g. Berner & Alves, 2021; Quinlan, 2004); and user experiences of STT draw attention to problems such as misinterpretation of spoken words, which can be exhausting for the users and lead to feelings of frustration or stigmatisation (Almgren Bäck et al., 2023; Brüggemann, 2018; Söderström, 2016). Furthermore, research literature from the field of media education indicates that the successful implementation of a tool is challenging for the teachers and schools (e. g. Kerres et al., in press).

Research contributions from Arcon et al. (2017), MacArthur (2009) and Liu et al. (2019) identify conditions for success. Accordingly, an STT tool will only deliver the desired benefits if the learners are motivated and willing to learn to use the tool and agree to use it consistently. In addition, the learners should receive a thorough introduction to become familiar with the tool. Once STT and other technological tools are introduced, they are helpful for teachers, because students become more independent (Svensson et al., 2019).

This thesis addresses the question of whether STT is an effective tool, both to assist and enhance writing, for primary school students with dyslexia and, if so, what teachers and specialists need to know for its implementation so that the benefits of STT enhance the students' learning. The current state of research with regard to the question of the effectiveness of STT for students with dyslexia contains some research gaps, which the following studies identify. Matre and Cameron (2024) argue, "there is a great need for more robust research on the use of speech-to-text technology (STT) in educational settings, especially on its effect on writing skills" (p. 1). Svensson et al. (2019) agree on the current lack of solid studies but also consider that there are no reliable quantitative and qualitative instruments to investigate the possible benefits of ATs. Investigating changes in written language abilities due to the use of a tool such as STT, compared to traditional teaching and writing approaches,

1 The expression 'poor writing skills' is often used synonymously with 'struggling writers' including different subgroups such as learning disabilities or dyslexia.

requires specific instruments, i. e., to measure gains in dictated texts. Furthermore, due to the limited empirical base in the field, there is insufficient evidence on how interventions should be designed in a school setting (Nordström et al., 2019). Matre and Cameron (2024) and Almgren Bäck et al. (2023) identify technical, contextual and emotional barriers that impede the application of STT for primary and secondary school students.

Research in this field also has to cope with the rapid pace of technological development (H. Schneider & Anskeit, 2017; Schüler, 2021). While the quality of the Automatic Speech Recognition technology (ASR) has improved substantially and continued technological developments in STT may potentially resolve many challenges (Filippidou & Moussiades, 2020), some questions remain that require further exploration. One question is the accuracy in different languages, because STT quality varies from one language to another (Matre & Cameron, 2024; Silber-Varod et al., 2021). STT as a mode of text production has not yet been fundamentally investigated in German-speaking countries (H. Schneider & Anskeit, 2017; Schüler, 2020). Furthermore, Matre and Cameron (2024) state, in their scoping review on the use of STT with struggling writers at secondary level, that little is known about the age that suits best for introducing the use of STT. These authors assume that STT does not suit all students, which Almgren Bäck et al. (2023) confirm. This observation, firstly, reinforces Schröder et al.'s (2022) claim that research should be more specific about the use of a tool by individual groups of students. STT has been studied for various groups of students, e. g. students with learning disabilities, struggling writers, students in second language acquisition, secondary school and college students, or students after traumatic brain injury (Arcon et al., 2017; Haug & Klein, 2018; MacArthur & Cavalier, 2004; Matre & Cameron, 2024; Noakes et al., 2019; Schüler, 2020). Secondly, no studies on STT appear to have been conducted specifically on students with dyslexia. Thirdly, while most of the limited number of studies investigate the usefulness of STT for writing in the dictating modality (e. g. Almgren Bäck et al., 2024) or differences between modalities, even fewer studies attempt to investigate changes in modality-independent writing and writing-related skills (Svensson et al., 2019). The present thesis project aims to fill these research gaps. In addition, the present research includes the question of the challenges met during implementation in schools.

1.2 Research goal and questions

In line with the above-elaborated research gaps, this thesis research investigates the effectiveness of speech recognition technology for primary school students with dyslexia. Situated in the field of intervention research, it addresses the following overall research question, according to methods of inferential statistics (Bortz & Schuster, 2010):

Is the use of STT-tools for students in fifth grade with dyslexia an effective method to improve text quality and other literacy skills and increase their writing motivation?

According to Cognitive Load Theory and motivation theories, as well as the state of the research in this field, this thesis hypothesises that text quality, and other literacy skills such as spelling and basal reading skills, improve, and that writing motivation increases, among students who use an STT tool.

Given the state of the research and the complexity of writing processes, challenges during the implementation in the classroom must, however, be expected. This project also aims to identify these, and asks:

Which challenges accompany the implementation of STT in the classroom?

The explanatory methods to investigate the effectiveness of the tool are, therefore, supplemented by an explorative approach. The mixed-methods approach in this project combines quantitative and qualitative research methods to provide insights into this topic and substantiate the scope of the results (P. Schneider & Niederberger, 2020).

1.3 Research outline

This thesis study investigates the questions above with a quasi-experimental study with a mixed-methods design. It examines the development of written language skills and writing motivation among Grade 5 students with dyslexia. The study takes place in the German-speaking part of Switzerland during the school year 2022/2023. The study includes a pre- and post-test and attempts to meet the conditions for a successful use of STT, i. e., a thorough software instruction for those using the tool, and specific writing instruction. Three groups are compared to each other. One group receives the software introduction and uses STT in their writing classes. This group also receives writing instruction by videos provided to the teachers. This group is compared to two distinct groups that write in a traditional manner, i. e., without STT. While these latter two groups write by hand, they differ in terms of the writing instruction they receive: one receives the writing instruction by videos; the other, conventional writing instruction by the teacher. At the end of the intervention period, interviews are conducted with the teachers, including special needs teachers and speech pathologists, who worked with the participants of the experimental group that use STT in their writing classes.

As there are no guidelines or manuals for the introduction of speech recognition software, appropriate materials needed to be developed for the student group and evaluated for use in this study. Teaching materials for writing lessons do exist, but due to the design of this study, it is important that the writing lessons in the experimental groups are comparable. Thus, specific materials for the present study were also designed and piloted in advance. While existing instruments can be used to as-

sess text quality, writing fluency and reading fluency, there is no suitable instrument in German to assess motivation to write. A questionnaire was thus translated from English, adapted, and evaluated prior to the study.

1.4 Structure of the thesis

This thesis is structured as follows. The first part of Chapter 2 focuses on the definition and importance of writing competence. Following this, relevant models are discussed that illustrate the complexity of writing. S. Graham's (2018) "Revised Writer(s)-Within-Community Model of Writing" conceptualises writing as a social activity within specific writing communities. It emphasises that writing does not occur isolated from the environment and includes the writing tool. The writing model by Hayes (2012) illustrates the separation of lower-order and higher-order processes. Olive's (2014) model addresses the complexity of the coordination of writing processes. The second part of this chapter focuses on dyslexia and its impact on school, work and everyday life. It then discusses ways of addressing dyslexia, such as intervention or compensation, with STT being one option.

Chapter 3, firstly, outlines major developments in technology over the last few decades. Secondly, the use of STT for students with dyslexia is discussed in a theoretical approach, namely from the perspectives of Cognitive Load Theory (Sweller, 1994), motivational theories (e. g. Bandura, 1997), New Literacies Theory (Leu et al., 2013) and media education (Kerres, 2018). Chapter 4 elaborates on the benefits and challenges of STT for writing. While substantial literature for students with dyslexia is scarce, studies from neighbouring fields can provide some useful information for the question of how students with dyslexia may benefit from using STT. This chapter also delves into advantages and challenges of this type of tool. According to the literature, such tools need to be carefully introduced and combined with effective writing instruction. Therefore, this chapter maps key elements of conditions for the successful application of STT.

Chapters 5 and 6 constitute the empirical research report. Chapter 5 presents: the subsidiary research questions to the above-mentioned overall questions in detail, the research design with the methodological approach of the quasi-experimental study and the interview study, including the field access, the participant recruitment and the sample for each, and the quantitative and the qualitative study component. Within the research design, the integration strategy of both components is also outlined. Due to the self-assignment into study groups, the comparability of the groups is presented in detail. This chapter also includes the rationale for the selection of the assessment instruments, and discusses adaptations made to meet the research purposes. The final section of Chapter 5 presents the hardware and software provided to the group using STT, and all instructional materials for the software instruction and writing instruction by videos. Chapter 6 reports all results for the inferential statistical and the qualitative content analyses, in order of the research questions, including

reporting on any breaches of the required statistical assumptions. The quantitative and qualitative data allow for complementary findings which are presented in the last part of each section.

Chapter 7, firstly, brings the results of the two components together and discusses these within the context of the literature. This first section also elaborates implications for further research and practice for the use of STT for students with dyslexia, followed, in the second section, by limitations of this study. Chapter 8 is dedicated to the thesis conclusions; Chapter 9 closes the thesis with a summary of the research.

2. Writing despite Dyslexia

In Chapter 1, the importance of skilled writing was emphasised as a key competence for school, professional and everyday life. This chapter elaborates the terms writing or writing competence and situates writing in its contexts. Writing is commonly known as a complex task, which is why, at this point in the thesis, the processes and components necessary to accomplish a writing task are highlighted. An overview of the development within the norm of writing and writing instruction is presented, which provides a basis to discuss the impairment of writing acquisition and possible measures.

Writing – a flexible, versatile and powerful tool – is undoubtedly very important for school and everyday life. However, S. Graham and Harris (2019b) also describe writing as a very complex competence and activity, as their following statement illustrates:

> Writers must juggle and master a commanding array of skills, knowledge, and processes, including knowledge about topic and genre; strategies for planning, drafting, revising, editing, and publishing text; and the skills needed to craft and transcribe ideas into sentences that convey the author's intended meaning (2019b, p. 19).

In addition to its multidimensionality and complexity, writing is also a social activity shaped by cultural, societal and historical factors which takes years of training and effective instruction (Girgensohn & Sennewald, 2012; S. Graham & Harris, 2019b).

Section 2.1 elaborates this task, commonly considered complex with its numerous processes and components. To the end of this overview, this chapter elaborates on the term, writing competence, by the presentation of three models. The first model presented is the Revised Writer(s)-Within-Community Model (WWC) by S. Graham (2018), which provides a broad perspective on writing within a writing community, such as in writing instruction. This model also considers the writing tool. The second model presented here is the current version of the writing model, by Hayes (2012) illustrating the elements of the internal processes that are important in writing. Although the differentiation of lower-order and higher-order writing processes is at the core of the discussion presented in this chapter, further cognitive processes such as working memory are also developed. However, these two models don't provide a detailed description of how the writing processes described are coordinated by Olive's. The coordination of these processes is considered to be of interest for the present thesis research, because the use of a technology might impact such processes and their coordination. Therefore, a third model, by Olive (2014, p. 173), "toward a parallel and cascading model of the writing system", is discussed and summarised. Sections 2.1.2 and 2.1.3 then elaborate key points of writing instruction including the writing tasks and writing assessment within the normally expected development of student writing abilities. On this basis, Section 2.2, next, focuses on the

impaired acquisition of written language and possible interventional, adaptational and compensational measures.

2.1 Writing competence

Philipp (2015) assumes that 'writing competence' is a challenging term to define. While the author assumes that the term 'writing' can be defined, he points out the difficulty of defining the term 'competence'. Becker-Mrotzek (2014) summarises the term 'writing' as commonly used to describe a written-language action that involves spatially and/or temporally extended text-based communication between a writer and a reader. Philipp (2015) describes writing, at its prototypical core, as a communicative action. By contrast, the term competence is difficult to define because there are several different understandings, e. g. cognitive competences, key competences or meta-competences, in research and commonly used in everyday life (Grabowski, 2014; Philipp, 2015). Harsch et al. (2007, 45, Transl. Flütsch Keravec) define writing competence[2] as the "ability to formulate texts in a way that is appropriate for the target audience and, depending on the writing goal, to inform precisely, argue convincingly or use language in an aesthetically accurate and creative way".

Writing competence requires various linguistic and cognitive resources as well as knowledge about writing and content (Becker-Mrotzek, 2014). Philipp (2015) distinguishes writing competence from a purely graphomotor or (meta-)cognitive action and proposes a narrow and a broader definition of writing competence. While the broad definition includes motivational factors as well as the social context and the individual's interaction with the developing text, the narrow definition focuses on the text production performance but still includes motivational factors.

With regard to students with dyslexia both definitions may be of interest. One the one hand, considering students' participation in the community by the means of writing, accordingly a broad perspective of writing seems important for considerations about whether or not to apply digital technologies as AT. On the other hand, instruction in literacy acquisition and remediation, often focussing single elements of writing, take a narrower perspective of writing. This perspective seems worthwhile when considering research questions about the effectiveness of digital technologies on writing such as in the scope of this thesis. It puts the writing skills and writing motivation of students with dyslexia at its core and focusses on text production[3]. Nevertheless, the supplementary research question in this thesis, about accompanying challenges for the implementation of a transcribing technology in school, invites a broader perspective because, for example, the social context, well-being

2 Not differentiating between competence and ability; Harsch et al. (2007).

3 Spelling and handwriting/typing are mentioned when it seems relevant to the understanding of (text) writing skills or their acquisition.

of students with dyslexia, or teamwork among teachers and specialists, all play an important role.

2.1.1 Models of writing

A scientific model is an approximation of a phenomenon (S. Graham, 2018) and is used to describe, explain and predict that phenomenon (A. Sturm & Weder, 2016). According to the perspective different models are possible (S. Graham, 2018). This thesis starts from a broad perspective, of writing within the community. Subsequently, the perspective is narrowed to the cognitive and motivational processes of writing, in the current version of the writing model by Hayes (2012), and then to the perspective of the model by Olive (2014) representing the interaction and coordination of processes during the writing process.

2.1.1.1 Written products, writers and the community: The Writer(s)-within-Community Model

S. Graham (2018) presents a model that he calls the Revised Writer(s)-within-Community Model (WWC), with the aim of improving understanding of how writing occurs. He illustrates the complexity of writing with two figures, one (Figure 2.1) putting writing at the core of the community, the other (S. Graham, 2018, p. 265) showing the cognitive mechanisms involved in writing.



Figure 2.1: Basic components of a writing community (s. Graham, 2018, p. 264)

By merging the sociocultural and cognitive perspectives, this model places the writing process within a large context, i. e., on a macrolevel, and emphasises the importance of social interaction and collaboration (Ekholm et al., 2018). All components of the writing community interact reciprocally, shown in Figure 2.1 by the arrows. The second circle, moving from the outside to the centre, shows that a writing community includes its purpose for engaging in writing, its members, i. e., writers, collaborators, audiences, its typical practices, its physical and social environments, and its own collective history. A language arts class, where students are learning to write, is one example of a writing community; but an online website where writers share and support one another's writing efforts is also one, or a parent and a child who write stories together to entertain someone (S. Graham, 2018). According to the third circle in Graham's WWC, "writing is simultaneously shaped and constrained by context, the capabilities, and perceptions of writers and collaborators, and the interactions between the two" (2018, p. 258). These multiple interacting characteristics of the community simultaneously influence, at its core, the written product and how writers and possible collaborators achieve the desired writing goals, which writing tools they apply, and which actions they use (S. Graham, 2018). Tools for writing include not only paper and pencil or typewriters, but also dictation with a human or technologies such as word processors and speech recognition technology (Gabrial, 2008; S. Graham, 2018).

Within the writer, the cognitive, affective and motivational resources interact among themselves but also with all resources within the community. These include: 1) long-term memory resources such as knowledge and beliefs; 2) control mechanisms such as attention or working memory; 3) production processes such as transcription or ideation; and 4) modulators, i. e., emotions, personality traits, and the physical state (S. Graham, 2018).

S. Graham (2018) takes, with this second part of the model, a closer look at the resources within the writer that influence the written product. Due to similarities with other models, S. Graham's second part is not presented here. Hayes (2012) and his colleagues have proposed models to illustrate the resources and interacting processes in writing since 1980, firstly with the Hayes-Flower (1980) model. While this model presented adults' writing, more recent work also addresses children's expository writing (Hayes & Olinghouse, 2015). "Although the different models of the writing processes cannot be regarded as equivalent" (Olive, 2014, p. 174), several models conceive writing as an engaging process that includes different levels of processing, in which conceptual, linguistic and motor processes must be orchestrated (Olive, 2014). The following section aims to disentangle this engaging process by highlighting the most relevant processes and their development for this thesis.

2.1.1.2 Writing processes in models including development of processes

While the WWC model contains the writing products, the models from Hayes and colleagues intend to describe the writing processes. According to Philipp (2015), the

distinction between both in the field of competences is beneficial for understanding and addressing them coherently in research and instruction. The two approaches should therefore be seen as complementing rather than competing with each other (Philipp, 2015; A. Sturm & Weder, 2016).

Figure 2.2 shows that the model by Hayes (2012) divides the writing process into three levels. Several factors, processes and functions are involved in writing (Hayes & Olinghouse, 2015). The components considered central for this thesis are elaborated and complemented by consideration of writing development in children and adolescents, since the (2012) model describes[4] writing in competent adult writers.

The three levels of the model, from the top to the bottom, are the control level, the process level, and the resource level. The control level includes factors that shape and direct the writing activity. Its components are 'motivation, goal setting, current plan' (both further mentioned in Section 2.1.2 about writing instruction) and 'writing schema'. 'Motivation' to engage in writing, and to maintain throughout the writing process, is an important component to make writing happen at all. However, Hayes and Olinghouse (2015) state that hampered transcription skills limit writers from engaging in writing, especially in long essay writing. This is therefore interpreted as an interaction between motivation and transcription, which illustrates that the processes of different levels and resources interact with each other, although the model in Figure 2.2 contains no arrow between the levels. Since students with dyslexia have, per definition presented below in Section 2.2.1 difficulties in transcription, dyslexia might reduce writing motivation, which is thus discussed in detail in Chapter 3, Section 3.3.2.

'Writing schemas' represent the writer's knowledge about strategies, as well as knowledge of text properties such as genre, length, format and tone. Schemas guide experienced writers in creating a variety of genres, but they also support the construction of written plans and guide the carrying out of revisions. Included in these schemas are knowledge about how the writer can generate ideas or how to structure a text. The various genres require different schemas which are acquired at different times and rates. Parts of genre knowledge may be taught and learned piecemeal. The use of schemas is a conscious activity that is learned through years of instruction and experience in reading and writing (Hayes & Olinghouse, 2015). Writing schemas are not intended to be automated processes (Hayes & Olinghouse, 2015); but according to cognitive load theory (see Chapter 3, Section 3.3.1), previously acquired schemas can be automated to some extent, freeing working memory space (Paas et al., 2003). The *process level* includes, on the one hand, the internal writing processes and, on the other hand, the environmental factors, which influence each other. While the writing processes, one component of the process level, are discussed in the following

4 For full details, please refer to further literature. See Hayes (2012); Hayes and Olinghouse (2015); A. Sturm et al. (2017).Klicken oder tippen Sie hier, um Text einzugeben.

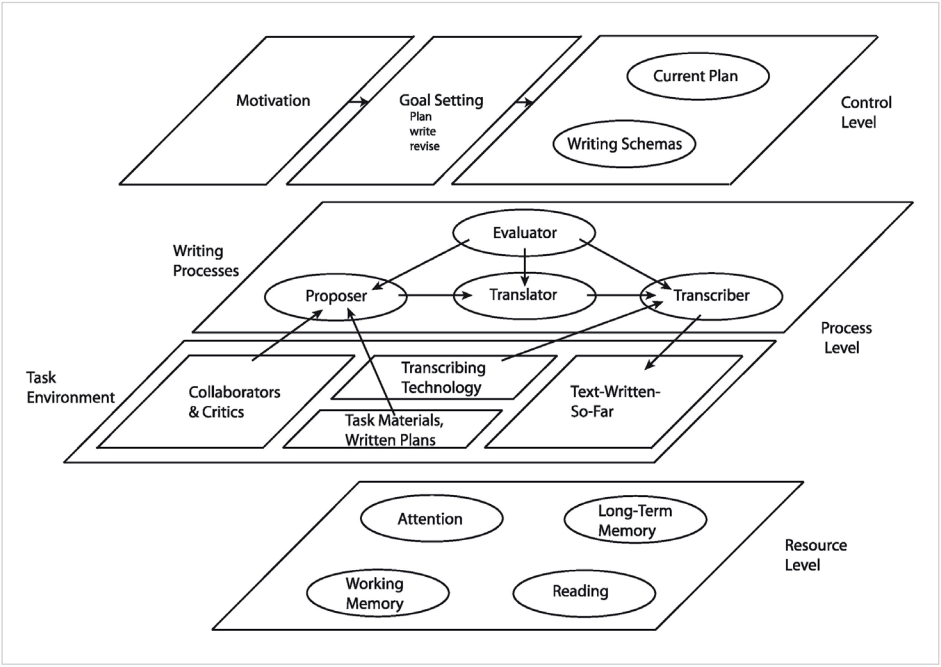


Figure 2.2: Current version of the writing model (Hayes, 2012, p. 371)

paragraphs, the other component, the task environment, is discussed within Section 2.1.2 on writing instruction (Hayes & Olinghouse, 2015).

Writing processes represent the internal mental processes going on in writing activities. These are not observable by themselves, but by combining the four internal writing processes, an outcome in the form of a written product can result. The four internal writing processes are, according to this model, the 'proposer', the 'translator', the 'evaluator', and the 'transcriber' (Hayes & Olinghouse, 2015).

The 'proposer', in Hayes (2012), suggests ideas for inclusion into the text, which are triggered by stimuli such as the sensory environment, memory, goals, or collaborators. These ideas are then processed by the 'translator'. The ideas are translated into grammatical strings of language; in other words, the translator translates nonverbal ideas into verbal form (Hayes & Ann Chenoweth, 2007). The 'transcriber' applies rules of spelling and good written form, such as well-formed letters and appropriate capitalisation and punctuation, to these grammatical strings and turns them into written text (Hayes, 2009). This process plays a critical role in the development of children's writing in early school years (Hayes, 2012) and also for children with learning difficulties (A. Sturm et al., 2017). "Successful transcription requires skills in forming letters (handwriting or typing), in spelling, in capitalisation, and in punctuation" (Hayes & Olinghouse, 2015, p. 487). This is why this process requires a significant number of resources and is described as the major bottleneck limiting writing fluency (Hayes & Olinghouse, 2015). In contrast to Hayes' (2012) model, writing fluency is distinctly

illustrated in the translated and adapted model for German by A. Sturm et al. (2017, p. 85). Handwriting or typing and spelling are two entities that form, together with the retrieval of linguistic knowledge, writing fluency. Retrieving linguistic knowledge from the long-term memory refers to the translator in Hayes' model, shown in Figure 2.2. A. Sturm et al. (2017), which describe the development of these skills with a focus on their automatization, also known as 'lower-order' writing processes (Philipp, 2015). In contrast to the higher-order writing processes, these processes can and must become more automatic to relieve the burden on the cognitive system (A. Sturm et al., 2017), in other words, to require less working memory and interfere less with other writing processes (Hayes, 2012). Therefore, writing fluency plays a mediating role between the higher- and lower-order processes, i. e., between the control and process levels (Y.-S. G. Kim et al., 2018). Hayes and Olinghouse (2015) cite research literature showing that specific training in handwriting or spelling increases the length and quality of texts. However, the acquisition and automatisation of these transcription skills take years and training (Bazerman et al., 2017), and these skills are mode-specific, meaning that the transcription training must also be mode-specific (Hayes & Olinghouse, 2015). Regarding spelling and orthography, the orthographic skills are, according to the developmental model of written language acquisition by Frith (1985)[5], acquired last, after logographic and alphabetic skills. Frith assumes that orthographic knowledge can only be transferred to the spelling output system when it becomes precise enough; in other words, internal representations must be exact in terms of letter-by-letter detail. However, it is important to point out that the application of the automated spelling process is only of interest in the transcription process, whereas the revision process needs access to rule knowledge, targeted rehearsals and correction strategies, which are particularly important in German (A. Sturm et al., 2017).

The 'evaluator', presented in the model by Hayes (2012) with arrows to the three other processes, examines their outputs and judges their adequacy. The evaluator can act on ideas, on translation into language, or on transcription, by rejecting or keeping stages of the developing written product. Not illustrated in the model, but important parts in writing, are 'planning' and 'revision'. Both are complex activities, each with its own writing schema and each making use of some of the basic processes to accomplish their own special goals. The processes of planning and revision differ between children and competent adult writers. While developing writers spend little time on planning, the relation between planning and composing develops with age. While adults spend time on revision while composing, this type of revision, but also post-composition revisions, are relatively rare in children, even if children are able to revise when specifically asked to do so (Hayes & Olinghouse, 2015).

5 The orthographic skills refer to the units that are internally represented as abstract letter-by-letter strings. These units make up a limited set that can be used to create, by recombination, an almost unlimited number of words. The orthographic strategy is distinguished from the logographic and the alphabetic ones by Frith (1985).

The *resource level* in Hayes' (2012) model contains four components, 'attention', 'long-term memory', 'working memory' and 'reading', which are used in writing and many other activities (Hayes & Olinghouse, 2015). In contrast to the first model by Hayes and Flower (1980), the model of 2012 places 'reading' as a resource. Editing and revising typically involve the repeated reading of the text written so far. Thereby, the reading process depends on the writer's goals and the specific task (Hayes & Olinghouse, 2015). According to Linnemann et al. (2022) and R. D. Abbott et al. (2010), reading and writing are closely related, because they are based on comparable cognitive skills and common knowledge (Linnemann et al., 2022). 'Attention' "is the ability to maintain focus on a task in the face of distraction" (2015, p. 486). Attention, also referred to as 'executive function' or 'executive control', is essential in writing high-quality texts. 'Long-term memory' consists of topic knowledge, schemas, linguistic knowledge, events etc. (Hayes & Olinghouse, 2015). Long-term memory has no known limits of capacity or duration. A signal from the environment can trigger the transfer of information from the long-term memory back to the working memory, where the appropriate action is executed. Regarding writing, this applies, for example, in the activation of writing schemas (Sweller, 2023), discussed in detail in Chapter 3, Section 3.3.1 about cognitive load theory. The 'working memory' is presented as a short-term memory system with separate storage for verbal and for spatial information, which is used to handle information requirements needed to carry out tasks (Hayes & Olinghouse, 2015). According to their paper, 'The phonological loop as a buffer store: An update', Baddeley and Hitch (2019) state that this perception has lacked further development, and the authors elaborate a multicomponent model of working memory with a number of inter-related buffer stores. The authors propose the current working memory model as a hierarchical structure based on a series of interacting buffer stores (Figure 2.3).

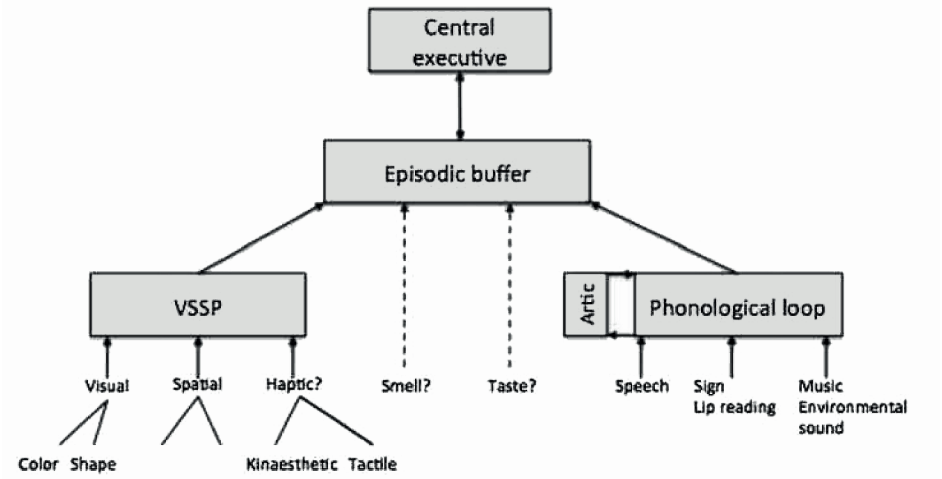


Figure 2.3: Current version of multicomponent model of working memory (Baddeley, 2012, p. 23)

This model retains the episodic buffer, the two elements of visuo-spatial sketchpad (VSSP) and the phonological loop, as well as the central executive. In contrast to previous models of working memory, it distinguishes between the purely attentional function of the central executive and the capacity for passive storage of the episodic buffer. The buffer represents the aspects of the visuo-spatial and phonological subsystems that are accessible to conscious awareness. These buffer stores are potentially able to combine information from two or more sources and are likely to process information. This applies for the sketchpad and the phonological loop. Regarding the latter, the authors assume it to provide a merging point for language-related material. It processes not only sound and phonology but also associated language-based features such as lip reading. These types of input have different but relatable codes and different input rates. This allows the input to be bound into more complex verbal linguistic episodes. Combining codes and coordinating input rates probably require temporary storage. This update on the phonological loop as a buffer store illustrates, on the one hand, that the complex processes executed during writing are not yet completely known. On the other hand, the model offers an account of how memory and attention combine to facilitate complex cognition (Baddeley & Hitch, 2019). In this sense, and regarding the research question in this thesis about transcription technology, the question is raised as to the extent to which the transcription technology affects the demands on the storage and the coordination of processes in the working memory. At this point, it is assumed that the demands of handwriting and thinking about spelling when producing text put a different strain on working memory than in dictating text. While the model by Hayes (2012) illustrates the multitude of interacting components and processes, however, it does not delve into the question of how these processes are coordinated. Therefore, an overview of the model by Olive (2014), about the parallel and cascading architecture of the writing system, which is based on studies that examine the impact of handwriting skills on higher-order writing processes, is presented next and concludes this section about writing models.

2.1.1.3 Toward a parallel and cascading model of the writing system – Olive (2014)

Olive's (2014) article, 'Toward a parallel and cascading model of the writing system', takes numerous proposals for models into account and proposes a conceptual architecture of the writing system, which is a domain-specific model. The model (Figure 2.4) specifies that the processes in writing do not take place successively but that they partially overlap (Linnemann et al., 2022).

The conception of this model is in line with incremental speech production models such as Levelt's (1989) model which offers, according to Olive (2014), a more sophisticated model for speaking than that which exists for writing. The author proposes a model that contains three levels of processing: the conceptualising level, the formulation level,

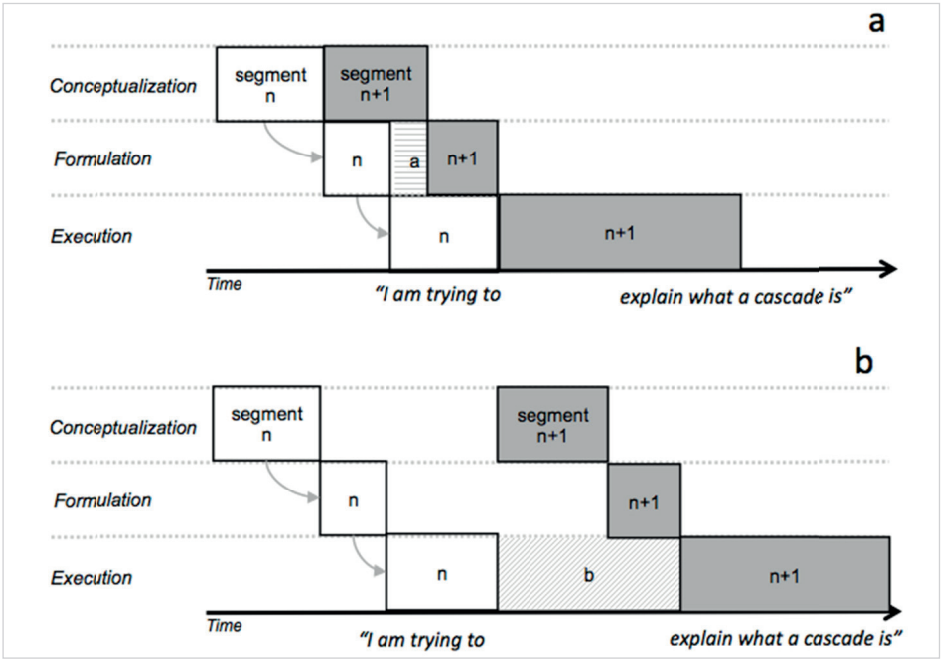


Figure 2.4: Schematic representation of writing processes coordination (Olive, 2014, p. 178)

and execution level. These levels are organised hierarchically, from more central cognitive processes, i. e., conceptual, to more peripheral ones, i. e., execution processes. Olive (2014) calls this flow, from the centre to the periphery, 'cascading'. During the writing process, the three levels operate simultaneously on different segments, illustrated in the upper part of Figure 2.4, labelled as (a). Regarding the production of written language, this corresponds to converting ideas into well-elaborated language before transforming it into written text via hand and finger movements. In addition to the cascading process, Olive (2014) assumes that different segments of text are processed in 'parallel', when sufficient cognitive resources are available. He argues (2014, p. 176):

> if writing processes were only sequentially coordinated, then writers would first plan a segment of text, next transcribe it, and they would not be able to prepare the following segment of text until the handwriting of this segment of text was finished"

He illustrates such sequential processes in the lower part of Figure 2.4, labelled as (b). By contrast, observations gathered from several studies show, firstly, that language preparation occurs not only during (handwriting) pauses[6], and secondly, that rapid interactions between planning and translating observed in skilled writers are possible. Therefore, Olive summarises the conceptualisation of writing as follows (2014, p. 177): "Each segment of text is sequentially processed from central to peripheral processes, and different levels of processing operate simultaneously on different

6 The light grey field with the letter b in the lower model 'b' indicates the writing pause.

segments of text". From the top down, operations at a particular level may interact with a subsequent level, because information cascades between levels of processing. At the same time, bottom-up interferences are possible, as operations at low levels of processing may affect higher levels because several levels of processing operate simultaneously. For example, the time taken to physically write down thoughts might cause the writer to forget the ideas on the conceptual level or formulations. The observation that handwriting might affect quality of composition in these two ways was already described in 1987 by Bereiter and Scardamalia, which is therefore in line with Olive's (2014) model. As a result of a particular need at a particular level of processing, or due to less experience in writing, fluency decreases or becomes more hesitant because writers focus on a specific level of processing and therefore adopt a more sequential coordination of the writing processes (Olive, 2014).

Working memory and the executive functions, although domain free, in this model receive the role of mediators for process coordination[7]. The processing demands of each level, placed on working memory capacity influence to what extent the levels of processing are parallel. Observations of children and adults, the former preferentially adapting a step-by-step strategy with pauses to plan further text segments, the latter tending to prepare longer text segments, suggests a development from sequential to more concurrent activation due to practice and age (Olive, 2014).

In summary, several points of these three models influence the present thesis and its research approach. S. Graham's (2018) WWC illustrates that, in writing research and instruction, the environment must be considered, including among many other factors the writing classroom and the writing tools. Hayes' (2012) model presents many processes and components that interact. The application of a transcribing technology potentially influences lower-order writing processes. Subsequently, working memory resources might be freed and possibly used by higher-order processes. At the same time, an application that acts on transcribing potentially interacts also with motivation. Olive's (2014) model illustrates that different levels and different segments are processed simultaneously. At this point, it is therefore assumed that the transcribing technology impacts the simultaneous processing because the type of input and the input pace vary, which thus needs to be considered.

2.1.2 Writing instruction and writing task

Girgensohn and Sennewald (2012) wrote that to become a virtuoso at writing takes many years or even decades; and S. Graham and Harris (2019b, p. 24) state, "Writing is hard work and learning to write well is even harder". To become a skilled writer, learners need writing instruction that allows them to flourish. This instruction must make sure that students acquire the skills, strategies, and knowledge they need to

7 The light grey field with the letter a in the upper model 'a' represents a period where buffering is necessary while the following processes get ready to operate.

meet the challenges on the path to becoming a skilled writer (S. Graham & Harris, 2019b). The quality of writing tasks is also an important component of writing instruction, especially for motivation and engagement (Bruning & Horn, 2000). The key points of both writing instruction (Section 2.1.2.1) and the task environment (Section 2.1.2.2) are elaborated in the following.

2.1.2.1 Writing instruction

The models presented above provide a roadmap for writing instruction (S. Graham & Harris, 2019b)because both writing instruction and writing development should consider the context, as well as the role of cognition and motivation in writing, to understand them adequately (Bazerman et al., 2017). This section provides an overview of evidence-based writing instruction, referring by writing instruction to the whole spectrum of writing, i. e., not only lower-order processes such as spelling or higher-order processes such as writing strategies. This section provides the foundation for the selection of an adequate method for the present thesis, which will be presented in Chapter 5, Section 5.4.3.

A supportive writing environment, where the teacher supports students and students support one another, is an important basis for effective writing (S. Graham et al., 2015), both according to the scientific literature and teacher-based best practice (S. Graham & Harris, 2019b). Scientifically supported practices that stand for a supportive writing classroom are, firstly, when teachers reinforce a positive feature of students' writing. However, such praise must be specific. Secondly, setting clear, specific and reasonably challenging goals improves the quality of what students write. Thirdly, instructional arrangements where students write together improves the quality of their writing. This applies to planning, drafting, revising and editing, but it requires specific directions and guidelines for what they will do when working together, and it should directly teach them how to apply these procedures (S. Graham & Harris, 2019b; S. Graham et al., 2011; MacArthur, 2019). Fourthly, the most critical element to grow as writers is, according to S. Graham and Harris (2019b), that the students write frequently and regularly. However, Applebee and Langer (2009) show that American students spend surprisingly little time writing in school, and that writing assignments often are short. Philipp (2019) assumes, despite fewer available studies, a similar picture in German-speaking areas. Fifthly, a supportive writing environment also requires consideration of tools that students use when they write. While, for example, word processors have several advantages over writing by hand, digital writing is still relatively less used for composing in school, compared to the everyday and workplace habits of the 21st century (S. Graham & Harris, 2019b).

Philipp (2017) emphasises three areas of instruction and/or support that are especially efficient to improve writing: highly demanding, higher-order writing processes (planning and revising); lower-order writing processes (writing fluency); and relieving writing processes (by technology or human assistance). Regarding Philipp's first area

mentioned, scientific experiments support the practice of explicitly teaching writing strategies to students. The strategies of planning, drafting, evaluating and revising text strongly improve the quality of writing. This applies to students of all grades, from elementary to high school (S. Graham & Harris, 2019b; S. Graham, McKeown, et al., 2012; Philipp, 2017). Writing strategy instruction is, at the most basic level, to explain the purpose and rationale of a strategy, including when and where to use it. Modelling the strategy, ideally several times, and providing students with assistance in applying the strategy, are further possibilities of strategy instruction. The aim is to use these strategies independently and effectively. Facilitating the continued and adaptive use of such strategies by explaining, modelling and providing guided practice is enhanced when students are shown how to regulate the learnt strategies. Therefore, the impact of a strategy has to be monitored, which requires the teacher to set goals and evaluate them (S. Graham et al., 2015). In addition to writing strategies, to support this first area named by Philipp (2017), it is especially important for writers to acquire two types of knowledge: topic knowledge and knowledge about the text type, i. e., genre. Prewriting activities such as brainstorming, reading or other informational-gathering procedures help students to acquire information on topic knowledge to write about. Teaching students about the characteristics of specific text types and providing them with good models for the expected genres, are scientifically based best practices toward this end (S. Graham & Harris, 2019b; S. Graham et al., 2015; Philipp, 2017).

The second area of writing instruction is, according to Philipp (2017), lower-order writing processes. It was emphasised in Section 2.1.1.2 that a lack of fluency in handwriting or typing and spelling interferes with other writing processes (Bereiter & Scardamalia, 1987). Effective instruction is therefore essential. Handwriting, or typing and spelling, together result in writing fluency, which Linnemann et al. (2022, p. 193) define as the "automatized ability to produce legible letters, correct words and, on the basis of information in lemmas, syntactically correct phrases/sentences (= Transcription fluency) as well as locally coherent text (= Text generation fluency) at a reasonable pace with sustained attention." These foundational writing skills must be taught, and such instruction has been shown in writing research to be efficient (S. Graham & Harris, 2019b; S. Graham et al., 2015; S. Graham & Santangelo, 2014; Santangelo & Graham, 2016). Philipp (2017) summarises as key examples, fostering writing vocabulary, combining sentences, and repeated short training in transcribing skills, that have shown significant results with moderate effect sizes resulting from meta-analysis research (e. g. Gillespie & Graham, 2014; S. Graham & Santangelo, 2014; Santangelo & Graham, 2016). The only approach found with a negative effect on text quality was traditional grammar instruction (Philipp, 2017).

Regarding the third area of instruction and support, Philipp (2017) names several efficient measures that relieve the writing process, all of them with moderate to strong effect sizes in meta-analyses. Such an example is cooperative writing. Cooperative writing means that, usually, two persons plan, transcribe and/or revise together. In addition, Phillipp indicates, especially for young adults, dictating with

a human scribe or a device to provide effective assistance. Formative feedback[8] by humans is also a possibility with positive impact on text quality, whereby feedback from adults is the most effective, followed by peer-feedback and self-assessment. Until Philipp's (2017) publication, feedback by a computer had only a small effect.

2.1.2.2 Task environment and writing task

The task environment is, next to the writing processes, the second component of the process level in Hayes's (2012) model. The task environment includes the immediate physical and social surroundings that influence writers when they write. Examples of the physical context of the writing task are task materials, the transcribing technology, and the text-written-so-far. The social task environment includes all of the inputs, for example, from peers or teachers, but also any background conversations and social distractions in a classroom (Hayes & Olinghouse, 2015). However, the writing task itself also needs attention (Bruning & Horn, 2000; Feilke, 2017). Although Feilke (2017) states that, as a response to the PISA[9] studies, the writing task itself has recently been receiving more attention in writing research and writing instruction. A. Sturm (2017) writes that practice-orientated publications and intervention studies rarely go into detail about the writing tasks used as part of the instruction. Bachmann and Becker-Mrotzek (2017a) address this question by asking about the conditions of instructional writing settings and writing arrangements that make the writing activity particularly effective for learning. Since solid research regarding this question is scarce, they base their approach on interdisciplinary quality criteria for effective instruction and the available theoretical models about writing processes and writing acquisition. Important in their reasoning is that any oral or written action is a form of communicative problem-solving. Therefore, a writing task must initiate the motivation to solve a communicative problem by forming and executing a plan in a written form and with the particularities of writing. According to these authors, four points are particularly important for meaningful writing assignments: 1) the text to write needs an identifiable function which is recognisable for the students, which must also identify the communicative problem and know the goal and the addressee(s) of the text, to make meaningful decisions regarding genre, content, linguistic, style etc.; 2) students need the opportunity to acquire the necessary linguistic and content knowledge to know how and what they can write about; 3) the social context discussed above must give students the opportunity to write their text with social interaction; and 4) students need the opportunity to check the effect of their text on the reader, as they are used to doing in oral communication. When these four conditions are met, Bachmann and Becker-Mrotzek (2017a) conclude that writing tasks

8 In contrast to summative feedback, formative feedback is not only given at the end of the text production process but early and throughout the process, see Philipp (2017).

9 PISA: Programme for International Student Assessment, see Wallner-Paschon and Haider (2002).

can be defined as 'profiled'. The authors explain the term 'profiled task' as being one that embeds the learning process in an authentic and social context in which writing has a recognisable meaning for the students.

While Bachmann and Becker-Mrotzek (2017a) focus on the task and its effect on learning to write, Boscolo and Gelati (2013) write from a motivational perspective about the task. They argue that the interest in the writing task is important for their motivation to write and thus the engagement and success in writing. However, they emphasise that finding an interesting topic, although it might be a good starting point, is not enough to engage students in writing. Despite this different perspective, the authors suggest some similar points as the above-mentioned authors. Firstly, they assume that the students need the awareness that writing on that topic is worthwhile, as well as awareness of the multiple functions of writing such as writing as a tool for learning. Secondly, what is written, individually or in collaboration, should have a value or relevance for the students. At this point, it is estimated that Bachmann and Becker-Mrotzek's (2017a) conditions for a profiled writing task meet the latter condition of relevance by Boscolo and Gelati (2013). It is thus concluded that a profiled writing task is both motivating and engaging at the same time.

2.1.3 Assessment of writing skills

Due to the complexity and multidimensionality of writing, writing performance is one of the more challenging skills to assess (Bouwer et al., 2024; Hopfenbeck, 2019). A text can be well or badly written because of its linguistic features and/or because of features such as fulfilling communicative purposes and/or well-developed or logical structure (Bouwer et al., 2024). It is therefore important to consider the dimensions of writing and to specify the goal and the focus of the assessment.

In this thesis, the first focus is set concerning the writing processes and the genres. Regarding the lower-order writing skills, multicomponent writing fluency[10] is prioritised over the single subskills, handwriting/typing and spelling. Regarding the higher-order writing skills, the focus is set on narrative writing, since each text genre demands adapted assessment (Brouwer, 2012). The second specification of the focus is the aim of the assessment. There is no doubt that writing assessment within writing instruction should, on the one hand, meet the claims of Bazerman et al. (2017) for fairness and authenticity and, on the other hand, foster learning, which would speak in favour of formative assessment[11] (S. Graham et al., 2011). Nevertheless, this

10 'Writing fluency' is, in this thesis, understood as a multicomponent ability comprising productivity and accuracy measures, in contrast to a narrow definition taking into account only the motor component, i. e., fluent, legible handwriting; see A. Sturm and Weder (2016).

11 'Formative assessment' is understood as an assessment for learning with the aim of adjusting instruction to maximise learning, and stands in contrast to 'summative assessment'; see Wilson (2019).

section focuses on the context of interventional research and therefore on evaluating the degree to which students have progressed, which is referred to as summative assessment. Results of summative writing assessments are often set in relation to one of the three comparative values: a reference sample, such as the class or an age-group standardised sample; with one's performance at an earlier time point; or in relation to a criterion (Lenhard & Lenhard, 2017). Lastly, unlike the diagnostic assessment of potential special needs in Section 2.2, this section addresses assessment in the sense of the educational diagnostics of writing (Gebhardt, 2023).

2.1.3.1 Assessment of lower-order writing processes and products

The basic subskills of handwriting or keyboarding and spelling can each be assessed in isolation. Test tasks that do not require text production can be used for this purpose: examples are copying and dictating tasks (A. Sturm et al., 2017), and dictation with gaps to fill in such as the "Hamburger Schreibprobe" (HSP) (Lenhard & Lenhard, 2017). In German, some standardised assessment tools exist for the lower-order writing skills (Winkes & Schaller, 2022). This means that results can be compared to a norm group (Gebhardt, 2023).

The analytical rating of a text presents a further possibility to analyse a product according to pre-defined criteria (A. Sturm et al., 2017), with one example in German being the "Oldenburger Fehleranalyse" (OLFA) (Lenhard & Lenhard, 2017). This type of rating allows statements about criteria defined prior to the assessment, but not about writing fluency, as this method generally does not contain strict time restrictions. Therefore, to assess the multidimensional construct of writing fluency, one can examine either the writing products or the processes. The investigation of the writing processes can, for example, be of interest in order to learn about the challenges a writer encounters during transcribing, formulating or revising (A. Sturm et al., 2017). Some of these processes occurring during writing can be observed or asked for, e. g. by think-aloud protocols and/or retrospective reports (Piolat et al., 1999; Weder, 2021). Other processes require equipment such as tablets, smartpens or keystroke-logger software – each with or without eye-tracking (A. Sturm et al., 2017). The equipment then allows one to analyse so-called 'bursts', which are periods of handwriting or typing separated by pauses. Length and structure of bursts potentially allow drawing inferences about the processes occurring, i. e., planning, translating and revising (Alves et al., 2008).

Writing fluency can also be assessed in writing products with so-called curriculum-based measurement (CBM). CBM-writing reflects both productivity and accuracy (Dockrell et al., 2015). Students receive mainly story starters as a writing prompt. The writing time is strictly limited, usually between three and ten minutes, with a short period for mental planning, often exactly one minute. The time for planning intends to generate the main ideas before the writing period in order to focus on formulation and transcribing during the writing process (Hosp et al., 2016). CBM aims to record progress in basic skills, as well as small increases in performance, and

to assess writing with a stronger alignment with teaching and the curriculum than standardised tests (Saddler & Asaro-Saddler, 2013; A. Sturm et al., 2017). CBM assessment allows one to record the individual learning of a student and to compare it with an individual norm, which proves to be superior for learning than comparison with a social norm such as an age-group norm (A. Sturm & Weder, 2016). Hosp et al. (2016) have developed norms for writing assessment, which can provide useful information to monitor writing development and/or interventions, not least for research. However, there are currently no standards for writing in German.

Different categories of variables can be collected by CBM-Writing. The first category is productivity-dependent measures, which means that they depend on the length of text, i. e., total words written (TWW), words spelled correctly (WSC), correct writing sequences (CWS), and incorrect writing sequences (IWS). The second category is productivity independent, meaning that it is a percentage value independent of text length, i. e., percentage of correct writing sequences (%CWS) and percentage of incorrect writing sequences (%IWS). The last category is a value for accuracy and fluency. To obtain this, the incorrect writing sequences are subtracted from the correct writing sequences (CIWS) (Malecki & Jewell, 2003). This last value corresponds to the above-discussed understanding of writing fluency. By contrast, the TWW does not include the accuracy of the produced words. Nevertheless, the total number of words produced in a text has been found to be a reliable and valid measure of overall writing ability, in some previous studies (Troia et al., 2019). This could speak in favour of collecting more comprehensive data within the framework of CBM.

2.1.3.2 Assessment of higher-order writing products

For the assessment of writing skills beyond the level of lower-order writing products and processes, there is an array of different rating procedures, varying in the amount and type of information they provide (Feenstra, 2014). Wolcott and Legg (1998) state that writing assessment is discussed by several terms that are often almost opposing, such as the above-mentioned terms, formative versus summative. Two of these polarities, the standardisation of the assessment task and the scoring procedure, will be discussed here, because they are estimated as useful to give an overview of the various options for writing research and instruction.

Wolcott and Legg (1998) present the tasks of writing assessment on a scale with two polarities. At the most extreme polarity in writing assessment stands the norm-referenced standardised tests, also referred to as indirect testing. An example of indirect testing is the evaluation of correct and incorrect sentences. At the other end are performance assessments or authentic assessments, where students demonstrate competence by producing text rather than selecting an answer, and where projects are prioritised rather than items. However, the tasks that involve text production can vary greatly. For example, a portfolio assessment is completed over a period of time and is a purposeful collection of student's work that exhibits efforts,

progress or achievement. A specificity of performance assessment is that they attempt to consider not only writing products but also writing processes. By contrast, a more direct assignment to produce text is an 'on-demand writing task' such as a direct assignment to produce an essay on a specific topic. Such impromptu essays are often timed and used for a variety of examination purposes, because the controlled conditions allow comparisons to be made across or within reference samples (Wolcott & Legg, 1998).

With regard to the writing product, any type of text assessment, portfolio or impromptu essay, requires 'informed judgement' rather than mechanical scoring. Neumann (2017) describes the scoring procedures – referred to above as informed judgement – also on a scale, placing the analytical ratings on one side and holistic ratings on the other. There are variants with combined procedures between the two poles. Wolcott and Legg (1998) name such combined variants, e. g. ratings with an exemplary essay, as a 'reference point', or 'primary trait', which means that a text is judged on how well a goal is achieved. The following paragraph names and highlights advantages and disadvantages of different scoring methods. It also considers possibilities in German.

'Holistic ratings' give an overall quality evaluation by jointly considering multiple facets of an essay. The key advantage of holistic ratings is the relatively short amount of time needed for the scoring (Feenstra, 2014; Lai et al., 2015). While Feenstra (2014) estimates the scores as relatively reliable, Lai et al. (2015) question their consistency because individual raters may weigh the multiple facets differently. Another disadvantage emphasised by both is the little diagnostic information available to students concerning the basis of the score or how to improve their writing. Holistic scoring compares one text to texts from a reference sample (Wolcott & Legg, 1998).

The so-called 'comparative judgement', which can be based on holistic information on the overall quality of the text or on specific traits, compares texts directly to each other in a series or pairwise comparison. The similarly comparative approach with exemplary essays, or so-called 'benchmarks', compares each text to prototypical texts on a continuous rating scale and scores the text accordingly (Bouwer et al., 2024). Bouwer et al. (2024) state that the comparative judgements in writing research are associated with less rater variance than holistic ratings, and with less task-specific rater variance than analytic ratings, but that they require several comparisons to reach a minimum level of reliability.

As already mentioned, 'primary-trait scoring' is somewhere between holistic and analytic scoring. This procedure requires the precision of a scoring guide and should, in contrast to holistic scoring, enable the scoring to be replicated. This method was developed to respond to the limitations of holistic scoring. Thus, the primary-trait method considers particularities of different text genres and evaluates texts based on specific criteria (Wolcott & Legg, 1998). However, the disadvantage of poor information applies, according to Wolcott and Legg (1998), to some extent also to the primary-trait method. Nevertheless, it is also an efficient method and might illus-

trate the essence of writing (Feenstra, 2014). The assessment method using so-called 'rubrics' has certain similarities, aiming for more analytical information to serve to directly link instruction and evaluation within the formative assessment framework. A rubric summarises different criteria in a checklist and classifies what distinguishes different levels of performance for each criterion. The aim is to evaluate the presence of elements and/or the performance levels (Wilson, 2019). Depending on the content and precision of a rubric, this method can be categorised as analytical scoring.

'Analytical scoring' methods evaluate different facets of the writing and assign a separate score to each trait such as structure, voice or development. This procedure requires more time to score compared to a holistic rating and requires the construction of detailed scoring schemes (Feenstra, 2014), but it is suited to providing more extensive information on the writing performance. Theoretically, this procedure should result in better agreement between raters; but this benefit, as well as the benefit of provision of diagnostically useful information, depend on whether the raters apply the distinct analytic scoring rubrics accurately (Lai et al., 2015).

From these presented options it is concluded that many questions must be considered when choosing a task and a scoring method. On the one hand, there is the question of cost and benefits, meaning how much time is invested for the development of the scoring guide and the rating, and how much information is obtained. However, for writing research, the scientific quality criteria of objectivity, reliability and validity are prerequisites (Bortz & Döring, 2006) and should also be striven for in the educational context in order to ensure a fair assessment (Neumann, 2017). On the other hand, in addition to advantages and disadvantages of each procedure, the risk of rating biases, such as sequence effect[12] or halo effect[13], are important in writing assessment (Feenstra, 2014). For the selection of an appropriate assessment procedure or tool for the purpose of research, these issues must be kept in mind. Regarding the assessment of writing processes in German, there is the further difficulty of the lack of appropriate, well-established tools.

Regarding the assessment of higher-order writing processes, such as planning and revision, there are, according to Lenhard and Lenhard (2017), no psychometrically well-founded instruments in German. And there is also a lack of standardised assessing tools for higher-order writing products in this language. However, at this point it is estimated that a holistic rating (referring to a non-standardised reference sample) is independent of the language and could therefore be administered in German. Comparative judgement, comparing to a class or a study sample, would be possible. By contrast, comparing to an exemplary text would only be possible if benchmark texts were available. However, no exemplary essays or benchmark texts

12 The sequence in which texts are scored influences the rating; see Feenstra (2014).

13 Irrelevant aspects of the performance influence the rating, e. g. poor spelling affects the score of text quality; see Feenstra (2014).

for narrative writing in German are known. Thus, such prototypical texts would have to be created first.

An example for a differentiated assessment of content and linguistic aspects in German is the 'Zürcher Textanalyseraster', presented as a basic catalogue that needs to be specified, i. e., a schema of specific features to assess must be developed from this basic catalogue. The development of such a catalogue is time consuming but has then the advantage that it becomes easier and quicker to assess a text, especially for unexperienced raters. The results also allow detailed feedback to the writer, allowing one to illustrate weaknesses and strengths on a group level (Neumann, 2017).

An example of an on-demand-writing task – where students write, during a limited period, a story to a picture prompt – with analytical scoring, are the Subtests 6 and 7 of the Test of Written Language-Fourth Edition (TOWL-4) which is norm-referenced in English (Hammill & Larsen, 2009). A non-standardised translation for this test in German exists (Winkes & Schaller, 2022).

To sum up, not only are the acquisition of writing competence and writing instruction challenging but also the assessment of this manifold competence, with numerous skills and processes involved. Many considerations are necessary to select an adequate assessment method, as well as a task and scoring method. In addition, assessment is task- and genre-specific. It is therefore not surprising that few well-established assessment tools are available, especially in a less widespread language other than English. This can be problematic when considering students who encounter problems in written language acquisition, since the recognition of students struggling in their development is an important part of adapted intervention.

So far, writing competence has been elaborated without regard to particular educational needs. In the following section, the focus will be on students with literacy difficulties.

2.2 Dyslexia

A useful definition of dyslexia in the educational context is as follows: "Literacy difficulties, when they are not caused by the lack of education, are known as dyslexia" (Snowling, 2019, p. 1). Although this definition may seem clear and plausible, a consensus on the definition of dyslexia, also called developmental[14] dyslexia[15], has not yet been reached in the research and education community. In recent decades,

14 The term 'developmental dyslexia' is used to distinguish it from 'acquired dyslexia' following brain injury or illness after the acquisition of literacy skills; see Baddeley et al. (1982). In favour of readability, with a few exceptions only, the term dyslexia is used in this paper instead of developmental dyslexia.

15 *Dyslexia* is an alternative term used to refer to a pattern of learning difficulties characterized by problems with accurate or fluent word recognition, poor decoding, and poor spelling abilities. If dyslexia is used to specify this particular pattern of difficulties, it is important also to specify any additional difficulties that are present, such as difficulties

much has been published about dyslexia. Not only have various research disciplines published texts about dyslexia; the popular media have also made it a topic of discussion. This has brought more attention to the topic but has also led to confused definitions of the condition (Kirby & Snowling, 2022).

Therefore, this section gives an overview of the current state of the research, and of debates on the classification and definition(s) (Section 2.2.1), prevalence and aetiology (Section 2.2.2), and primary and secondary symptoms, as well as comorbidities, of dyslexia (Section 2.2.3). Comorbidities occur frequently, and this work addresses some of these. Section 2.2.4 focuses on evidence-based assessment of dyslexia and interventions, adaptational measures, and possibilities for compensation. Preventive measures will not be discussed here, as this thesis focuses on manifested dyslexia.

2.2.1 Classification and definition(s)

According to the two most recent and most widely used classification systems, the International Classification of Diseases (ICD-11, i. e., the 11th version), from the World Health Organization (WHO), and the American Psychiatric Association's Diagnostic and Statistical Manual (5th version), reading, writing and mathematical disorders are currently classified under an umbrella term. The ICD-11 classifies the disorders as 'developmental learning disorders', while DSM-5 categorises them as 'specific learning disorders'. Both recognise the existence of cases with a more specific form of learning disorder, by providing 'with impairment in' reading, written expression, or mathematics (Di Folco et al., 2022). Concerning reading, an isolated reading disorder can, lately, be classified as in ICD-11, which has been approved by researchers, for example by Schulte-Körne (2021). By contrast, Schulte-Körne (2021) criticises the term 'learning disorders with impairment in written expression' which was adopted into ICD-11 from DSM-5. Schulte-Körne (2021) estimates that this diagnosis, which expands the spelling disorder considerably by including grammar, text production and punctuation, is not sufficiently empirically substantial and harbours diagnostic problems. There is little evidence for the newly added core problems of punctuation, text production, and grammar. Thus, there is a risk of misunderstandings and confusion in clinical practice when diagnosing difficulties as 'learning disorders with impairment in written expression' (Schulte-Körne, 2021). Considering the difficulty of assessment of higher-order writing in general, and especially in German as elaborated in Section 2.1.3.2, the statement by Schulte-Körne (2021) is by no means surprising. How should practitioners diagnose 'learning disorder with impairment in written expression' if there are no appropriate instruments and norms for assessing written expression?

---

with reading comprehension or math reasoning; see American Psychiatric Association (2013).

Although the 11th revision of the WHO follows APA's DSM-5 quite closely, the two classifications crucially differ in terms of certain criteria used in the definition, which can lead to the identification of different individuals and to different prevalence rates depending on the defining criteria (Di Folco et al., 2022). Whether reading and writing disabilities should be defined relative to the level of intellectual functioning is one major difference of the two classification systems, but also in research. Snowling (2019, p. 2) writes, "there is a continuing debate regarding the boundaries of the 'disorder', its precise characterization, and even its name". On the one hand, ICD-11 applies the double discrepancy criterion. This states that the individual's performance in the affected academic skill(s) is markedly below what would be expected for chronological age and general level of intellectual functioning (WHO, 2024). Schulte-Körne (2021) criticises this sharply, because the state of research does not allow us to determine distinct differences between the two groups of students experiencing problems learning to read and write, one with a discrepancy of the IQ and the other with a discrepancy in relation to the age group. On the other hand, DSM-5 (American Psychiatric Association, 2013) renounces an IQ discrepancy and recommends the use of an age discrepancy criterion: "The affected academic skills are substantially and quantifiably below those expected for the individual's chronological age" (2013, p. 67) and have persisted for at least six months, despite the provision of interventions targeting the difficulties. The problem of the dimensional nature of the disorder thus remains, meaning that there is no clear cut of values, but DSM-5 asserts various exclusionary criteria and advocates for the importance of clinical judgement (Snowling et al., 2020). Subtypes of dyslexia have been widely discussed in English. However, this scientific discussion of dyslexia subtypes in German is considered to be too weak and is therefore not discussed further here.

In sum, these classifications do not answer all the questions about the definition of reading and writing disorders, and they are not congruent to the question of the role of IQ for the diagnosis of dyslexia. While Schulte-Körne (2021) claims, in his critique of the ICD-11-revision, that there is a lack of additional information about the significance of severity, comorbidities and, above all, the development of dyslexia, Catts et al. (2024) make recommendations for the diagnosis of dyslexia. In rethinking the definition of dyslexia of the International Dyslexia Association, these latter authors recommend recognising the persistence of word-reading, decoding and spelling difficulties, acknowledging the multifactorial causal basis of dyslexia, clarifying exclusionary factors, and denoting comorbidity with other developmental disorders (Catts et al., 2024). Meanwhile, Snowling et al. (2020), who also consider the difficulty of defining and understanding dyslexia and the advantages and disadvantages of dropping the IQ-discrepancy definition, go a step further and propose four specific points for the practice of labelling a child or an adult as having dyslexia:

1. "The term dyslexia should not be used as a shorthand for 'reading disorder' but should be used to refer to a difficulty with decoding and spelling fluency which is

evident from the early school years and persistent over time" (2020, p. 507). However, according to Snowling et al. (2020), DSM-5 does suggest that achievement, such as literacy skills, more than 1.5 standard deviations below the mean assures the greatest diagnostic certainty.
2. To define the difficulties as dyslexia, they should affect academic functioning, i. e., progress in school as being less in literacy-based areas of the curriculum than the progress of peers in a similar setting.
3. Co-occurring features should be labelled as such, but they should not be considered core to the diagnosis.
4. The diagnosis should be qualified as mild, moderate, or severe. These authors would classify dyslexia as mild when the difficulties can be fully compensated when appropriate arrangements are in place. They hypothesise that students with 'severe' difficulties are often those with a range of comorbid conditions (Snowling et al., 2020).

Both the delimitations and the recommendations by Catts et al. (2024) and Snowling et al. (2020) are, at this point, considered to be largely accepted in the field of speech pathology. Therefore, this framework is used for the further description of dyslexia in this thesis.

2.2.2 Prevalence and aetiology

The prevalence of dyslexia heavily depends on the definition which, as elaborated above, differs in terms of classification systems (Di Folco et al., 2022; Schulte-Körne & Galuschka, 2019). Schulte-Körne and Galuschka (2019) summarise the prevalence of dyslexia, for combined reading and writing disorders, as being between 2 to 6 %, for isolated writing disorders between 2 to 9 %, and for isolated reading disorders between 4 to 7 %.

Boys are more affected by dyslexia than girls. An early Isle of Wight population study of 1968 showed that boys were more affected by 'unexpectedly' poor reading skills, 'unexpectedly' referring to the expectation of the reading skills compared to the IQ (Snowling, 2019). Moll et al. (2014) confirm this detrimental impact on boys in a study about specific learning disorders including also arithmetic deficits. Accordingly, in particular isolated spelling deficits, as well as combined reading and spelling deficits, are more common in boys than in girls.

The aetiology of dyslexia is also, to some extent, uncertain (Schulte-Körne & Galuschka, 2019). Klicpera et al. (2020) write that, in most of the cases, several factors interact in developmental dyslexia. Catts et al. (2024) also acknowledge the multifactorial causal basis of dyslexia, but they emphasise that dyslexia is not primarily caused by environmental factors (e. g., poor instruction) or by lack of motivation. While the research situation thus does not yet allow a clear understanding of the cause, Schulte-Körne and Galuschka (2019) describe the research efforts, in this re-

gard, which provide some models for a pathomechanism based on molecular genetic findings. These authors divide the assumed causes into the following influencing factors: 1) genetic; 2) neurobiological and neuropsychological correlates; and 3) psychosocial. The following paragraphs give an overview of these areas.

Several authors, e. g. Schulte-Körne and Galuschka (2019), assume that 'genetic factors' show a causal link between gene-function and brain-function, although the state of research does not yet allow a causal link between specific candidate genes or candidate gene regions. This genetic factor explains, to some extent, the hereditary component that has been demonstrated and confirmed by various studies (Schulte-Körne & Galuschka, 2019). According to Klicpera et al. (2020), the significance of genetic factors is greater for spelling, compared to reading, and tends to remain high or even increase in older children and adolescents, while the significance of genetic factors decreases for reading and with age.

Regarding 'neurobiological and neuropsychological correlates', Schulte-Körne and Galuschka (2019) write that literacy skills require a functional neural system that integrates the processing of orthographic, phonological and semantic-lexical features of words. These authors summarise deficits in the neural system as follows. On a neurobiological level, there are impaired functions in regions of the occipito-temporal cortex, the dorsal temporo-parietal cortex, and the inferior frontal gyrus. The impaired functions are mainly situated in the left hemisphere of the brain. Allocating this effect to the behavioural level, the authors write that reduced visual and auditory stimulus processing is discussed as a correlate of the reading and/or spelling disorder.

Although the evidence is quite strong that phonological processing difficulties often underlie dyslexia – these difficulties are generally believed to be constitutional in origin (Castles & Coltheart, 2004) – there is notable inconsistency in this association. Not all individuals with dyslexia have difficulties with phonological processing, and not all of those who demonstrate such difficulties develop significant word-reading problems (Castles & Coltheart, 2004; Catts et al., 2024). Schulte-Körne and Galuschka (2019) also name deficits in phonological awareness, automated memory recall, expressive and receptive language skills, and deficits in working memory and attention. The authors summarise these as deficits at a cognitive level. Further deficits of this type are, according to Catts et al. (2024), for example limitations in oral language, processing speed, visual processing or procedural learning.

'Psychosocial factors', e. g. the child's family, school environment and socio-economic environment, have not been investigated sufficiently to estimate their role in the development of dyslexia. At present, researchers assume that the family environment has a significant influence on literacy development, and that literacy can be stimulated in the family environment, but only when reading and spelling skills are explicitly taught. However, as mentioned above, the lack of a rich family reading environment is not a causative factor for dyslexia. The same applies to school instruction and methods which can be important influencing factors. These are most

probably significant for the development of reading and/or spelling disorders, but in the opposite direction, it cannot be assumed that optimal instruction can prevent dyslexia based on neurobiological and genetic differences (Catts et al., 2024; Schulte-Körne & Galuschka, 2019). Emotional and motivational factors in the child, such as the disposition to learn and to work hard, or reduced self-concept and self-efficacy, can also have an impact on achievement and therefore on the manifestation of dyslexia. However, these are more likely to be internalised disorders or secondary disorders (see next section) and are not viewed as primary causal factors of dyslexia. Other individual and environmental factors also play a role in the multifactorial causal model of dyslexia, e. g. comorbidities such as AD(H)D, or mental disorders. These factors do not explain dyslexia but can aggravate it (Catts et al., 2024).

2.2.3 Symptomatology, secondary symptoms, and comorbidities

2.2.3.1 Symptoms of dyslexia

The symptoms of dyslexia have already been partially outlined by the recommendations by Catts et al. (2024) and Snowling et al. (2020). Thereafter, dyslexia can be defined as a problem with learning that primarily affects the development of reading accuracy and fluency and/or spelling skills. Schulte-Körne and Galuschka (2019) specify that these persistent and significant difficulties labelled as dyslexia are not attributed to a far below average intelligence or developmental age, insufficient ability in the instructional language, difficult psychological circumstances, uncorrected visual and hearing impairments, or neurological or mental illness. Dyslexia is a lifelong condition with an early onset (Snowling et al., 2020). The difficulties are also manifest in other school subjects when literacy skills are involved, as well as when learning foreign languages (Schulte-Körne & Galuschka, 2019). The above-presented prevalence rates are classified as isolated and combined disorders. Moll and Landerl (2009) state that reading difficulties are often, but by no means always, associated with poor spelling (Moll & Landerl, 2009)[16].

Schulte-Körne and Galuschka (2019) and A. Mayer and Lindberg (2016) describe the symptoms of reading disorders as characterised by pronounced difficulties in reading accuracy, e. g. characterised by a large number of errors when reading aloud, as well as significantly reduced reading speed, due to a lack of automated word recognition. Both difficulties have a negative impact on reading comprehension and other everyday tasks for which reading skills are required (Schulte-Körne & Galuschka, 2019). However, A. Mayer and Lindberg (2016) specify that these difficulties in reading comprehension are not original symptoms of dyslexia. Therefore, the authors'

16 In favour of the readability, this thesis does not always mention isolated and/or combined disorders, although their existence is not questioned. The term dyslexia is used as a common term for both isolated and combined disorders.

view is not in line with the ICD-11 classification that includes reading comprehension difficulties as symptoms of dyslexia.

Spelling disorders are characterised by high frequency of spelling errors that persist despite intensive learning efforts. Children with a spelling disorder also find it difficult to deal with well-practised words and tend to spell the same words in differently incorrect ways. There is no characteristic error type for spelling disorders. The diagnosis can therefore not be derived from the type of error (Schulte-Körne & Galuschka, 2019). In addition to poor spelling, Sumner et al. (2014) observe that students with dyslexia produce fewer words than their peers without dyslexia during the same period of time. Afonso et al. (2019) and Sumner et al. (2014) explain these differences mainly by the higher number of pauses produced within misspelled words and by longer pauses. The differences are therefore rather situated in the spelling processes than in differences in handwriting speed.

Further basic functions and skills may impact spelling skills of students with dyslexia in a negative way (Hebert et al., 2018). According to Hebert et al. (2018) difficulties experienced by students with dyslexia may be related to poor reading skills, and problems in working memory and/or executive functions such as goal-directed behaviour, planning, and organisation. These difficulties therefore potentially affect overall writing skills. From this perspective, difficulties in the overall writing skills would not be categorised as original symptoms but as secondary symptoms. Again, this is not in line with the above presented core symptoms in the ICD-11-classification (WHO, 2024) which was criticised by Schulte-Körne (2021) regarding the existing lack of evidence and potential difficulty of diagnosis.

The symptoms of dyslexia are present from the start in school and stay relatively stable throughout the school career (Schulte-Körne & Galuschka, 2019). 70 % of the dysfluent readers in Grade 1 were, in a follow-up study in Grade 8, still poor readers. The assessments showed that reading speed, rather than accuracy, impacted reading fluency (Landerl & Wimmer, 2008). Regarding difficulties in spelling, these stay present throughout the school years and into adulthood. People with dyslexia produce a higher number of mistakes at word and sentence levels (Schulte-Körne & Galuschka, 2019). Such persistent spelling difficulties represent the major reason for complaint in adults with developmental dyslexia (Afonso et al., 2019; Di Betta & Romani, 2006). However, the spelling errors change over time. Students in Grade 3 committed various kinds of errors; while students in Grade 5 produced mainly errors that were phonologically plausible. However, the errors of adults cannot be explained by the phonological deficits usually reported in children (Afonso et al., 2019). The assumption that the difficulties in adulthood are only due to reduced exposure has not been confirmed and can therefore explain the difficulties in reading and writing in adulthood only to a limited extent (Di Betta & Romani, 2006).

Early, high-quality and intensive support and teaching can have a positive influence on the progression of reading and/or spelling disorders (Schulte-Körne &

Galuschka, 2019); but the risk of secondary symptoms and heavy consequences for school and professional career is high, as the following sections illustrate.

2.2.3.2 Secondary symptoms, consequences on school and professional career

While ICD-11 classifies reading comprehension and text production as core symptoms of dyslexia, many researchers categorise these as consequences of primary disorders (e. g. Hebert et al., 2018; A. Mayer & Lindberg, 2016; Schulte-Körne, 2021). While some students manage to hide difficulties to some extent, the reading and/or spelling disorder becomes more obvious when the demands increase and students' school performance declines rapidly. Those affected experience repeated academic failures as well as difficulties in extracurricular learning. This can lead to serious fears of failure and further, indirectly related challenges such as negative self-concept, low self-esteem, anxiety, and depression (Catts et al., 2024; McArthur et al., 2000; Scheerer-Neumann, 2018; Schulte-Körne & Galuschka, 2019). However, the secondary symptoms of dyslexia do not stop at the higher vulnerability to experiencing poorer school achievement and overall school success: affected people are also more vulnerable to experiencing consequences such as behavioural problems, psycho-social problems, truancy, or dropout from high school (Catts et al., 2024). DSM-5 by the American Psychiatric Association (2013, p. 70) classifies dyslexia among specific learning disorders, and states: the "Specific learning disorder is associated with increased risk of suicidal ideation and suicide attempts in children and adolescents, and adults". These negative consequences are just as concerning as the primary symptoms, because they can seriously impact the overall quality of life of people with dyslexia. Therefore, these consequences need explicit recognition and attention to potentially prevent or reduce their occurrence (Catts et al., 2024).

2.2.3.3 Comorbidities

The co-occurrence between two (or more) disorders in the same individual is called a comorbidity (Angold et al., 1999). Regarding dyslexia, comorbidity can occur between disorders within the same diagnostic category such as a mathematical disorder which is, in DSM-5, equally grouped within the subcategory of specific learning disorders among the neurodevelopmental disorders, or between disorders from different diagnostic categories (American Psychiatric Association, 2013; Angold et al., 1999). The prevalence rates of comorbidities vary widely, but Moll et al. (2019) estimate that about 40 % of the children with a reading disorder or dyslexia have another disorder as well. Dyslexia commonly but not exclusively co-occurs with a developmental language disorder (DLD), speech sound disorder (SSD), mathematical disorder (MD), or attention-deficit/hyperactivity disorder (AD(H)D) (Catts et al., 2024). Since these comorbidities show an increased prevalence in people with dys-

lexia, they are each discussed in the following paragraphs. The already mentioned secondary symptoms – anxiety disorders, depressive symptoms, and absenteeism from school – are also named among the comorbidities by Galuschka and Schulte-Körne (2016).

Developmental language disorder (DLD): According to Catts et al. (2024, p. 12), "numerous studies have demonstrated the increased prevalence of DLD in children with dyslexia". These language problems, which extend beyond phonology, are frequent in individuals with dyslexia, some of them as "late talkers", but the problems vary in intensity from clinically mild to significant (Price et al., 2022). The comorbidity rates for children with dyslexia and DLD observed by McArthur et al. (2000) in their sample were even higher than the rates given by Moll et al. (2019). McArthur et al. (2000) found that over 50 % of children can be equally classified as having a specific reading disability (SRD) or DLD[17]. Adlof et al. (2021) and Duff et al. (2023) state that children with dyslexia and comorbid DLD have more severe difficulties in word reading and general reading than those with dyslexia alone. Early spoken-language difficulties should therefore be considered as a risk factor for dyslexia (Catts et al., 2024).

Speech sound disorder (SDD)[18] involves the persistent and systematic deletion, substitution and distortion of speech sounds (Farquharson et al., 2021). Cabbage et al. (2018) recognise that the prevalence of SDD in dyslexia are inconsistent and that SDD is often mediated by oral language skills. Therefore, the authors do not estimate SDD as a major risk factor of dyslexia, but they argue that an isolated speech sound disorder also poses an increased risk for dyslexia that should be considered.

Mathematical disorder (MD) is also often comorbid with reading disorder[19] (RD), with prevalence rates varying greatly depending on the study (Moll et al., 2019). Joyner and Wagner (2020) found, via a meta-analysis, that children with MD are twice as likely to have dyslexia as those without MD. However, Moll et al. (2019) claim that the large overlap between MD and RD does not concern all mathematical skills but is restricted to those mathematical subskills that build on language skills, i. e., arithmetic fluency. However, the deficits in basic number processing and in literacy are distinct. Thus, "Comorbidity between RD and arithmetic deficits can be explained by domain-general risk factors (i. e., language skills) and by symptom overlap (Moll et al., 2019, p. 292)".

Regarding *Attention-deficit/hyperactivity disorder (AD(H)D)*, Catts et al. (2024) describe a bidirectional comorbidity between AD(H)D and dyslexia and name a 25–40 % prevalence rate. Schuchardt et al. (2015) investigated the comorbidity rates of AD(H)D symptoms and of specific learning disorders, including dyslexia and

17 Former specific language impairment (SLI).

18 Children with persistent speech sound disorders-(SSD) are those who continue to experience speech sound errors past a typical age of acquisition; see Farquharson et al. (2021).

19 Aspects of writing are not mentioned explicitly in the cited studies.

MD. While they assume a prevalence rate of AD(H)D of 5 % in children without a specific learning disorder, they find considerably higher prevalence rates between AD(H)D and reading and/or writing disorder (around 20 %), but not with MD. In addition, these authors describe a much higher risk of comorbidity for boys than for girls.

Several authors recommend to carefully identify these comorbid conditions in diagnostic evaluations, although these disorders are not a defining characteristic of dyslexia. Firstly, the comorbidities should be noted as such to improve the diagnosis and adjust the subsequent treatment. Secondly, the precise picture of the conditions might serve as additional criteria for identifying dyslexia in diagnostic evaluations, meaning that MD might lead to the detection of dyslexia (Catts et al., 2024; Joyner & Wagner, 2020; Moll et al., 2019; Price et al., 2022; Snowling et al., 2020).

2.2.4 Assessment, intervention, adaptation, and compensation

2.2.4.1 Assessment of dyslexia

Already in the 1960s, it was accepted that dyslexia required an educational intervention. However, before teaching could begin, a 'diagnosis' was sought that was based on a medical model of referral and treatment (Snowling, 2019). Therefore, this section, firstly, provides an overview of the diagnosis of dyslexia; and secondly, presents the options for intervention and/or adaptations and compensation. In distinction to Section 2.1.3 where the educational assessment of writing skills was elaborated, this section focusses on special educational diagnostics, aiming to identify and classify disabilities and impairments and to support and promote people with disabilities (Gebhardt, 2023). As this thesis concentrates on already manifest dyslexia, the topic of early detection and prevention is not in the scope of this thesis.

According to Scheerer-Neumann (2018), diagnosing reading and/or spelling disorders should follow a hierarchical approach that starts with the assessment of whether a disorder is present. Subsequently the diagnosis is refined, and further features are assessed. Schulte-Körne and Galuschka (2015) wrote, in the now outdated S3-Guideline, that diagnosis should be based on the results of psychometric performance tests that measure reading speed, reading errors, reading comprehension, and spelling errors in word and/or text writing. The problem is that dyslexia is not a specific medical syndrome with a clear cut-off value (Snowling et al., 2020), and as discussed in Section 2.2.1, there are no (inter-)nationally recognised standards (Schröder et al., 2022). Therefore, Schulte-Körne and Galuschka (2019) recommend using any of the three discrepancy definitions: the age norm, the class norm, or the IQ norm. A difference of at least one standard deviation (SD) below the age or class norm equals a below-average performance in reading and/or spelling.

Once the deficits have been identified, the diagnostician determines the current state of reading and/or writing development of the child or adolescent (Scheer-

er-Neumann, 2018). The diagnostic objective, the age of the person, the grade level and the type of school, as well as the methodological quality of the psychometric tests, determine the selection of the appropriate test procedure (Schulte-Körne & Galuschka, 2019). Regarding reading skills, a distinction is required between basic reading, such as word recognition and decoding at word level, reading fluency at sentence or text level, and reading comprehension[20]. The same distinctive procedure is recommended for writing skills in order to assign the developmental stage and a detailed error analysis. This precise assessment of phonetic, orthographic and morphemic writing skills not only enables a statement about specific deficits but can provide indications for specific intervention planning and also allows one to derive the precise support measures (Scheerer-Neumann, 2018; Schulte-Körne & Galuschka, 2019). For the diagnosis of these lower-order writing skills in German, several standardised tests are available for all grade levels and from the end of primary school onwards, and also for different types of school. Nevertheless, Galuschka and Schulte-Körne (2016) indicate a lack of standardised spelling tests in German that can be used for the entire school year because of their standardisation with time intervals or at specific periods. Regarding the diagnosis of the core symptom – according to the classification systems, 'impairment in written expression', here used synonymously with 'higher-order writing skills' – diagnosis in German is fraught with problems because no up-to-date standardised test is available (Lenhard & Lenhard, 2017; Schulte-Körne, 2021), as already highlighted in Sections 2.1.3 and 2.2.1.

As presented in Section 2.2.1 Snowling et al. (2020) recommend qualifying dyslexia as mild, moderate or severe. In addition to psychometric performance tests, the assessment of developmental and school history and, as stated above, of co-occurring features and comorbidities are important parts of the diagnostical process and must be considered in the treatment plan (e. g. Catts et al., 2024; Galuschka et al., 2014; Galuschka & Schulte-Körne, 2016; Schulte-Körne & Galuschka, 2019; Snowling et al., 2020). Clinical examinations, such as hearing or visual tests, and the observation of learning strategies, are also recommended (Scheerer-Neumann, 2018). Schulte-Körne and Galuschka (2019) call this a 'multiaxial approach' which includes the assessment of the core difficulty, differential diagnostics, and the clarification of specific comorbidities. They recommend an interdisciplinary diagnostic for the diagnosis of dyslexia.

This section has so far presented the diagnostic procedure and its elements to determine, firstly, the presence or not of dyslexia and, secondly, the aims to evaluate the development at a specific point. However, assessment does not stop there but also takes place in the process of an intervention in order to evaluate the response to an intervention (Gebhardt, 2023), as discussed under the following subheadings.

20 Tests meeting the standards for reading assessment are presented in detail by: Schulte-Körne and Galuschka (2019); Scheerer-Neumann (2018); and A. Mayer and Lindberg (2016).

2.2.4.2 Interventions

A vast arsenal of different treatment methods for dyslexia is available, but many methods have not been evaluated or only to an unsatisfactory degree. A meta-analysis including treatment methods, up to April 2015 (Galuschka & Schulte-Körne, 2016), shows an improvement in reading and spelling performance but only for symptom-specific approaches. Galuschka and Schulte-Körne (2016) recommend these evaluated methods for treating reading and spelling disorders, requesting, at the same time, more solid research for all interventional approaches and methods.

Systematic instructions on letter-sound correspondences and letter-syllable-synthesis, as well as on morpheme synthesis, are the best for developing reading performance. A more recent meta-analysis on phonics training for English-speaking poor readers, by McArthur et al. (2018), confirms the former findings. Phonics training teaches the identification of each letter of a letter-cluster in a new word, transposes each letter or letter-cluster into its corresponding speech sound, and blends those speech sounds into a word. On the writing side, Galuschka and Schulte-Körne (2016) recommend, as the most effective methods to improve spelling performance, systematic instructions on sound-letter correspondences and exercises analysing sounds, syllables, and morphemes. In addition, applying a training procedure to enable the acquisition and generalisation of orthographic regularities has proven to be effective (Berninger et al., 2008; Galuschka & Schulte-Körne, 2016; Ise & Schulte-Körne, 2010; Kirk & Gillon, 2009). However, most of the studies included in the meta-analysis stem from English-speaking countries. Galuschka et al. (2014) state that it is not yet clear whether these interventions are equally effective across languages, among others in German. Important to add is that there is a dearth of evidence on interventions for spelling and writing fluency, as pointed out by Snowling et al. (2020). This meta-analysis also did not confirm the effectiveness of programs for auditory or visual perception and processing, training, medication, or Irlen lenses. The same applies to alternative approaches (Galuschka & Schulte-Körne, 2016).

Interventions to support children with dyslexia should start as early as possible and continue until the child or young adult has reached a reading and spelling ability that enables their participation in social and public life in an age-appropriate way (Galuschka & Schulte-Körne, 2016). Although the state of research on this aspect is not currently sufficient, some evidence indicates that interventions with higher amounts of treatment or longer durations of treatment seem to be more effective in improving literacy skills than therapies with small amounts of treatment or short-time interventions (Galuschka et al., 2014). The same applies for the success of interventions regarding subgroups of severity. Thereafter, children and adolescents with mild reading disabilities show more improvement in literacy skills than more severely impaired children (Galuschka et al., 2014).

In addition, Galuschka and Schulte-Körne (2016) investigated, by the means of a meta-analysis, the setting, i. e., individual or small groups, and the person that

provided the therapy or intervention. The authors do not identify a difference in effectiveness between individual sessions or group sessions, but the profession of the therapist did affect the effectiveness of the intervention. Measures administered by teachers and study authors were more effective than those by pupils, parents and university students.

2.2.4.3 Adaptational measures

As elaborated above, early intervention is strongly recommended for students with dyslexia. In addition to remedial intervention, instruction in school and potential adaptations of instruction to students who are struggling with reading and writing play an important role (S. Graham et al., 2022). S. Graham et al. (2022) describe instructional adaptations for the weakest writers as modification with the underlying idea that the success of weaker writers is aided by making instruction responsive to students' needs.

An information sheet (in German, 'Infoblatt H, Lese- und/oder Rechtschreibstörung (Dyslexie)'), of the Swiss Foundation Centre for special needs education, presents such measures for students with dyslexia for the attention of the teachers. On the one hand, this document presents instructional recommendations for all students that are at the same time helpful for students with dyslexia, e. g. the choice of simple fonts without serifs for better decoding. On the other hand, it presents examples of adaptations that can be applied in the classroom for students with dyslexia, such as digital teaching materials with an additional accessible format. The authors recommend planning these adaptations as a team of teachers and specialists, and discussing them with parents (Ayer & Joss Almassri, 2023). However, such recommendations for teachers give no information about what actually is applied in the field; but a few studies about instructional adaptations for struggling writers have been conducted in North and South America (S. Graham et al., 2022; S. Graham et al., 2016). These have investigated the type[21] and the frequency of adaptations applied by primary and middle school teachers. Accordingly, extra encouragement and extra time to complete writing assignments are the most common adaptations applied by teachers of Grades 1 to 3 in the USA. Adaptations involving the use of technology or dictations are the least commonly applied. S. Graham et al. (2016) report that teachers knew about the adaptations and how to implement them but were less positive about the time needed to apply them. In addition, "teachers further evidenced some concern about dictation having negative effects on their struggling

21 Adaptations named by S. Graham et al. (2022): extra handwriting, grammar/spelling, planning/revising, sentence construction, and text structure instruction, as well individual mentoring, extra writing instruction using computer technology, adaptations for facilitating writing such as weaker writers working with a peer, and providing choice or alternative writing assignments.

writers" (S. Graham et al., 2016, pp. 12–13). Almost the same picture emerged when investigating the same topic for middle school teachers in the USA. Teachers mentioned the lack of hardware and software in school as reasons for the limited use of technology (S. Graham et al., 2014). The study by S. Graham et al. (2022) about adaptations of writing instruction to the weakest writers also shows more frequently applied adaptations in Chile than in the USA. Regarding differences of adaptations between grade levels, the authors found no significant differences in the type or the frequency of adaptations.

Locke et al. (2015), who investigated effects and adaptive strategies of dyslexia in and after training of doctors and nurses, conclude that there is insufficient research currently available about adaptations in dyslexia and that the existing studies are not sufficiently recent to map the possibilities of new technologies. Beer et al. (2014) draw a more positive picture for students with dyslexia, by assuming that the educational possibilities for children and adolescents with dyslexia have been improved during the last decades through assistive technology, by providing extra time, and other adaptations such as recorded textbooks or bigger font. However, these authors see still major challenges for people with dyslexia in the transition from education to professional life.

Svensson and Lindeblad (2019) state, with regard to students' academic and professional future, that at a certain point in the learning process, the question arises as to whether therapy and differentiation are sufficient or whether certain processes should be compensated for. The adaptations and differentiations discussed in this section support learning, but adaptations as mentioned by Beer et al. (2014) also touch on measures that could be classified as compensational measures which relieve the learner with dyslexia of certain processes. These measures usually do not change the objectives of learning, because the compensation allows the learner to direct the full attention to the essential learning content and therefore perform as expected without a disadvantage such as dyslexia. In Switzerland, under certain circumstances, these compensational measures are classified as 'compensation of disadvantages' (Ayer & Joss Almassri, 2023).

2.2.4.4 Compensational measures and compensation of disadvantages 'Nachteilsausgleich'

The descriptions above in Section 2.2.3 of the secondary symptoms of dyslexia illustrate the high risks that come with such difficulties. As also discussed, the symptoms of dyslexia remain often present throughout school and in adults and cause many risks and difficulties at work and in everyday life (Schulte-Körne & Galuschka, 2019). Section 2.2.1 about the classification of dyslexia outlined the disorder, from a medical perspective, as a developmental disorder (WHO, 2024), or as specific learning disorder (American Psychiatric Association, 2013). From a legal perspective, dyslexia is classified as a handicap (Ayer & Joss Almassri, 2023). Therefore, internationally,

the 'Convention on the Rights of Persons with Disabilities' (United Nations, 2024), and on the Swiss National level, the 'Behindertengleichstellungsgesetz' (BehiG SR 151.3, 2002/as at 1. Juli 2020), apply. Accordingly, persons with a disability are legally entitled to measures to compensate for their disadvantage. The resources invested in eliminating the disadvantage, and the effect achieved, must remain in a reasonable relationship and comprise individual measures to minimise the disadvantage caused by a disability. This does not refer to an adaptation of the content or educational objectives; instead, it refers to the adaptation of the conditions under which teaching and examinations take place. Technological aids, personal assistance, and adaptations of teaching and examination materials, but also adjustments in terms of time and space, can be approved (Ayer & Joss Almassri, 2023; United Nations, 2024).

While several cantons[22] provide policies on compensation for disadvantage (e. g. Staat Freiburg, 2023), there is no substantial research found on German-speaking countries about what compensational measures teachers actually apply in school and about their effectiveness. However, according to Sahli Lozano et al. (2016) and Schellenberg et al. (2017): firstly, dyslexia is the most common reason for applications for compensation for disadvantage; and secondly, additional time is the most frequently accorded compensation.

In sum, the literature on the adaptational and compensational measures presented here does not allow overall conclusions about the effectiveness, type and frequency of applied measures in educational and professional settings. Locke et al. (2015) state that up-to-date and substantial research is needed for the professional development of affected people. Regarding the educational setting of interest for this thesis research, Peterson-Karlan (2011) and S. Graham et al. (2016) make the following two points: on the one hand, critical examination of the tools and technologies that may provide compensatory benefit are necessary; on the other hand, it is also important to make sure that teacher education programs and professional development efforts by schools help pre- and in-service teachers to receive the necessary knowledge. The teachers must become aware of possible adaptations and know their strength and limitations as well as how to apply them in a busy and over-crowded curriculum (S. Graham et al., 2016).

22 The cantons of Switzerland are the member states and have their own constitutions, parliaments, governments and courts, EDA (2024).

3. Speech-to-text for Students with Dyslexia – Theoretical Approach

Writing is a complex process that involves several challenges for all students, but especially for students with dyslexia (e. g. Afonso et al., 2019; Mossige, Arendal, et al., 2023). Therefore, early intervention, and/or adaptational and compensational measures for reading and writing, are a necessity for learning and coping with the challenges of dyslexia (Ayer & Joss Almassri, 2023).

Difficulties in accurate and fluent spelling are the principal primary symptoms of dyslexia regarding writing (Rose, 2009), as elaborated in Chapter 2, Section 2.2.3. Because spelling is not automised, this process demands many of the limited resources in the working memory and thus leaves few resources for text generation and executive functions (Berninger & Amtmann, 2003; Mossige, Arendal, et al., 2023). Therefore, Mossige, Arendal, et al. (2023) assume that writers with dyslexia depend on technological tools which can help ensure accurate spelling. Section 3.1 presents an overview of technological tools that support transcription and explains why the focus in this thesis research is directed towards STT. Section 3.2 provides an insight into the technology of STT. Section 3.3 raises the question of whether, on a theoretical basis, STT is an appropriate technology for writing for students with dyslexia.

3.1 Overview of technology-supported writing tools for students with dyslexia

Various authors consider that technological tools for writing could be applied in a beneficial way to support students and adults with dyslexia (Locke et al., 2015; Mossige, Arendal, et al., 2023; Svensson et al., 2021). Peterson-Karlan (2011) assumes that, for example, word processors, due to their support, can advance planning, transcribing, and editing and revising. Such tools can be applied as adaptational measures in the differentiation of the educational setting; or, on the legal basis of the Convention on the Rights of Persons with Disabilities (United Nations, 2024), they can be approved as compensation for disadvantage (Ayer & Joss Almassri, 2023). Schröder et al. (2022) identify, as digital interventions and media to support the narrative writing skills of students with diagnosed special educational needs, the technologies of word processing, word prediction, and text-to-speech programs. Text-to-speech (TTS) compensates for difficulties in decoding, which can be due to visual impairment or literacy difficulties. Meanwhile, Svensson et al. (2021) refer also, for students with dyslexia, speech recognition or speech-to-text technology (STT) to compensate for impaired writing abilities. On a theoretical basis, the following discussion pursues

the question of the most adequate technology for learners with dyslexia. To this end, it firstly outlines the most prominent features of the above-mentioned technologies.

Chapter 2, Section 2.1.1 illustrated the spelling process as one of the lower-order processes. Lower-order processes receive assistance by spellcheckers, commonly integrated in 'word processing'. The term, word processing, has changed over time, and this technology is increasingly packaged with other technologies that support writing. A text editor, usually integrated in word processing, simplifies editing and revising text (Morphy & Graham, 2012). Perelmutter et al. (2017) assume, on the basis of a limited number of studies with adolescents and young adults, that a text editor can assist in writing. However, the person who produces text still accomplishes both the lower-order processes, by typing letters and words, and the higher-order processes such as planning and structuring text. Regarding assistance in relation to the challenges for a person with dyslexia, the assistance is limited and might direct the attention to spelling in a hampering way for the higher-order processes. In addition, Rønneberg et al. (2022) point out the problem of spellcheckers pinpointing misspellings with red underlining, which provides negative feedback for all writers but especially for those with dyslexia with poor spelling.

The technology of 'word prediction' assists the writing process by generating a list of the words most likely to follow in a sentence. The software then displays these in a word list, and the user selects the desired word. This process decreases the number of keystrokes required to type a word. It is therefore supposed to improve the rate and accuracy of text entry, reduce fatigue, and thereby enhance the productivity of writing (Tam et al., 2002). Using this technology, the machine thus takes over parts of the lower-order processes. Nevertheless, Mossige, Arendal, et al. (2023) criticise this procedure as time-consuming.

A technology that does mostly take over the lower-order writing processes is 'speech recognition technology', here used synonymously with STT, which translates oral language into written language. In other words, a person dictates oral text to a machine. The person thus neither types nor focuses on the spelling of words. However, the interpunctuation must still be accomplished by the person.

At the time of study planning in 2021, STT represented the technology that most consequently assisted in lower-order writing processes. With the release of ChatGPT-4 in November 2022 and subsequent advances in natural language generation, writers have gained access to another tool that supports the entire writing process – from idea generation to final proofreading and editing (Imran & Almusharraf, 2023). One could therefore argue that such a comprehensive tool is better suited to assist students with dyslexia across all stages of writing. However, as outlined above, the aim of this study is to understand the effect of reducing the burden of lower-order processes, by allowing the writer to handle the higher-order processes such as idea generation and text structuring. Consequently, the use of STT only remains appropriate for the purpose of the present thesis research despite recent technological developments.

This chapter explores the question of whether STT can be beneficial for students with dyslexia from a theoretical perspective. To this end, it develops the theoretical cognitive and motivational rationales that can argue in favour of or against the use of this technology for students with dyslexia. It then addresses the use of STT from an educational perspective, in particular media theory and new literacy theory. However, firstly, this chapter provides an insight into the most important steps in the development of the technology of STT.

3.2 Speech recognition technology

One of the primary means of human communication is speech. Two primary technologies concerning speech have developed: text-to-speech (TTS); and speech-to-text (STT) or automatic speech recognition (ASR). As its name suggests, TTS refers to the process of the conversion of text into speech (Filippidou & Moussiades, 2020). The software transforms written text to oral language, i. e., a speech synthesizer generates a synthetic spoken form without the necessity of pre-recording (Giannouli & Banou, 2020). However, since TTS assists, first and foremost, reading, it is only briefly mentioned in this thesis research. The counterpart to TTS is STT which "generates digital text from spoken language" (Matre & Cameron, 2024, p. 2). This technology converts sound waves into electric signals, using sound cards that digitalise the signals and by a speech engine software that transforms data into text. The technology user speaks into a microphone connected to a device where the translated words appear on the screen within seconds (Aldosari et al., 2023).

3.2.1 Development of the technology

Researchers at the Bell Telephone Laboratories (USA) built the first speech recognition system in 1952 (Davis et al., 1952; Filippidou & Moussiades, 2020). This early recognizer, called circuity, automatically recognised spoken digits, i. e., numbers

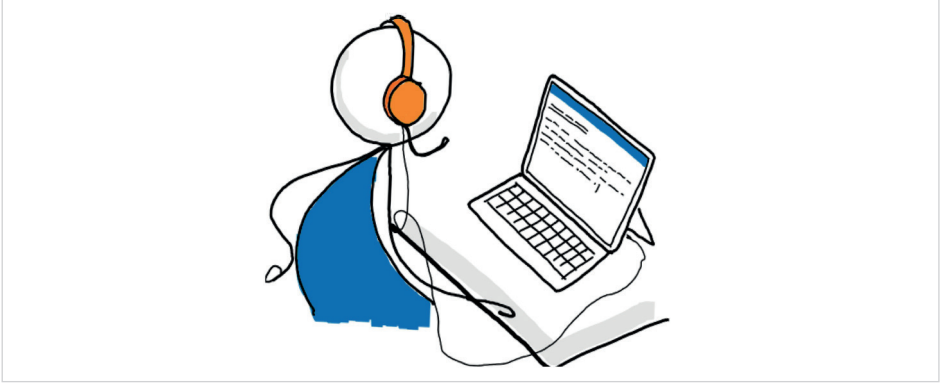


Figure 3.1: Generation of (digital) text from spoken language via microphone and speech recognition technology (own representation)

from 0 to 9, at normal speech rate by a person for whom the machine is adjusted. The circuity contained three separate sequential functions. Firstly, it compared unknown sounds to Format 1 plots versus Format 2 plots; and secondly, it matched these against similar plots of reference sounds. The technique for identification of the digits is part of the process called pattern matching:

> Pattern matching involves comparing some aspect or aspects of data derived from an unknown signal with the corresponding aspect or aspects of a number of known signals considered as standards to determine which of the standards was intended by the talker when speaking the unknown digit (Davis et al., 1952, p. 637).

The best matching techniques were employed in a way that normalised various lengths of digits and speeds of talking. Since human speech contains variability, the goal is not an exact match of an utterance to any standard sound but rather to decide whether a sound resembles one standard more closely than any other (Davis et al., 1952). Therefore, the speech recognition is in this regard the opposite of voice recognition with its unique identification of a speaker from his voice pattern, as applied in security devices (Hollier, 1983). Finally, the speech recognition machine presented the information visually. This matching technique from 1952 allowed the machine to deliver a determination in spite of a poor match. While short unintentional sounds and noises were eliminated, this circuity did not allow the machine to reject a word or sound but provided the output within the accepted vocabulary (Davis et al., 1952). Initially, this early system required speakers to pause after each digit (Filippidou & Moussiades, 2020).

The first recognition system of continuous speech, a major achievement, was constructed in the late 1960s in the USA; followed by the development of an algorithm in the UK that could recognise speech, creating a sequence of words contained in connected speech (Filippidou & Moussiades, 2020). This latter technology was intended for activities where free hands were of advantage, such as giving instructions or commanding controls in industry, and aimed to accelerate speed by oral entry as a faster alternative to keyboard entry (Hollier, 1983; Sensor Review, 1981). The technology also used pattern recognition approaches, including, among other techniques, the use of linear predictive coding, as the basic methods of spectral analysis (Karpagavalli et al., 2011). The pattern recognition approach or pattern-matching approach involves two main steps, namely pattern training and pattern comparison. The system described in the Sensor Review (1981) contained an initial vocabulary of 100 words but was extended up to 2000 words.

Karpagavalli et al. (2011) outline the advancement that occurred in the early 1980s by the introduction of the Hidden Markov Models[23] to the Automatic Speech

23 Hidden Markov Models (HMM) are powerful tools for probabilistic sequence alignment. They contain a doubly stochastic process with an underlying stochastic process that is not observable (hence the term, hidden), and a second stochastic process that produces a sequence of observations; see Furui (2010).

Recognition (ASR). These models provide a much more powerful tool for speech recognition and allow the use of larger vocabulary. "Recognizers with continuous speech capabilities are some of the most difficult to create because they utilize special methods to determine utterance boundaries" (Karpagavalli et al., 2011, p. 327). In contrast to the initial technology that had to be adjusted to a speaker, this new technology allowed speech recognition in a speaker-independent manner.

Nevertheless, the performance of speech recognition applications still had the challenge of reverberation and to overcome even low levels of ambient noise that deteriorated the performance of ASR. The robustness of recognition of speech in noise increased in the next generation of ASR. From 2000 onwards, the technology has used parallel processing methods that increase recognition decision reliability. The combination of Hidden Markov Models and Acoustic Phonetic Approaches[24] targeted to detect and correct linguistic irregularities. In addition, the techniques of machine learning advanced the technology by finding the optimal combination of models (Karpagavalli et al., 2011).

In their paper of 2011, Karpagavalli et al. estimate that the most successful solution is going to be the use of larger databases and trying to embed into the training set all the variability of speech and speakers. These authors assume that approaches mostly based on Artificial Neural Network[25] would provide the computational power required for continuous speech recognition (Filippidou & Moussiades, 2020).

Figure 3.2 and the following paragraphs illustrate the architecture of a typical ASR system based on Filippidou and Moussiades (2020) and Karpagavalli et al. (2011), before the introduction of the recent, so-called End-to-End technology.

An acoustic front-end processes the speech signal to extract useful features. Several feature extraction methods are used for this purpose (Filippidou & Moussiades, 2020). The most used is the so-called Mel-Frequency Cepstral Coefficients process (Karpagavalli et al., 2011). This involves the generation of a feature vector. The extracted information is then processed by a so-called 'decoder', or 'search algorithm', using an acoustic model, a lexicon, and a language model. The product is a hypothesized word or phoneme:

> An acoustic model is a file that contains statistical representations of each of the distinct sounds that makes up a word. The lexicon includes terms of the vocabulary of the current application. The language model consists of the limitations associated with the sequence of words that are accepted in a given language (Filippidou & Moussiades, 2020, p. 74).

---

24 The Acoustic Phonetic Approach postulates that finite, distinctive phonetic units in spoken language exist and that these are broadly characterised by a set of acoustic properties which are manifested in the speech signal over time. Several steps are conducted from the analysis of the sound to the determination of a valid word; see Karpagavalli et al. (2011).

25 Artificial neural networks (ANN) are classifiers that are used to learn, organise and improve pattern recognition; Malik et al. (2021).

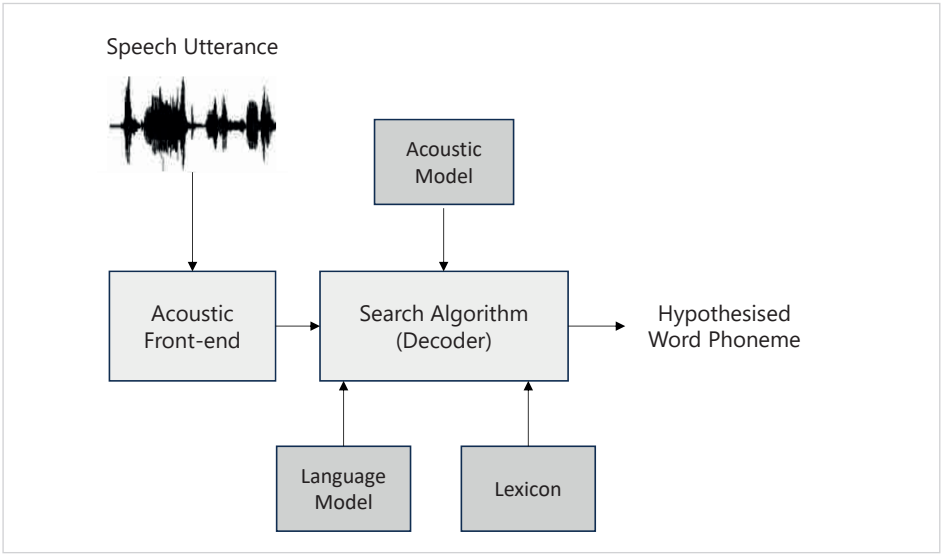


Figure 3.2: Conventional speech recognition architecture based on Filippidou & Moussiades (2020, p. 75) and Karpagavalli et al. (2011, p. 327)

The number of ASR systems is continuously increasing up to the present (Filippidou & Moussiades, 2020). Several international companies such as Microsoft SAPI, Natural and Google, and research groups, are currently working on ASR systems and continuing to develop these. Because conventional models have some disadvantages, e. g. the acoustic models are time-consuming to train and must be trained for each language, improved technologies have been innovated and introduced. The new models contain a technology that is based on an architecture composed of a whole and unique block that jointly optimises acoustic, lexicon and language models. The so-called End to End (E2E) model is an example of this technology. Figure 3.3 illustrates the combined processing of the previously separated procedures. One major advantage of this technology, in addition to the faster processing, is that it does not require training to a specific acoustic or language model (Ghayas et al., 2024).

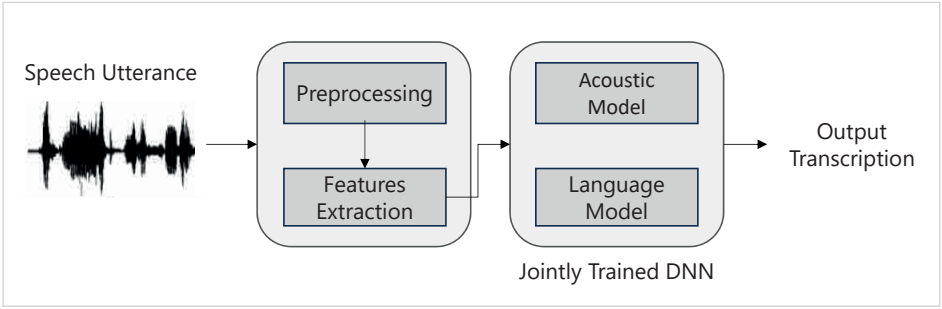


Figure 3.3: End-to-end ASR Architecture (Ghayas et al., 2024, p. 150) (simplified representation)

In sum, "speech recognition technology has improved rapidly, ranging from systems that needed to be adapted for individual users to more advanced programming that builds on deep learning algorithms …" (Matre & Cameron, 2024, p. 2). However, the question of whether this technology is appropriate for students with dyslexia is considered important for this thesis research.

Important in this regard is the improvement of ASR systems and, according to Kuhn et al. (2023), the reduction of the word error rate (WER)[26]. The application of new technological innovations such as End-to-End (E2E) models or self-supervised learning improves the accuracy of the tool. Google reported in 2017, for example, an accuracy rate for their speech recognition of 95 % (Palmiter Bajorek, 2019). Kuhn et al. (2023) concede that ASR systems outperform human transcriptions on some datasets, but they assume two problems with these high indications given by providers. Firstly, these indications tend to be too high because of overfitting for specific datasets and therefore perform lower on novel datasets. Secondly, these authors discuss studies that have investigated specific biases for speakers of under-represented groups, but Kuhn et al. (2023) deplore that there is no precise knowledge on a baseline of accuracy for the ASR systems. In addition, accuracy rates vary strongly between different providers. Feng et al. (2024) suggest, however, that research and developers should not focus blindly on lowering the error rates but should consider the speaker groups and demographics that are inherently present in society.

S.-E. Kim et al. (2024) and Kuhn et al. (2023) agree on the fact that the disparity between human and ASR performance has significantly diminished with recent developments in technology. However, they state that noise or variability in language are continuing issues of ASR. When these authors compared ASR-based intelligibility values with those of human listeners under some of the most challenging conditions, which were speech of second language (L2) speakers and in continuous broadband noise, they also found that the ASR system[27] outperformed human listeners on some accuracy measurements in spite of the variability in L2 speakers and the background noise. However, while the results are promising on some aspects of accuracy,

26 Word error rate (WER) is a metric used to evaluate the quality of transcriptions produced by Automatic Speech Recognition (ASR) systems. To calculate the WER, the 'minimum edit distance' between the reference text (original text) and the recognised transcript is measured. A distinction is made between substitutions (S), deletions (D), and insertions (I) as errors. The WER is the proportion of the total number of errors (S+D+I) over the number of words in the reference; see Kuhn et al. (2023); Schüler (2020).

27 The authors investigated four state-of-the-art ASR systems and found that the ASR system called 'WISPER' achieved the best overall accuracy values. The three other systems, GOOGLE SPEECH-TO-TEXT, HUBERT, and WAV2VEC 2.0, showed, overall, less positive results. The authors suspect this difference to originate in different training tasks and materials: while WISPER was trained with L2 speech in noisy environments, WAV-2VEC 2.0 and HUBERT were only pre-trained in English speech in quiet environments; see S.-E. Kim et al. (2024).

the authors still see challenges regarding observed 'hallucination errors' committed in speech input and think that further investigation of ASR systems is warranted to gather a deeper understanding of their strengths and limitations (S.-E. Kim et al., 2024; Kuhn et al., 2023). Other researchers in the field of ASR systems name several challenges in addition to L2-recognition. The following section addresses challenges that might impact empirical research on a German-speaking environment with primary school students with dyslexia.

3.2.2 Considerations about speech recognition for different user groups

Feng et al. (2024) show differences in state-of-the-art ASR systems in their performance according to speaker groups. These authors assume that many factors can cause bias, and investigate gender, age, and regional and non-native accents. However, before addressing these biases, the question about the performance of ASR systems in other languages is considered relevant for this thesis research. The above-discussed developmental advancements took place, on the whole, in English-speaking countries and on an English-language basis. The German-speaking environment, relevant for this thesis, requires an ASR system capable of processing another language than English.

While challenges remaining for English are stated by authors (e. g. S.-E. Kim et al., 2024), the recognition of other languages than English is even less satisfactory. Silber-Varod et al. (2021) assume, among other challenges, that ASR systems are language-dependent (Silber-Varod et al., 2021). These authors investigated four ASR engines[28] in the different languages of English (high-resourced), German (mid-resourced) and Hebrew (low-resourced), by four commonly used metrics, WER being one of them. Findings suggest that English ASR systems provide the best results. Surprisingly, the researchers find similar performances for ASR systems in Hebrew and German despite the estimation of German as a mid-resourced language. Kuhn et al. (2023) report also findings about Germany, comparing the German ASR-system with English L1 and English as L2. These researchers, interested in speech recognition as a possibility for access for deaf- and hearing-disabled people, analysed the accuracy of higher education lecturers among eleven ASR providers. Their findings show that accuracy is best in English L1, second in English L2, and the lowest in German. However, they state, firstly, variations in accuracy rates according to different factors, and secondly, that none are on top in every analysis.

The lower performance of speech recognition in German must be considered in the present thesis research, in which the state of research is mostly based on studies conducted in English, while the environment of the present thesis research is a German-speaking environment. In addition, to date there is an unknown proportion of the strength of any provider (Silber-Varod et al., 2021). For the present thesis

28 Google Cloud, Google Search, IBM Watson, WIT.

research this means, firstly, that there is no provider outperforming the others in a way that implies the selection of one tool or provider over another. Secondly, the strengths and weaknesses of current tools remain unknown to a certain extent.

Feng et al. (2024), whose goal is to uncover and mitigate bias in order to work towards a more inclusive automatic speech recognition for all people, show that, in addition to the bias of speakers in different languages, there are also biases for different speaker groups that are independent of the state-of-the-art ASR architecture. The investigated groups are namely different genders, age groups, native vs. non-native speaker groups, and dialect speaker groups. The two speech test modalities were a reading condition and a human-machine-interaction. These authors examined biases in Mandarin and Dutch; with the latter being considered informative for the present thesis research.

The results on Dutch indicate, firstly, an age bias. Among native speakers, the speech of adolescents was recognised the best, followed by adults, while children's speech reached the lowest rates. This is an important finding for the present thesis, since primary students are targeted (Feng et al., 2024). Higgins and Raskind (1995) consider that the higher pitch of younger students' voices may hinder accuracy. Feng et al. (2024) assume there is bias at the high volume and with the disfluency and/or hesitations in children's speech. For non-native speakers, however, the speech of children is found to be better recognised than that of adults.

The second bias is one against non-native speech, for whom the accuracy rate was significantly lower with an absolute WER degradation of 15 to 24 % according to the testing condition (Feng et al., 2024). This bias must also be considered for the present thesis, as the population under study is composed of both native and non-native German speakers. However, while the age error is disadvantageous for children, the non-native speaker error seems to have a less negative effect on children.

A third bias described by Feng et al. (2024) is one against regional accents, with the strongest biases against Flemish and speech from the south of the Netherlands. The authors assume the reason for this is a lack of the use of ASR training speech data for these regional accents (Feng et al., 2024). For two major reasons, this bias needs also consideration in the present thesis research. Swiss German is classified as a variety of German. Within the Swiss German variety, several regional accents are present and largely spoken in everyday life and in informal situations. Formal situations and written language take place, with a few exceptions, in Standard High German which is also the language of education. According to many authors, the mastery of both, Standard High German and Swiss German Dialects, is necessary in Switzerland (Berthele, 2019). At this point, two assumptions are made that might impact the accuracy of speech recognition. Firstly, Standard High German, learned as the educational language in Switzerland, is characterised by peculiarities in contrast to the standard language in other areas of Germany or Austria (Variantengrammatik, 2020). The German-speaking part of Switzerland is considered a relatively small region of the entire German-speaking area. It is therefore assumed that speech

data of Standard High German spoken by Swiss people is probably underrepresented. Secondly, due to the situation of diglossia in German-speaking Switzerland, the Standard High German spoken by Swiss people and especially by primary students might be less fluent than the German in Germany and Austria. Therefore, in addition to the comparable situation of a lack of Flemish ASR-training, speech recognition in the German-speaking region of Switzerland might be affected by disfluency or hesitations.

According to Feng et al. (2024), a last bias is the gender bias. Because earlier research results on gender bias cited by these authors are mixed, the present thesis research only presents the findings reported by Feng et al. (2024) regarding children and adolescents, focussed on by this thesis research. These authors find a bias against male adolescents in the reading condition and a bias against female children in human-machine interaction.

In sum, speech recognition for the user group forming the focus of the present research may be affected by certain factors as presented above. The primary students with dyslexia need a speech recognition software that works for German; and they may, due to their age, tend to speak less fluently than adults. This could be particularly true for those students with comorbidities such as oral language disorders or AD(H)D. In a representative sample of primary school students with dyslexia, both genders, as well as mono- and multilingual students, are likely to be present. These many factors play a role in the accuracy, and thus the performance, of speech recognition. They must, therefore, be considered carefully in the approach and the design of the present study.

3.2.3 Definition and characterisation of assistive technology

The rationale, elaborated at the beginning of this chapter, of speech recognition as the most consequential technology to assist students with dyslexia, on their journey through school, professional career and everyday life, is based on the concept of using STT as an assistive technology. However, the development of the technology, as shown in Section 3.2.1, reveals that this technology was not, or not specifically, intended for people with disabilities. While studies have emerged in the field of special education on STT as an assistive technology from the late 1980s onwards (Matre & Cameron, 2024), other sectors, such as industry or aviation, have also been interested in this technology (Hollier, 1983). This illustrates that whether a technology is considered as assistive or as mainstream technology goes far beyond the focus on machines (Blackhurst, 2005). Matre and Cameron (2024) state that STT can be applied as an assistive, a remediation or an instructional tool. Thus, it depends on the aim or intended effect of implementing the tool in education or research. However, Matre and Cameron assume that there has been insufficient research about theoretical, empirical and practical distinctions between STT as an assistive or as

an instructional technology, i. e., a learning medium. Therefore, the following paragraphs frame the theoretical approach taken in this thesis research.

C. Abbott (2007) presents three approaches to using digital technologies, elaborating the question of the inclusion of people with disabilities through access to digital technologies. The first category in this work about e-inclusion comprises technologies to train or rehearse. C. Abbott criticises this kind of much used technology, which focuses on drill and practice, as being limited because it is often associated with a behaviourist model of learning that does not sufficiently consider social inclusion and social justice. C. Abbott categorises the second approach as using technology to assist learning. He firstly points out that the term, assistive technology, is problematic since it is differently defined in various contexts. His second critique is that the use of a technology linked to the need to compensate is as an adjunct to learning rather than a key agency through which learning takes place. The third category comes closer to what he calls a 'catalyst for learning itself'. The author describes this category as technology to enable learning. This use of technology makes learning possible where it was not possible before, because it facilitates the creation of collaborations and communities. This approach, associated with social-constructivist models of learning, differs from the others because the use of technology transforms rather than modifies the learning context.

In considering these categories, he recognises the limitations of drill and practice as well as purely compensational software and discerns the potential of socially collaborative use of digital technologies (C. Abbott, 2007). This separation between assistive and enabling technologies is not emphasised in the definition of 'assistive technologies' (AT) by D. L. Edyburn (2015, p. 9) who focuses on AT and frames the term as follows: "Assistive technology has the potential to enable people with disabilities to live, learn, and work more independently through the application of specialized technologies that reduce, eliminate, or minimize the impact of a disability."

This definition is interpreted in such a way that the participation is directly fostered both by compensating for impaired functions and, at the same time, by increasing learning through the use of the technology. Therefore, this thesis research draws on this definition of AT and strives to investigate the question of the effectiveness of the use of speech recognition for students with dyslexia accordingly. In terms of writing, compensation (assistive technology) refers to the efforts to compensate for a lack of written skills and enhance the student's ability to plan, compose and revise text. By contrast, remediation or instructional technology aims to improve skills by enabling students to produce more text and increase exposure to writing activities (Matre & Cameron, 2024).

The present investigation of a technology intervention for a particular group of students confirms the classification of STT as an assistive technology because only the primary beneficiaries benefit from the intervention. However, STT and TTS are also conceivable as being applications within the framework of universal design for learning (UDL). A technology has quantifiable evidence as an UDL-application

when it is used by most students in a classroom and they benefit from it. UDL implies that instructional materials and environments are accessible for all students (D. L. Edyburn, 2021). STT is, next to TTS, an example of a UDL that also affords benefits to secondary beneficiaries, i. e., further learners might improve in their learning due to these technologies. D. L. Edyburn (2021, p. 311) states, "every technology platform has universal accessibility features that support the writing composition process by allowing users to speak their thoughts and have the device correctly spell the words". Therefore, STT can be considered as a UDL from a provider perspective (D. L. Edyburn, 2021); but dictation also meets the framework of UDL based on the legal perspective. The 'Individuals with Disabilities Education Act' in the USA asks for the provision of flexibility: in the ways in which information is presented; in the ways students respond or demonstrate knowledge and skills; and in the ways students are engaged. It also demands a reduction of the "barriers in instruction, provides appropriate accommodations, support, and challenges, and maintains high achievement expectations for all students, including students with disabilities and students who are limited English proficient" (D. L. Edyburn, 2010, 34).

3.3 Reasons for STT as assistive and/or enabling technology

People are using speech recognition technology more and more frequently. Such technology has demonstrated its enormous potential in Human-Computer Interaction (HCI) systems and is now playing an important role to assist people in various fields (Dhanjal & Singh, 2024). Examples where speech technology nowadays is used are speech commands in the car or in smart homes, in the military domain, in industry, and more recently in oral search commands and usage of STT in social media where videos are aggregated by text. In addition, the healthcare system and people with disabilities are increasingly benefitting from this technology. People affected by deafness or hearing disabilities use ASR systems to transcribe oral speech into written text. People with physical disabilities can command their computer and search on the internet without using their hands, and people with visual impairment use TTS and STT to listen to and produce text (Chen & Jokinen, 2010; Dhanjal & Singh, 2024). An increasing number of studies about students struggling with reading and writing demonstrate the possible benefits of such technology for students with disadvantages at school, such as dyslexia or learning disabilities.

Before the presentation of the state of literature in Chapter 4, three major theoretical approaches are discussed in the following sections. Chapter 2, Section 2.1 illustrated the complexity of writing and stated the enormous demands occurring simultaneously for the several writing processes as well as for the executive functions, working memory etc. This demand for cognitive functions is also known as the cognitive load. Section 3.3.1 following therefore asks whether Cognitive Load Theory (CLT), by Sweller (1994), gives reasons to support the application of STT for students with dyslexia. Section 3.3.2 addresses general motivation and writing

motivation, which is illustrated as another component that plays an important role in the writing process. In addition, the description of secondary symptoms of dyslexia, elaborated above, potentially affect self-esteem, self-efficacy and, thus, writing motivation. Section 3.3.3 moves away from the (neuro)cognitive and sociocognitive theories of learning and seeks reasons in favour of or against STT for this group from an educational perspective. To this end, this section discusses consequences of digital tools such as STT, for literacy and educational settings.

3.3.1 Cognitive Load Theory

3.3.1.1 Theoretical framework

Fayol wrote, in 1999, that writers must decrease demands that writing processes place on working memory to avoid overloading their limited working memory. This assumption is based on CLT which claims that resources in the working memory are limited, and that working memory can be overloaded. This theory assumes that instruction must adapt to this and reduce the cognitive load, to facilitate learning. The following paragraphs present the current version of CLT and illustrate the relationship between the cognitive load and the working memory. The importance of cognitive load considerations for the writing process, writing instruction, and the compensation of lower-order processes to improve higher-order processes, are discussed.

CLT, introduced by the Australian educational psychologist, John Sweller in the late 1980s, focusses mostly "on how the objective characteristics of the task affect cognitive load and, in turn, learning" (Moreno & Park, 2010, p. 11). It is a psychological theory that attempts to explain psychological or behavioural phenomena resulting from instruction. Skills, such as writing skills, are psychological constructs. The theory goes back to early psychological theories that recognise the multidimensional nature of the mental load which results from the interaction of subject individual characteristics and objective task characteristics. Since CLT is concerned with the question of how to design instruction to promote problem solving, it is now also referred to as an instructional design theory. Other theories, such as problem solving and schema theories[29], inspired CLT from its inception and are important concepts of the theory (Sweller, 2023).

An integral part of CLT is a basic assumption that heavily influences the theory: the theory assumes that human cognitive architecture is characterised by a very limited short-term memory, and a very large long-term memory. 'Short-term memory'

29 "A schema is a cognitive construct that organizes the elements of information according to the manner with which they will be dealt"; see Sweller (1994, p. 296). "Knowledge and intellectual skill based on knowledge is heavily dependent on schema acquisition"; Sweller (1994, p. 297). Sweller (1988) concludes that different types of analysis to solve problems leave more or less cognitive resources for schema acquisition.

is now mostly known as 'working memory' (Moreno & Park, 2010). This assumption differs between automatic and controlled processing, which are two qualitatively distinct ways to process information. Both processes must work together to enable learning. Since the awareness of this interplay is considered a prerequisite for the elaboration of CLT, working memory and long-term memory are described as follows, before the three categories of cognitive load are presented.

3.3.1.2 Working memory and long-term memory in CLT

Moreno and Park (2010, p. 14) summarise controlled processing as "relatively slow, mentally demanding, and highly dependent on working memory capacity". This controlled part of memory, where all conscious processes occur, is called working memory or short-term memory. It can only handle a very limited number of novel interacting elements, possibly no more than two or three. According to Paas et al. (2003), this number is considerably smaller than the number of interacting elements that most substantive areas of human intellectual activity contain. On its own, working memory would thus allow only relatively trivial human cognitive activities.

However, long-term memory completes the working memory by providing humans with the ability to vastly expand the processing ability (Paas et al., 2003). Moreno and Park (2010, p. 14) describe this automatic processing, in contrast to the working memory, as "fast, effortless, not limited by working memory capacity, and developed with extensive practice". In other words, the long-term memory can store information and can contain vast numbers of schemas. Schemas are "cognitive constructs that incorporate multiple elements of information into a single element with a specific function" (Paas et al., 2003, p. 1). These constructs may consist of a large number of lower-level, interacting elements. The incorporation of these elements into a schema and stored in the long-term memory allows people to bring the schema back to the working memory where it is processed and manipulated as one whole element instead of several (Paas et al., 2003). This process, as well as the automation of schemas, reduce the load for working memory during problem solving. The storage of automated schemas is therefore a function of learning, according to CLT (Sweller, 1994). Especially when one must deal with higher cognitive activities, schema acquisition and automation are major learning mechanisms because they are designed to circumvent the limited working memory (Sweller & Chandler, 1994).

3.3.1.3 Categories of cognitive load

Paas et al. (2003) describe three categories of cognitive load: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Figure 3.4).

The intrinsic cognitive load is inherent to the complexity of the material and information, and depends on the learner's prior knowledge (Sweller, 1994). The demands on the working memory capacity depend on the extent to which relevant

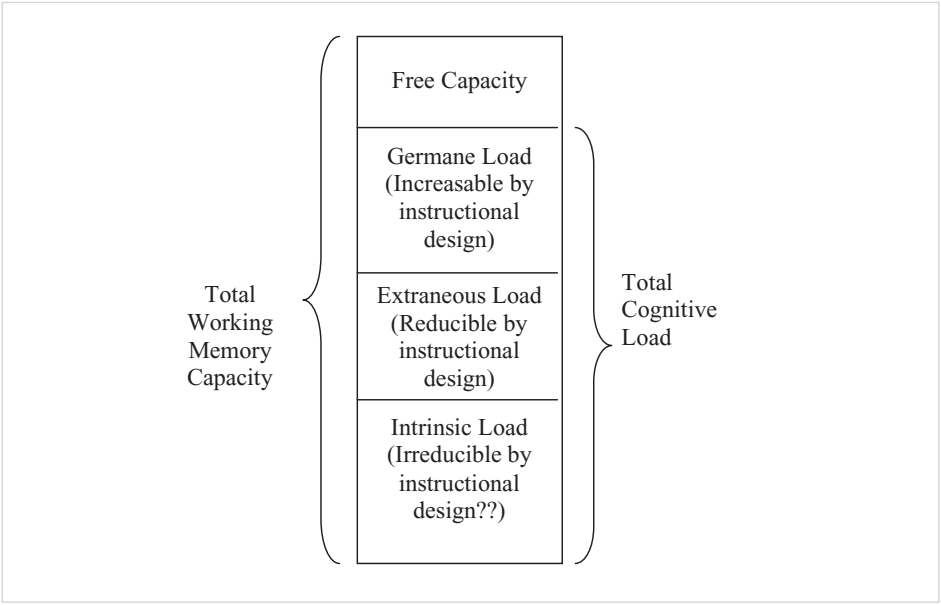


Figure 3.4: Visual representation of the assumptions underlying the CLT development (Moreno & Park, 2010, p. 18)

elements, which must be processed, interact. These demands are intrinsic to the material being learned, and figure as a base load. It is currently considered questionable in the literature, and thus labelled with question marks in Figure 3.4, whether and to what extent intrinsic cognitive load can be altered. It can be argued that instructional manipulations cannot alter the intrinsic cognitive load, because different materials have different levels of element interactivity. However, additional schemas and automating previously acquired schemas could theoretically reduce the intrinsic cognitive load to some extent (Moreno & Park, 2010; Paas et al., 2003).

The extraneous cognitive load, also called ineffective cognitive load (Paas et al., 2003), is caused by the way in which information is presented, i. e., the instructional design used. Since many conventional instructional procedures were developed without any consideration or knowledge of the structure of information or cognitive architecture, such procedures may contain activities that are irrelevant to schema acquisition and automation (Moreno & Park, 2010). In this case, working memory resources are used in vain. Thus, the cognitive load can be reduced by alternative instructional designs and procedures that reduce this extraneous cognitive load (Paas et al., 2003).

The third category is the germane or effective cognitive load. The germane cognitive load is, like the extraneous cognitive load, influenced by the design of the instruction. The difference between both lies in how instruction influences the load. Instruction enhances learning by the germane cognitive load, while instruction

interferes with learning due to the extraneous cognitive load. Increased effort or motivation can increase the cognitive resources devoted to a task (Paas et al., 2003).

As Figure 3.4 illustrates, the "intrinsic, extraneous, and germane cognitive loads are additive in that, together, the total load cannot exceed the working memory resources available if learning is to occur. The relations between the three forms of cognitive load are asymmetric" (Paas et al., 2003, p. 2). Thus, by using a more effective instructional design, the extraneous cognitive load can be reduced and thus the germane cognitive load can be increased. Consequently, learning through schema acquisition and automation might increase, which could therefore reduce the intrinsic load and the total cognitive load. This therefore freed working memory capacity allows the learner to use the newly learned material in acquiring more advanced schemas. The repetition of this cycle allows one to acquire advanced knowledge (Paas et al., 2003).

Instructional techniques that have been investigated to reduce extraneous cognitive load are goal specificity, worked examples, completion, split-attention, redundancy, and modality effects. While various techniques are effective for different learning activities and in different stages of expertise, their effectiveness alters with the development of experience in the learner. However, two scaffolding techniques consider this challenge: "the intrinsic aspects of cognitive load can be reduced by the scaffold of simple-to-complex sequencing, whereas the extraneous aspects can be reduced by providing the substantial scaffolding of worked examples initially, followed by completion problems and then full problems." (Paas et al., 2003, p. 3).

Whether the amount of extraneous cognitive load is important depends on the intrinsic element interactivity. If this is low, extraneous cognitive load caused by instructional design may not be important. However, by contrast, when dealing with materials with an intrinsically high element interactivity, the extent of the extraneous cognitive load is important. Intrinsic load can be estimated by the number of elements that must be considered simultaneously in the learning activity (Sweller & Chandler, 1994). Chapter 2, Section 2.1 emphasised the complexity of writing, and illustrated how many, mostly mental processes must be performed simultaneously. Since several of these processes are highly effortful and entail the temporary storage and processing of information, working memory plays a crucial role in text production (Kellogg, 2001; Linnemann et al., 2022). In the following, what this means for writing processes is discussed from a CLT perspective, and especially for writers with dyslexia. The following discussion also mentions possibilities and limitations for instruction or compensation.

3.3.1.4 Writing processes from a CLT perspective

Olive (2014, p. 175) writes, "producing written language requires converting ideas into well-elaborated language before transforming it into visual signs via hand and finger movements". The writer must therefore apply both lower-order and higher-or-

der processes. At the same time, as Kellogg et al. (2013) state, composing a written text entails far more than language production, and they refer, in addition to Olive's observations, further elements such as decision making or monitoring. Due to all these demands, writing places considerable demands on the working memory: knowledge and beliefs from long-term memory and external information are brought into working memory. There, all this information is processed and acted upon, while attention, writing processes, writing tools, motivations, emotions, personality traits, and the environmental and social situation in which writing takes place, must be regulated (Olive, 2014).

As emphasised above, CLT claims that working memory is a limited and temporary storage system (Sweller, 1994). However, as long as these processing demands do not exceed working memory capacity, the writing processes can be orchestrated (Olive, 2014). As stated above, schemes and automation free working memory capacity. An example of automation regarding writing is a relatively automatic motor execution. This helps writers to attend more to the higher-order writing processes required in text composition (Olive & Kellogg, 2002). By contrast, when motor processes, i. e., handwriting or typing, consume too much of the available attention, writers perform more weakly in the higher-order processes. Therefore, Linnemann et al. (2022) conclude that handwriting is associated with saving cognitive resources. According to several authors, this applies not only to handwriting or typing but also to spelling (Fayol, 1999; Olive & Kellogg, 2002).

Students with dyslexia often experience challenges in this regard. Thus, they consume more cognitive resources for spelling in the writing tasks (Beers et al., 2017). As a result, the working memory capacity available for the above-mentioned further cognitive processes in writing is limited. It is therefore emphasised in this thesis that there is a risk on two levels. Firstly, the completion of higher-order writing processes is hindered by the limited free resources. Secondly, the repetition of the cycle of learning through acquisition and automation of schemes, considered necessary to gain knowledge (Paas et al., 2003), is affected. In sum, the hampered lower-order writing processes in students with dyslexia bear the risk of hindering also the higher-order writing processes. The following discussion raises the question of how this risk, for the acquisition of higher-order writing processes, can be mitigated for students with dyslexia.

3.3.1.5 Deduction from CLT perspective regarding the application of STT for students with dyslexia

When the limited working memory capacity is at risk of overloading, the demands placed on it must be decreased (Fayol, 1999). An adequate instructional design can free working memory capacity. While there is less agreement on the question of whether intrinsic cognitive load can be altered – some authors claim that reducing intrinsic cognitive load artificially reduces also understanding, in other words,

a reduction in the material hurts learning – there is agreement that the extraneous load can be decreased and germane cognitive load can be increased by adapted instructional design (Paas et al., 2003). Referring to writing, in particular lower-order processes such as handwriting and spelling are favourable for automation (Fayol, 1999). A. Sturm and Weder (2016) write, concerning spelling instruction, that spelling develops and automates partly in an informal and casual way through many writing experiences, but that it nevertheless needs formal instruction. These authors name three instructional principles to foster spelling in a formal and explicit way: a) guided practice; b) systematic instruction of rules, application of spell-based revision checks, and strategies; and c) reflective activities for a deeper understanding of the spelling system. An important pillar of automation to keep in mind is, according to writing researchers, not only specific instruction but also the time on task (S. Graham & Harris, 2019b). Automation only occurs with many hours of practice. This is true for both lower-order processes and higher-order processes. In addition to the writing time, specific instruction also for higher-order tasks is suitable for reducing cognitive load. As mentioned in Chapter 2, Section 2.1, writing strategies divide the process into portions that facilitate the task, and which can be used in similar assignments (A. Sturm & Weder, 2016). This is emphasised as an example of constructing schemes to affect the extraneous and germane cognitive loads in a beneficial way.

However, the Chapter 2, Section 2.2 has demonstrated in detail that dyslexia is characterised by specifically poor skills and automation of spelling that remain problematic throughout the lifespan and result often in secondary symptoms. Despite remediation and adaptive instruction, these students achieve only reduced levels of automation of the lower-order writing processes. The often-experienced challenges in transcription skills limit the working memory capacity which is therefore not available for other cognitive processes in writing, such as idea generation and organisation (Linnemann et al., 2022; Yeung, 2022). Sturm and Rankin-Erickson (2002) contend that technology interventions can make writing less onerous for students by providing supports that reduce the cognitive burdens placed on them.

Therefore, this thesis emphasises that, from the CLT perspective, it is reasonable to assist spelling to reduce the extraneous cognitive load. STT, which produces written text from spoken text and, thus, takes over the lower-order writing processes, is argued to be the most consequential technology to assist spelling. While the intrinsic cognitive load of the complex tasks of writing is assumed to remain high, STT could free some cognitive load by relieving the lower-order processes. In the words of Paas et al. (2003), the freed cognitive resources can then be used to engage in schema acquisition activities. This thesis does not provide arguments about any impact of the technology on the germane cognitive load of writing. Nevertheless, it assumes that the germane cognitive load could, and should, be fostered by a suitable instructional design of writing instruction. Specific writing instruction has been shown to be effective for all students and for struggling writers, such as instruction on writing strategies or self-regulation strategies (Harris et al., 2003). It would, therefore, foster

the acquisition of higher-order processes and subsequently increase the germane cognitive load.

This thesis does not, however, question the importance of remediation and specific, adapted instruction for students with dyslexia. While STT assists the text production, the writer still must be knowledgeable for the text revision.

3.3.2 Writing motivation

Chapter 2 has emphasised at several points that the writing ability is a difficult and complex competence. This fact is important in considering writing motivation because motivational issues will assume particularly prominent status. Bruning and Horn (2000, p. 26) state, "writers need to develop strong beliefs in the relevance and importance of writing and, as they grapple with writing's complexities and frustrations, learn to be patient, persistent, and flexible".

Camacho et al. (2021), in their systematic research on empirical studies of writing motivation[30] published between 2000–2018, describe a growth of publications on the topic of writing motivation over this period. They interpret this growth as a sign of an increasing awareness of the importance of motivational constructs in writing research and point out that many researchers have highlighted the influential role of motivation in writing. However, while Camacho et al. (2021) and Ekholm et al. (2018) praise this growing research interest in the field of writing motivation, they are critical about the conceptualisation and terminology related to writing motivation. This difficulty is not only present on the side of empirical research, but also on the side of capturing and illustrating writing motivation in a writing model. Motivation figures in the latest model from Hayes (2012) who is convinced that researchers must learn how to integrate motivation into models of writing, to adequately account for how people write. However, since Hayes published his latest model in 2012, some doubts have remained about the adequacy of this model to capture the influence of motivation on writing. Therefore, the discussion in Section 3.3.2.1 attempts to clarify the concept of writing motivation for this thesis by elaborating, firstly, the difficulty of defining the concept of writing motivation. Secondly, this discussion following presents the rationale for the selection of an adequate conceptualisation required for the present thesis. It subsequently presents the model for writing motivation by Wright et al. (2019). Section 3.3.2.2 gives an overview on students' development of writing motivation throughout the school years. Lastly, Section 3.3.2.3 delves into the relationship between writing achievement and writing motivation regarding group differences and the effects of digital tools.

30 This thesis does not differ among the three terms, "motivation to write", writing motivation", and "motivation towards writing", but uses them synonymously, according to context.

3.3.2.1 Framing the construct of writing motivation

According to the above-mentioned systematic reviews by Ekholm et al. (2018) and Camacho et al. (2021), there are various matters that complicate comprehension of the construct of motivation to write. Firstly, Ekholm et al. (2018) assume that researchers have not systematically declared the underlying concepts. This may have led to measurements and definitions that fluctuate from study to study, and it is therefore questionable how researchers could fit the construct into a coherent theory with any explanatory power. Secondly, Camacho et al. (2021) report a profusion of implicit definitions of motivational constructs. Ekholm et al. (2018) also report, as a third complicating matter, that some research situates motivational constructs such as self-efficacy as components of writing attitudes, i. e., self-efficacy as part of writing attitudes, whereas other research subsumes writing attitudes and self-efficacy as two distinct components on the same level within general motivation. Therefore, both Camacho et al. (2021) and Ekholm et al. (2018) attempt to clarify these concepts to conduct meaningful research.

According to Ekholm et al. (2018) it is essential to define the construct of motivation and to reach conceptual and operational agreement among researchers on what 'writing attitudes' means. To the same end, Camacho et al. (2021) consider, firstly, generic motivation. They synthesise motivation as a multidimensional construct comprising several interrelated factors such as emotions, needs, beliefs, values, desires, and goals. Several different theories of motivation, for example the social cognitive theory (Bandura, 1997), the self-determination theory (Ryan & Deci, 2000), the expectancy-value theory (Eccles & Wigfield, 2002), or the interest theory (Hidi & Renninger, 2006) have influenced conceptualisations of writing motivation. Wright et al. (2020) agree on the expectancy-value theory and the sociocognitive theory, together, explaining the various components of writing motivation; but these authors do not name further theories in their study about writing motivation in middle school. In contrast to the list of theories by Camacho et al. (2021), Boscolo and Gelati (2013) disagree on the usefulness of the self-determination theory to understand writing motivation. In that theory, the terms 'intrinsic' and 'extrinsic' motivation are widely used to explain human behaviour. Boscolo and Gelati (2013, p. 288) argue "that a motivated-to-write student is not necessarily eager to write, but, with differing degrees of awareness, wants to write because he or she thinks that writing is a worthwhile activity, although not always an enjoyable one". They consider that the conceptualisation of writing motivation requires a more complex view; which is why this thesis does not further explore the terms of intrinsic and extrinsic motivation although as stated by Wright et al. (2019), it is not questioned here that motivation includes internal and external forces that influence an individual's decisions to engage in and persist through tasks.

Regarding the tasks themselves, interest and task value are components of human motivation to which researchers accord an important role for writing motivation.

Task interest and value are situated within the expectancy-value theory, highlighted by both Wright et al. (2019) and Camacho et al. (2021) and were briefly mentioned within Chapter 2, Section 2.1.2 about writing instruction and the writing task. Task interest and value are emphasised as important to be included in the design of writing instruction and planning of the writing task (Boscolo & Gelati, 2013; Eccles & Wigfield, 2002; Troia et al., 2013).

In addition to their theoretical approach, Camacho et al. (2021) attempt to clarify the key concepts from research in the field, on the one hand, by identifying motivational constructs, and on the other hand, by analysing explicit versus implicit definitions of motivational constructs. They identify 24 motivational constructs that confirm the multidimensional nature of writing motivation. Out of these, they select 12 motivational constructs, with explicit definitions provided by experienced scholars and summarise that researchers have focused on generic motivational constructs (e. g., attitudes) as well as on specific ones (e. g., self-efficacy). The two most studied constructs are self-efficacy and attitudes, which are discussed as follows.

In contrast to the approach above used to shed light on the complex construct of writing motivation by the means of systematic reviews, Wright et al. (2019) investigate the multiple dimensions of writing motivation through a combined approach, based upon a combination of theory and psychometrics, with the goal of proposing a model of writing motivation that aligns with current theories of motivation. Their results overlap with constructs described by the expectancy-value theory[31] (EVT) and aspects of Conradi et al.'s (2014) proposed hierarchy of reading motivation-related constructs. In relation to the latter study, Wright et al. (2019, pp. 74–75) observe:

> Their hierarchy conceptualized two categories related to beliefs – those about the self and those specific to the task. Furthermore, subsumed within beliefs about the self were self-efficacy and self-concept. Similarly, our model suggests that *self-beliefs as a writer* are comprised of students' self-concept and self-efficacy" [italics in original].

In contrast to EVT, Wright et al. indicate that attitude is a separate construct from students' beliefs towards writing and themselves as writers, as illustrated in Figure 3.5.

The theoretically based factors found by Wright et al. (2019) are '*beliefs* related to the task' and '*pre-dispositions* towards the task'. Wright et al. (2019) firstly defined a model with theoretically based factors for 41 items established for measuring effects on writing motivation. Then, they conducted both exploratory (EFA) and confirmatory (CFA) factor analyses to develop and confirm the model that best fitted the

31 EVT assumes that human behaviour can be explained by a person's beliefs about their ability to be successful at a task (i. e., expectancies) and the value they place upon that task (Eccles & Wigfield, 2002). *Writer Self-Beliefs* are representative of the former, and *beliefs about writing* specifically addresses the value placed upon the task; Wright et al. (2019, p. 75)

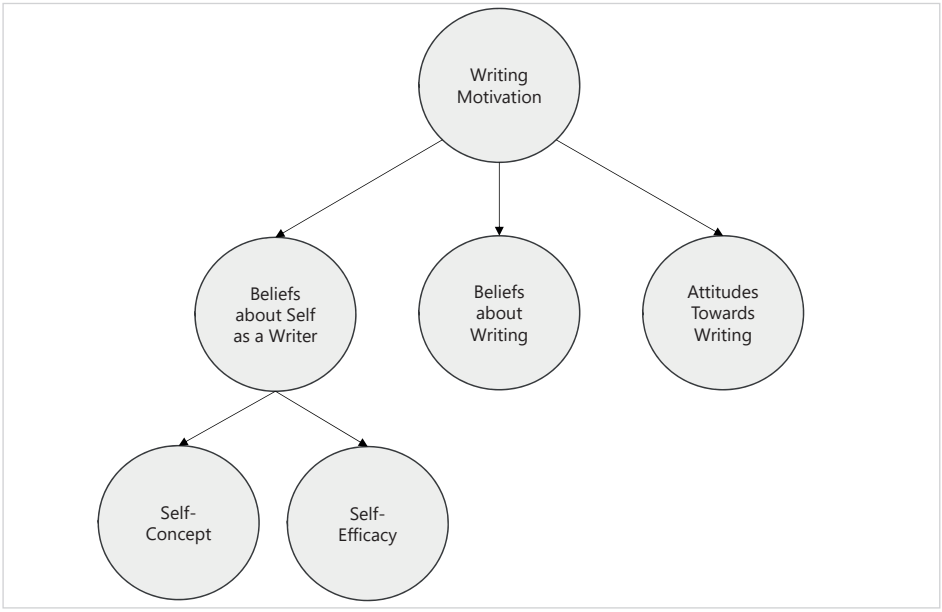


Figure 3.5: Proposed model of writing motivation according to Wright et al. (2019, p. 74)

data. The results of their analysis indicate a psychometrically based model as being the strongest fit. This model is based on 30 items that correlate logically and align with existing theory. It contains three factors, which are beliefs about self as a writer, beliefs about writing, and attitudes towards writing. In addition, the factor 'beliefs about self as a writer' contains two sub-factors: 'self-efficacy' and 'self-concept'.

The request to disclose underlying constructs, the lack of precise definitions of motivational constructs, and the relatedness of the variables pointed out by Camacho et al. (2021), Ekholm et al. (2018), and Pajares and Valiante (1997), are taken into account in this thesis in applying Wright et al.'s (2019) model. Therefore, the following discussion addresses the components of writing motivation according to this three-factor model. Furthermore, it is thus acknowledged in this thesis: firstly, that conceptualisations are disparate and include combinations of affective, motivational, and cognitive components; and secondly, that motivation and writing motivation are complex and multifaceted constructs (Wright et al., 2019).

Beliefs about self as a writer

'Beliefs about self as a writer' includes, in the model by Wright et al. (2019) illustrated in Figure 3.5, the two subfactors, self-efficacy and self-concept, which are presented separately.

Self-efficacy

Bandura (1997) explains that effective personal functioning is not simply a matter of knowing what to do and being motivated to do it, nor is efficacy a fixed ability that one does or does not have in one's behavioural repertoire. Rather, efficacy is a generative capability in which cognitive, social, emotional and behavioural subskills must be organised and effectively orchestrated to serve innumerable purposes. Bandura (1997, p. 3) emphasises the importance of self-efficacy because people's beliefs in their efficacy influence:

> "the courses of action people choose to pursue, how much effort they put forth in given endeavours, how long they will persevere in the face of obstacles and failures, their resilience to adversity, whether their thought patterns are self-hindering or self-aiding, how much stress and depression they experience in coping with taxing environmental demands, and the level of accomplishments they realize".

Since beliefs in personal efficacy depend on contextual factors, they vary according to the specific situation. Some situations require greater self-regulatory skills and more arduous performance than others (Bandura, 1997).

Klassen (2002) summarises that self-efficacy describes how someone perceives their own capability to perform specific tasks, which strongly influences their engagement in a task and its successful completion. Klassen points out that the self-efficacy beliefs are more task-specific than related constructs such as self-concept. Bandura (1997) specifies, furthermore, that they are established through normative criteria rather than through comparison with others. While self-esteem is concerned with judgments of self-worth, perceived self-efficacy is concerned with judgments of personal capability.

Regarding tasks in the domain of written expression, their demands are manifold, and correspondingly, the need to believe in one's own capabilities to monitor and execute the individual skills is high (Klassen, 2002). Writing self-efficacy refers, therefore, to the belief that one has the skills and abilities to overcome a writing task, even when that task is challenging (Wright et al., 2019). Klassen (2002, p. 174) illustrates this by the exemplary item "How confident are you in writing a paragraph on this subject?" which may be included in a scale measuring writing self-efficacy.

Self-concept

Self-concept for writing is a more global evaluation of one's abilities. It relates to whether writers generally consider themselves as strong or struggling writers (Wright et al., 2020). Conradi et al. (2014), who claim in their work on 'motivation terminology' in reading research that the term self-concept has often been misused, stress that the term, self-concept, properly denotes an individual's general self-perception related to a domain. Thus, it denotes a broader consideration than self-efficacy. Corresponding to the reader self-concept by Conradi et al. (2014), in relation

to writing, one possesses a self-concept as a writer. Boscolo and Gelati (2013), who write about 'self-perception' rather than 'self-concept', confirm the frequent synonymous use of 'self-efficacy' and 'self-perception', although these refer to two different motivational variables. They define self-perception of writing as an individual's evaluation of their writing ability. Therefore, scales measuring writing self-concept or writing competence might ask "Are you a good writer (compared to others in your class)?" (Klassen, 2002, p. 174) or "I am able to express my ideas clearly in written form" (Boscolo & Gelati, 2013, p. 286).

Beliefs about writing

Wright et al.'s (2021) model describes 'beliefs about writing' as a separate construct from the above-discussed 'beliefs about self as a writer' and from 'attitudes towards writing'. While expectancy-value theorists emphasise the more prominent role of perceived value constructs, Bandura (1997) considers that individuals who expect success in a particular enterprise tend to value those enterprises. Therefore, he argues, the perceived usefulness or outcome expectations are unlikely to make a strong independent contribution if self-efficacy perceptions are controlled (Pajares & Valiante, 1997). However, according to Bruning and Horn (2000), writers embrace a range of important motivation-relevant beliefs that go far beyond ideas of beliefs in the value of writing and in their own writing competence. Writers must develop strong beliefs in the relevance and importance of writing to carry themselves through the difficult and often emotion-laden process of writing.

This thesis understands 'beliefs about writing' as being the perception that writing has value and the beliefs regarding its importance for achieving academic and vocational goals (Bruning & Horn, 2000).

Attitudes towards writing

The definition of domain-general attitudes as "a learned predisposition to respond in a consistently favourable or unfavourable manner with respect to a given object" (Fishbein & Ajzen, 1975, p. 6) appears to be widely accepted and used in a variety of educational contexts (Ekholm et al., 2018). Nevertheless, the relationship between attitude and motivation has been somewhat unclear in the literature; and even though, or because, it has been the second most studied concept of writing motivation, measurements and definitions have fluctuated. While some researchers have considered attitude and motivation as related but theoretically distinct concepts, others considered attitude as a subconstruct of motivation (Ekholm et al., 2018). The analyses by Wright et al. (2019, p. 75) "indicate that attitude is a separate construct from students' beliefs towards writing and themselves as writers". Ekholm et al. (2018) point out that several researchers (e. g. S. Graham, Berninger, & Abbott, 2012; Troia et al., 2012) have defined writing attitudes as a primarily affective construct

that describes how a person feels while writing or about writing. Some definitions also include further psychological components and other motivational constructs.

For an explicit definition of attitudes towards writing, this thesis refers to S. Graham et al. (2007) who conceptualise these attitudes along a continuum with negative and positive extremes. These authors define writing attitudes as an affective disposition involving how the act of writing makes the author feel, ranging from happy to unhappy. Writing attitude cannot be described as a personality trait, because it is not that stable. However, it cannot be equated with a temporary emotional state either, as it is more stable. Therefore, it is more likely to lie somewhere in between the two (S. Graham et al., 2007).

Considering attitudes towards writing, two important points must be emphasised. Firstly, attitudes themselves are domain-specific and should therefore be studied as such (S. Graham, Berninger, & Abbott, 2012). Secondly, writing at its core is a social activity that takes place within a writing community. According to Camacho et al. (2021), S. Graham's recent Writer within Community (WWC) model of writing (see Chapter 2, Section 2.1.1.1) provides a useful framework for understanding writing attitudes, as it merges sociocultural and cognitive perspectives to offer a more complete understanding of how writing occurs.

3.3.2.2 Development of writing motivation

An influential article on developing writing motivation by Bruning and Horn (2000) proposes four clusters of conditions to enhance writing motivation in the classroom, some of them mentioned in Chapter 2, Section 2.1.2.2 about writing instruction and the writing tasks. These clusters are: nurturing functional beliefs about writing; fostering engagement using authentic writing tasks; providing a supportive context for writing; and creating a positive emotional environment. Furthermore, the teachers' own conception of writing plays a crucial role for adequate conditions in writing contexts. On the one hand, this list of areas to enhance writing motivation suggests that writing motivation is malleable and develops over time (Camacho et al., 2021). On the other hand, it illustrates that writing motivation can rather be fostered in an indirect way, i. e., by improving the context of writing, than in a direct way, i. e., by influencing writing motivation itself.

Regarding the development of writing motivation, Boscolo and Gelati (2013) formulate two main assumptions. One states that 'attitude toward writing' and 'attitude toward self as a writer' decrease over the school years. These authors assume that this is due to the increasing complexity, and awareness of the difficulty, of the activity of writing. The other assumption is in line with Bruning and Horn (2000), stating that instructional practices may influence attitudes either positively or negatively. Pajares (2003), firstly, reports a decrease in writing self-efficacy beliefs in middle school age, which is consistent with findings from expectancy-value researchers and with patterns of other language art skills. Secondly, he concludes that confidence in writ-

ing skills is not, in fact, nurtured as students become better writers. Ekholm et al. (2018) support that writing attitudes decline naturally with age. By contrast, Klassen (2002) and Camacho et al. (2021) disagree because they have not found single trends for grade-level differences in self-efficacy for writing. Thus, some point to greater self-efficacy among younger students, while other studies indicate an advantage for older ones.

The literature offers some explanations for this inconsistent state of research. As mentioned above, the lack of explicit definitions of conceptualisations in writing motivation, for example 'attitude towards writing', might have a negative impact on assessment and comparison of scores (Ekholm et al., 2018). The deficient accuracy and validity of self-efficacy measures may complicate comparison of grade-level findings of different studies. In addition, there is a lack of research focusing on causes underlying grade-level differences, and a lack of longitudinal studies examining the development of different writing constructs (Camacho et al., 2021).

While the state of research on the first assumption of the development of writing motivation is unclear, there is more agreement on the second assumption that instructional practices may influence writing motivation. Bandura (1997) notes that there have been no successful interventions that have used attributions to affect motivation, but that socio-cognitive programs aimed at promoting personal empowerment have uniformly produced positive change. Persuasive attributions can, therefore, only serve as complementary motivators but not as the primary mode for generating human motivation and achievement.

S. Graham, Berninger, and Abbott (2012) assume that self-efficacy (Klassen, 2002), topic interest (Boscolo & Gelati, 2013) and attitude towards writing (S. Graham et al., 2007) each also play a role as contributor to, and as an outcome of, writing development. Bruning and Horn (2000) suggest that writing efficacy increases as a result of interventions that provide students with tools for improving their writing skills. Teaching strategies for writing stories and essays, for example, increase both writing skills and self-efficacy of students with learning disabilities (S. Graham & Harris, 1989). As for writing attitude, there is also some evidence suggesting that it is a malleable belief that is sensitive to various types of writing interventions including technology interventions, strategy interventions, and changes in the general approach to writing in the classroom (Boscolo & Gelati, 2013; Ekholm et al., 2018).

In summary, students' perceptions of the value of writing as an intellectual and social tool, and of its potential for learning and communication (Boscolo & Gelati, 2013; Bruning & Horn, 2000), as well as their beliefs and their attitudes, are all important for writing. Therefore, teachers need to rely on evidence-based practices in writing instruction and writing task planning (Chapter 2, Section 2.1.2), to develop students' writing motivation along with relevant skills, strategies and knowledge (Camacho et al., 2021; S. Graham & Harris, 2019a).

3.3.2.3 Writing motivation and achievement

Modern theories of writing (Hayes, 2012; Kellogg et al., 2013) stress that students' motivational beliefs have an important influence on how students respond to writing assignments and engage in the act of writing. Theoretically, students with stronger beliefs about self as a writer and about writing, and with a more positive attitude, are more likely to write better and longer stories, because they are, firstly, more likely to persist when they are writing and, secondly, likely to put more effort into the task (S. Graham et al., 2018; S. Graham et al., 2017). Empirically, previous research indicates that self-efficacy and attitude towards writing predict writing performance (e. g. Bruning et al., 2013; Ekholm et al., 2018; Klassen, 2002; Pajares, 2003; Saddler & Graham, 2007). However, Camacho et al. (2021) highlight three points about writing motivation: firstly, writing motivation is of a multidimensional nature; secondly, motivational variables shape and are shaped by writing competence; and thirdly, these authors assume that little is known about the role of motivational constructs other than self-efficacy and attitude toward writing.

The following discussion delves into some of the evidence on correlations between motivational variables and achievement, with a focus on middle school students and struggling writers. In line with the present thesis research, findings on digital tools are mentioned. The topic of gender appears frequently in the writing motivation literature. However, this thesis does not primarily deal with the relationship between writing motivation and gender. Nevertheless, gender might need to be considered as a control variable, which is why it is treated briefly.

Motivational variables and achievement in middle school students

Previous systematic reviews show that attitude and self-efficacy predict writing performance (Ekholm et al., 2018; Klassen, 2002; Pajares, 2003). The effect size of attitude on writing performance ranges from $d = 0.5$ to $d = 0.9$. Text quality, text length and longest correct word sequences are commonly used measures to determine writing performance (Ekholm et al., 2018). According to Pajares (2003), research conducted on college students shows consistently that beliefs about writing self-efficacy, and writing performance, are related. The effect size for writing self-efficacy in multiple regression models ranges from .32 to .42. Writing self-efficacy beliefs are further correlated with variables such as writing anxiety, grade goals, depth of processing, and expected outcomes. However, another consistent finding is that neither writing apprehension nor other motivation variables are predictive of writing performance in regression models that include self-efficacy. Self-efficacy, therefore, remains an independent predictor of writing performance when the influences of previous achievement, gender, grade-level and other motivational constructs are analysed (Pajares, 2003).

While S. Graham, Berninger, and Abbott (2012) report that attitudes positively predict writing performance measures, S. Graham et al. (2018) examine whether attitudes toward writing and self-efficacy predict writing performance of middle school students collectively, and they confirm this. These two writing beliefs, combined, account for a statistically significant 7 % of the variability in students' standard scores on writing performance, once variance due to other factors (gender, free/reduced lunch status, students' first language and disability status) are controlled for.

Motivational variables in students with dyslexia, DLD and struggling writers

A uniform statement about writing motivation in groups with impairments is hampered due to the different concepts and components investigated. Brouwer (2012) reports, from a study investigating intrinsic writing motivation in 8 to 10-year-old students with language impairment, that their level of reported writing motivation and perceived writing competence, i. e., by self-evaluative judgment, was lower compared to typically developing peers. While language impairment is not a significant predictor of writing motivation, Brouwer states that perceived writing competence is a significant predictor of writing motivation; in other words, that writing achievement affects writing motivation. This is inconsistent with S. Graham et al. (2007) who postulate that writing achievement does not affect writing attitude in this direction or is bidirectional, but that attitude only affects writing achievement in a unidirectional manner. This is consistent with Yeung (2022) who investigates differences in writing motivation between Chinese Grades 3–5 students with and without dyslexia. The perceived value of writing, i. e., a motivational variable, is found to predict writing productivity in the students with dyslexia. Yeung found no significant difference in writing motivation between the two groups in elementary students but suggests that the influence of writing motivation on writing achievement become stronger over the school years. The above-mentioned study by S. Graham et al. (2018) investigates the impact of writing attitudes and self-efficacy on the writing performance in students with special educational needs. These authors find, in a series of exploratory analyses, that writing attitudes and self-efficacy, collectively, account even more for statistically significant variability in the writing performance of students with disabilities, once variance related to the above-listed variables are controlled. Writing beliefs account for 16 % of unique variance in writing performance for students with disability, which is even more than for students without disabilities. However, the limited number of participants with disabilities ($N = 53$) in their study limits the validity of the results. Nevertheless, Camacho et al. (2021) acknowledge the importance of writing motivation for all students including those with special educational needs and with migrant backgrounds. S. Graham et al. (2018) assume that writing motivational beliefs might be especially important to the writing success of students with disabilities.

Motivational variables and digital devices

Since digital technology is becoming increasingly prevalent in the classroom, writing researchers have been interested in the influence of various instructional technologies and students' attitudes towards writing (Ekholm et al., 2018). Boscolo and Gelati (2013) argue that new possibilities that have arisen with the internet may contribute to making school writing even less attractive. By contrast, Ekholm et al. (2018) cite a majority of studies that support the notion that students' writing attitudes are positively associated with digital technology use. These authors consider this to be particularly true when technology conditions are compared to pen and paper conditions. Measures of attitudes toward recreational reading/writing and academic writing in digital contexts have predicted statistically and uniquely writing performance (S. Graham et al., 2018). One explication given suggests that digital platforms allow students more choice in writing and thus lead to more positive attitudes (Pruden et al., 2017). Another study suggests that digital technology interventions can make writing less onerous for students by providing support that reduces the cognitive burdens placed on them (see Chapter 2, Section 2.1.1 and Chapter 3, Section 3.3.1) (J. M. Sturm & Rankin-Erickson, 2002). Both these studies investigate students with learning disabilities. These findings and the S. Graham (2018) WWC-model (see Chapter 2, Section 2.1.1) provide further support for the idea that the relationship between writing motivational beliefs and student's writing is related to the writing context.

The implication of these findings for the present thesis research is that the context, such as STT, needs to be considered when examining writing motivation.

Gender and writing motivation

Gender differences are not in the focus of the present thesis. However, since numerous studies have found that females typically have more positive writing attitudes than males (Ekholm et al., 2018), gender must be considered. Gender-based differences in writing attitude appear to exist from the start, and persist throughout school (S. Graham, Berninger, & Abbott, 2012). Camacho et al. (2021) agree with, and add to, these findings on gender differences by pointing out advantages of girls over boys in other motivational constructs. However, they advise about interpreting these findings with caution, since the role of other variables has not been examined in most studies. In addition, a study by Pajares and Valiante (2001, p. 366) shows that "all gender differences favouring girls in writing motivation and achievement were rendered nonsignificant when feminine orientation beliefs were controlled".

In summary, the state of research on writing motivation bears numerous challenges for conclusions regarding students with dyslexia. However, this thesis assumes, firstly, that these students' potentially weak writing motivation with traditional techniques may affect their writing achievement. Secondly, the application of

a digital writing tool such as STT may increase these students' writing motivation. By contrast, it is unlikely that writing motivation increases due to a potential effect of STT on, for example, text quality, because of the likely unidirectional influence of writing motivation on writing achievement.

3.3.3 Situating STT for students with dyslexia within further theoretical perspectives

So far, the use of STT has been discussed from the learner's perspective, and it has been argued that its application is reasonable from cognitive and motivational perspectives. However, this thesis assumes that the question of the use of STT for students with dyslexia cannot only be considered from the learner's perspective. On the one hand, the utilisation of tools influences the writing system (Leu, 1997). On the other hand, the application of an AT in the classroom raises questions such as the feasibility or inclusion and exclusion. The application of an AT in the classroom most likely does affect teaching in some way. Therefore, the overall research questions presented in Chapter 1, Section 1.2 are considered from linguistic and societal perspectives (see Section 3.3.3.1), and the challenges of technology implementation are raised from the media educational perspective (see Section 3.3.3.2).

3.3.3.1 Rationale for text production with STT from the perspective of New Literacies Theory

Leu et al. (2013) consider that literacy has always been changing, depending on the tools for reading and writing, naming the first writing system as cuneiform clay tablets many centuries B. C. These authors assume that, recently, this change has been accelerated by digital tools. They call the way in which digitalisation continuously affects literacy 'deictically' because "the nature of literacy changes even faster than we can develop adequate theory" (2013, p. 4). There are now many terms trying to capture these newly emerged literacies in connection with changes due to digitalisation and technologies. For this reason, Leu et al. (2013) propose a collaborative approach by a theoretical framework that recognises simultaneously the numerous changes and multiplicity of perspectives. Such a theoretical framework that considers the ways in which new technologies serve as cultural tools to mediate literacy is particularly important for studies of digital writing (West, 2019). Leu et al. (2013) frame 'new literacies theory' as a theory containing two levels: lowercase 'new literacies theories' explore specific areas of new literacies and/or new technologies, whereas uppercase New Literacies Theory defines common findings and patterns from the lowercase theories. Although it is too early to define a complete uppercase New Literacies Theory, multiple studies in the field of ICTs and the Internet have so far generated the identification of eight central principles of such an uppercase

theory. Examples of these principles that appear common across research and theoretical work are:

- New literacies are multiple, multimodal, and multifaceted.
- New forms of strategic knowledge are required with new literacies.
- Teachers become more important though their role changes within new literacy classrooms (Leu et al., 2013).

Regarding writing, Leu (1997) assumes that these technologies alter various language modalities, especially writing processes and products. That fact that technological tools affect the writing process itself seems widely accepted, and after the presentation of the writing model of Hayes (2012) in Chapter 2, Section 2.1.1.2, not surprising (Applebee & Langer, 2009; Leu et al., 2013; Morphy & Graham, 2012). These changes in writing processes continue at a rapid pace with the development and the distribution of generative, pre-trained transformer architecture, such as ChatGP[32], driven by advancements in artificial intelligence (AI) and natural language processing technologies. These deep learning techniques are applied to understand and generate text which is based on given prompts (Mabuan, 2024). Applebee and Langer (2009) wrote, several years before this recent development, that new genres and forms of publications have emerged due the tools and technologies for creating written text.

To summarise, theories regarding literacies, especially concerning writing, have witnessed significant advancements with the emergence of digital technologies. Two conclusions are drawn from this. Firstly, these transformed literacy practices and writing processes demand the acquisition of adapted abilities to use these writing tools. STT is such a recently emerged writing tool. Secondly, the educational system, the task of which is to educate literate people, must adapt to these changes. Therefore, the role of educators is to give access and to adapt instruction of writing to this reality. Schools and educators play a crucial role in shaping instructional practices and in determining the effectiveness of educational interventions with these technologies. However, this task is challenging and needs close consideration. The fields of educational research and practice that deal with the question of enabling the next generation to use technologies and media in an educationally responsible way are media pedagogy and media education (Schäfer, 2001). The following discussion explores this challenge from the media education perspective.

3.3.3.2 STT in the writing classroom from the perspective of media education

Kerres (2018) states that any technology can only exploit its potential when the technology matches the intended learning goals and further parameters in the instruc-

32 ChatGPT is an advanced language model developed by OpenAI that has gained momentum since its introduction to the public in November 2022; see Mabuan (2024).

tional field. While the goal of writing was elaborated in Chapter 2, Section 2.1, the parameters and influencing factors in the instructional field are addressed as follows.

Impact of digitalisation on learning and teaching

The vision of digitalisation on learning impacts heavily the teaching and, it is assumed, also any technology-mediated intervention. The impact of digitalisation on learning has been the subject of controversial debate in the educational environment (Kerres, 2018). Kerres (2018) names three different notions about the role of technology on humans' actions which are important for the further discussion about media education and, therefore, discussed first.

There is, firstly, the notion that 'digitalisation is associated with an epochal change' to which education must adapt with the aim of optimising educational work with digital media. The use of digital technology is, in this view, part of the educational process, and students become competent handling digital information and tools through the active application of and reflection on their use. According to this notion, the educational offers adapt to socially widespread expectations and routines. The hope in technology lies in improving teaching and learning by better practice and expanded materials. Since instruction must be designed, this can involve major changes and effort for the teacher, which is rejected by some. Since the educational setting does not change profoundly, this view of digitalisation has an 'affirmative function' (Kerres, 2018).

A second notion, seen as 'digitalisation with disruptive effects', accords digitalisation more far-reaching effects on learning and the entire educational system. This effect of digital technologies emanates from the artefact itself: a cognitive confrontation or conscious adoption, or even decision, by the actors is not necessary. According to this line of argument, the presence of technology in the social environment affects people's behaviour, and the computers in the classroom change not only teaching and learning but also the role of the teacher (Kerres, 2018).

The third notion discussed by Kerres (2018), 'digitalisation as an option for action', acknowledges digitalisation as a design field and recognises that it depends on the players to bring changes into the learning environment. In this view, technology and digital transformation open up a wide range of options for teaching and learning, but Kerres (2018) states that the technology itself does not trigger the change in education. On the contrary, according to the author, a pedagogical goal can be achieved with various methods and is largely independent of the medium. Accordingly, the people responsible for the design of instruction can influence education with the help of digital technology because the technologies allow organising education differently and effectively implementing certain teaching and learning scenarios. The teachers' role will change in this setting because they will increasingly become orchestraters of learning contexts rather than dispensers of skills. Nevertheless, teachers must develop new instructional strategies and proficiency with new

technologies, which requires time spent on professional development (Leu et al., 2013).

The impact of digitalisation, according to this last notion, is therefore more complex than in the first two. Ultimately, whether certain expectations associated with media can be realised depends on the preparation and the design of their implementation in an educational context. At the heart of an intervention is therefore the potential a technology can have for a student through the student's engagement (Kerres, 2018).

In sum, it seems obvious for this thesis research that the assumption of reaching learning objectives by the simple application of the tool is too short-sighted. In other words, the use of a tool and the equipping of students with STT is not enough to improve their learning. It is therefore emphasised here that the tool must be implemented in the instructional design, and that it must match as closely as possible the educational situation of students with dyslexia, in order to permit a positive impact. In the case of STT as an AT in the writing process, the tool must match the instructional design of writing lessons and, possibly, further writing situations. However, the investigation of varying visions on digitalisation and learning and teaching among schools and educators are not within the scope of this thesis.

Nevertheless, these visions may impact the effectiveness of the use of the tool, because the support of study participants using an AT for writing may vary in terms of acceptance and adaptation of the instructional design. The fact that STT is an AT aimed at a specific group of students, rather than a technology for the whole class, might also be an impeding factor. The attempt to implement the technology into the instructional design might thus not receive sufficient attention. Educators with the above-discussed vision of digitalisation as affirmation – according to which a digital tool reinforces students with special needs, but adaptions in the instructional design present a high level of effort for educators in their conventional way of working – could reduce their investment of effort into an adapted instructional design. The lack of sufficient attention to the instructional design might also be relevant for educators with a vision of digitalisation as a disruption. Educators who assume that the digital technology impacts learning by itself might be less willing to consciously take an active role in planning and organising the instruction accordingly.

Framework for technology-mediated writing

Finding an instructional design that supports educators in their task to teach literacy, in the present thesis, writing, and to foster the meaningful use of technologies, are major and highly important challenges for stakeholders in the educational system, i. e., teachers and specialists. According to Usman et al. (2024), a practical model for delivering lessons with technological integration is the TPACK-framework by Mishra and Koehler (2006). The letters of TPACK stand for technological, pedagogical, and content knowledge. The TPACK-model supports understanding

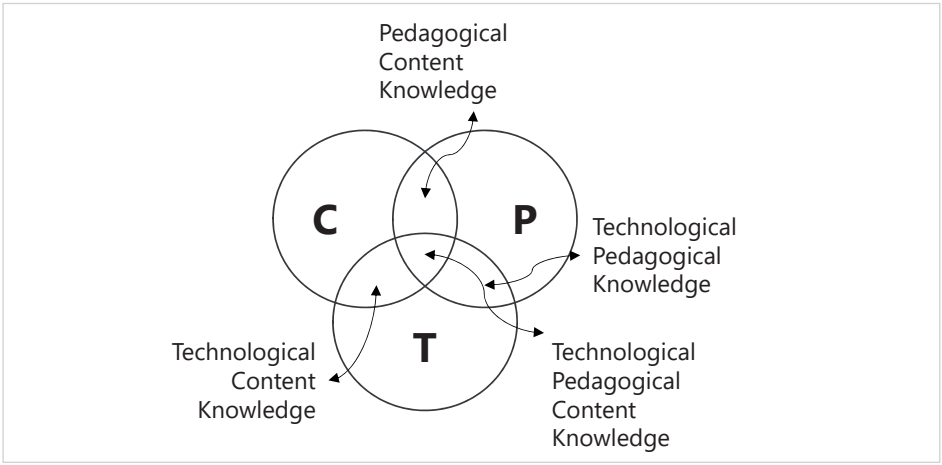


Figure 3.6: Technological Pedagogical Content Knowledge according to Mishra and Koehler (2006, p. 1025)

of the complex relationships between technology, content and pedagogy. Although Leu et al. (2013) call for further research about its efficacy, these authors assume that the TPACK-model facilitates teachers' growth in new literacies. Figure 3.6 illustrates that there are overlaps and interactions between the individual areas of knowledge (Usman et al., 2024).

Regarding writing, teachers must have content knowledge such as linguistic or text genres knowledge. The overlapping spheres of P and C refers to knowledge related to the learning process, such as teaching writing strategies or scaffolding techniques; whereas the overlapping area of technological and content knowledge refers to how various technologies can be used in learning to write. The inner part of the illustration refers to the knowledge teachers require to combine technology with learning and the writing areas (S. Graham, 2019; Usman et al., 2024).

This model illustrates the complexity of the tasks to bring technology into the writing classroom. This might be part of the reason why, according to S. Graham (2019), digital media have so far been underrepresented in writing lessons, although the majority of writing outside of school takes place digitally. An "indicator of the insufficiency of writing instruction in the typical classroom was the notable absence of the use of digital tools for writing" (S. Graham, 2019, p. 281). Another concern is that the writing needs of students with a disability are not sufficiently addressed (Dockrell et al., 2016). A further aggravating factor regarding students with disabilities, or low performing students, is that they receive less exposure to writing on the computer, both inside and outside of school (White & Robertson, 2015). This is problematic because, as Tate et al. (2016) found from the National Assessment of Educational Progress (NAEP) in the USA, regarding computer-based writing assessment, students' prior use of computers for school-related writing impacts writing achievement. School-related digital writing – in distinction to casual computer use

for more general writing – is found to have a greater impact on writing achievement because the tools must be integrated in a way that supports and extends the curriculum in meaningful ways.

3.4 Summary

In this chapter, it has been pointed out that digital technologies, independently, whether for all students or AT for specific students, are not efficient without adapted instruction (Kerres, 2018). By means of the TPACK model, it was then illustrated that technology-enhanced teaching is challenging and needs comprehensive teacher training and extensive knowledge (Mishra & Koehler, 2006). This is especially true for a challenging content such as writing instruction (Schröder et al., 2022). In Chapter 2, Section 2.2.3.2 it was emphasised that students with disabilities, and thus students with dyslexia, are more vulnerable to negative impacts of lesser prior computer use. It is therefore concluded that, for struggling writers, students with dyslexia included, it is even more relevant and necessary to receive both effective writing instruction and technology-mediated writing instruction and interventions (Regan et al., 2021). Boscolo and Gelati (2013, p. 286) write: "The internet has created a new literacy and provides people, particularly adolescents, with new possibilities of expressing themselves and communicating by writing". However, if the groups of students with special needs do not receive the necessary support to become competent in the use of media and digital writing, they are unable to exploit the potential of these new possibilities.

This is an example of the risk often referred to as the 'digital divide'. Digital divide(s) emerge through inequalities in access to and use of digital technologies for social, leisure and citizenship purposes (C. Abbott, 2007; I. Bosse et al., 2018). 'E-inclusion' is the term used by C. Abbott (2007) which stands within the context of debates around social inclusion and social justice. Social inclusion, a basic principle of human rights, is now understood in a much wider sense than when discussed initially in educational settings. While physical disabilities were prominent at first, there was then a shift to inclusion regardless of gender, race, age, gender, disability or class. E-inclusion refers to the use of digital technologies to break down all such barriers.

STT, as an assistive technology to improve both writing and knowledgeable use of digital tools, is at this point estimated as meaningful on a theoretical basis. Learning to apply these now widely used digital writing tools is important for the school and professional career. The introduction and supervised acquisition of such tools is especially important for students confronted with difficulties.

The development of STT has reached a state that allows application in different languages and for various groups of people. Some questions remain open, however, about its functionality for primary students with dyslexia and potentially further language difficulties or comorbidities. Nevertheless, according to CLT, motivational

theories and media theory, the compensation of the hampered lower-order writing processes offered by STT was argued to possibly be a suitable option to overcome barriers in learning due to dyslexia. Therefore, the following chapter elaborates on the state of research about STT for students with dyslexia.

4. State of the Art on Speech Recognition Technology for Learners with Dyslexia

Several arguments presented on a theoretical basis in Chapters 2 and 3 advocate for the application of STT for students with dyslexia: 1) STT is a powerful tool with the potential of further improvement and is nowadays easily accessible on various devices and/or online; 2) STT potentially frees cognitive resources at the level of lower-order writing processes, and the application of such a tool could thus have a positive impact on writing achievement and writing motivation; and 3) learning to use 21st century writing tools is supported for all students but especially for students with dyslexia. Nevertheless, several reasons could also be given against the use of STT for this group of students:

1) students with dyslexia potentially have secondary symptoms, e. g. avoidance behaviour, or comorbidities such as developmental language disorder (DLD), that could hamper the use of STT due to influent speech.
2) The presented writing models illustrate the complexity of the task and that many skills must be acquired and coordinated. Oral text production, which means by dictation, requires different processes and coordination which might, to some extent, bear challenges too. Chapter 2 also emphasised that writing fluency is mode-specific, and a new writing technology must thus be learnt and trained.
3) On the side of digital tools and learning in school, Chapters 3 highlighted that the successful implementation of a tool is demanding for educators and can be perceived as challenging.

However, the theoretical approach by itself does not allow a conclusion on the suitability of STT tools for students with dyslexia in the classroom. Therefore, this chapter delves into the state of literature on STT in the educational setting and specifically for the group of students with dyslexia.

4.1 Overview of the state of research regarding the focus user group

Research endeavours in the field of learning disabilities and speech recognition technology have been in progress since the first prominent study by Higgins and Raskind in 1995 which investigated the compensatory effectiveness of speech recognition on the written composition performance of postsecondary students with learning disabilities (Pennington et al., 2018). However, one conclusion of the scoping review by Berner and Alves (2023) is that research in this field greatly lacks empirical evidence. Their review of literature using STT by individuals with disabilities in multiple contexts shows that most of the studies are non-empirical or have lower levels of evidence. Matre and Cameron's (2024) conclusion from their scoping

review about STT for adolescents with learning difficulties points in a similar direction, stating that there is a great need for more robust research on STT use and especially on its effect on writing skills.

The implementation of STT in the classroom requires certain conditions to be beneficial (see Section 4.3). Regarding research in this specific field, two major challenges are added to those commonly present in experimental or quasi-experimental research studies in the educational field such as the recruitment of sufficient participants for statistical power, meeting the quality criteria, and obtaining authorisations. Firstly, STT has been presented as a rather recent technology, as discussed in Chapter 3, Section 3.2.1, and technological advancements have been and are still occurring. This implies that research findings in educational settings – which are time-consuming to obtain – are exposed to these rapid technological changes and thus rapidly outdated (Schüler, 2021), as already emphasised by Leu et al. (2013) and in the New Literacies theory. These rapid changes in STT complicate the identification of the most effective technology or the establishment of reliable outcome measures to conduct rigorous study designs (Berner & Alves, 2023). Secondly, according to Berner and Alves (2023, p. 1143), "An assistive technology must meet the complex needs of the individual, capitalise on the strengths of the individual, be compatible with multiple layers of context, and successfully support participation in the chosen activity or occupation". Studying AT in an empirical and standardised fashion limits, therefore, these requirements because it is difficult to account for variation in severity of disability and differences in desired outcomes and contexts.

In response to these calls for more solid research, Matre and Cameron (2024) acknowledge that, in their field on LD in secondary education, there is a small, yet important collection of studies on effects of the application of STT. However, in the specific field of students with dyslexia in primary school there are, indeed, only a few studies focussing on writing and writing-related skills. As a result of this lack, there is little knowledge about the influence of STT on texts written by students with dyslexia and about what the production process might look like (Mossige, Almgren Bäck, et al., 2023). Therefore, findings from neighbouring fields, presented in the following section, are considered in this thesis to elaborate potential benefits and challenges of STT in the educational setting, although these may be generalised for other groups only to a limited extent.

This thesis prioritises studies with primary and secondary students, while studies with post-secondary students and adults are considered only in specific cases. Schröder et al. (2022), who investigate digital technologies for writing but not STT, observe that most interventions with digital technologies for narrative writing are only implemented from Grade 5 onwards which might explain why only a limited number of studies on primary students is available. Since only one study on STT is available from a German-speaking country, attention cannot be paid to the language in which the study was conducted and the STT-technology employed. The transferability of research findings of symptom-based treatment approaches from

English-speaking countries to other languages is largely debated, e. g. the transferability to German with a more consistent orthography and less syllabic complexity (Galuschka et al., 2014). Since, in the present thesis the intervention is not a symptom-based treatment of spelling or writing accuracy, but higher-order processes are the focus, which are more comparable between related languages, the transferability is to some extent estimated as meaningful. Furthermore, some relevant research has been conducted in Swedish and Norwegian, thus strongly related to German. More problematic for the transferability of research findings between languages, but also over time, is the already mentioned fast development of the technology. On the one hand, STT in English, is presented as, to date, the most high performing technology outperforming lower-resourced languages such as German (Silber-Varod et al., 2021). On the other hand, research findings in English from, for example, two decades ago are themselves subject to limitations due to the enormous subsequent technological advances, which therefore reduces their transferability (Svensson et al., 2019).

Three research articles reviewing the state of the art in neighbouring fields are considered for the present state of research in this thesis. According to the scoping review by Berner and Alves (2021), there is no strong evidence supporting STT for individuals with various disabilities, although they conclude that STT may improve participation and independence. Matre and Cameron (2024) formulate the results of their own scoping review with the same caution, namely that STT may increase certain abilities and provide help with some processes. Meanwhile, Perelmutter et al. (2017) summarise the results of their evidence-based systematic review and meta-analysis regarding STT as having a small positive effect, despite some mixed results. These latter authors investigate AT interventions for adolescents and adults with LD. A first deduction is thus drawn from the reviews, that the state of research is ambivalent, meaning that there are not only advantages but also obstacles and significant challenges in the application of such technology in the educational setting.

Of the greatest interest for this thesis research are Scandinavian studies because they investigate digital technologies as AT for students with dyslexia. (Almgren Bäck et al., 2023; Almgren Bäck et al., 2024; Mossige, Arendal, et al., 2023; Svensson et al., 2019). However, the scope of the tools examined in the most comprehensive study, by Svensson et al. (2019), is broader than only on STT, and allows only a few statements on writing. Further literature considered for this thesis review investigates STT for: students with LD (Higgins & Raskind, 1995; Lee, 2011; MacArthur, 2009; MacArthur & Cavalier, 2004); individuals with disabilities in multiple contexts such as physical disabilities (Berner & Alves, 2023); struggling writers (Schüler, 2020); English language learners, meaning learners of a second language (L2) (Arcon et al., 2017); regular school settings (Haug & Klein, 2018); or students with traumatic brain injuries (Noakes et al., 2019). A second deduction is thus drawn from this general overview: to date there is insufficient research to assert that STT simplifies the entire writing process for students with dyslexia. Research findings must thus be sought

in neighbouring fields and assembled according to distinct writing processes and further components, for the purposes of the present, thesis research.

4.2 Benefits and challenges of STT for students with dyslexia in school

Examples of observed benefits of STT in the educational setting are presented in the following discussion, separately according to the writing and writing related skills, e. g., text quality, writing fluency, reading fluency, or writing motivation. Assumptions about challenges are elaborated, with the aim of making informed decisions for the subsequent deduction of the thesis research procedure. Conditions for the successful implementation of the assistive technology are then elaborated in Section 4.3 according to the larger context of the educational setting and the individual needs of the student.

4.2.1 Research findings on STT condition and higher-order writing skills

Students with dyslexia struggle in getting text on paper, because of difficulties in spelling, capitalisation, punctuation and/or handwriting which interfere with the higher-order processes of composing (MacArthur & Cavalier, 2004). Therefore, interventions or adaptations that potentially remove barriers at the level of spelling and handwriting or typing would, theoretically, enhance the higher-order writing processes and, thus, writing skills. Therefore, benefits and challenges for higher-order writing skills are presented. However, it should be noted that the difficulty of assessing higher-order writing skills, presented in Chapter 2, Section 2.1.3, is further amplified here because of the lack of assessment tools for the dictating process and dictated text products (Svensson et al., 2019).

4.2.1.1 Text quality

MacArthur and Cavalier (2004) and Arcon et al. (2017) both compare three conditions to each other: handwriting; dictating to a human scribe; and dictating with STT. Findings of these studies indicate that the best results are achieved with a human scribe, which is not, however, within the scope of this thesis. Both studies investigate text quality via a holistic scoring. The study by MacArthur and Cavalier (2004) – comparing high school students with and without LD – results in significantly higher text quality when dictated with STT than when written by hand for students with LD; with t (20) 2.36, and $p = .028$, $d = .42$. By contrast, the study by Arcon et al. (2017) investigates the use of STT for 9- to 14-year-old English language learners, with these authors reporting significantly higher text quality in the dictating than in

the handwriting condition.[33] This finding is not confirmed by the results of Quinlan (2004) who compares handwriting with STT in narrative writing of 11- to 14-year-old, fluent and less fluent writers, finding a significant interaction of group x mode of STT for text length (see below) but not text quality.

In contrast to these studies on students with LD and less fluent writers, Svensson et al. (2019) investigate TTS and STT for students with severe reading and writing difficulties, excluding comorbidities such as DLD and autism as well as students who had not attended school in Sweden from Grade 1. These authors specify that none of the assessment methods objectively capture writing skills but that the students and parents testified by self-reporting that, thanks to STT, the students with writing difficulties were as skilful at producing text as their typically developing peers.

4.2.1.2 Vocabulary diversity, text length, and word length as predictor variables of text quality

Almgren Bäck et al. (2024) seek to fill the gap of objective effects on writing skills in research, focusing also on students with dyslexia. However, these authors do not provide a holistic or analytic score for text quality. Instead, they endeavour to capture text quality by variables of text complexity commonly assumed as predictors for text quality, i. e., vocabulary diversity, word length (WL), and text length. Almgren Bäck et al. (2024) find, in their multiple-baseline single-case study, that seven out of eight students, aged 9 to 12 years, improve their proportion of diverse vocabulary, calculated by a non-overlap of all pairs (NAP)[34]. The NAP = .62-1.00, showing medium to large effect sizes. MacArthur and Cavalier (2004) as well as Arcon et al. (2017), also investigate vocabulary diversity in students with LD and conclude that this variable does not explain quality differences among the conditions. Furthermore, Almgren Bäck et al. (2024) and some other researchers investigate productivity, meaning 'text length' as a predictor of text quality. Almgren Bäck et al. (2024) state that STT improves the productivity of most students in their study, with NAP = .70–1.00, but that the magnitude of the effects varies greatly with median level changes among students, 15–218 words. Quinlan (2004) also finds an increase in text length. The repeated measures ANCOVA shows a mode x group interaction for text length, i. e., STT or handwriting, with F (1, 38) = 4.75, $p < .05$. This indicates that the STT writing mode allows less fluent writers to produce longer texts, which is not the case for fluent writers. By contrast, studies by Kraft et al. (2019) and MacArthur

33 Arcon et al. (2017) indicate the ANOVA results, with $F(2, 34) = 18.99$, and $p < .001$, $\eta_p^2 = .53$, but do not provide the post hoc results comparing handwriting to STT.

34 NAP is an index of data overlap between phases in single-case studies, and quantifies the extent to which measurements for each phase are the same; see Parker and Vannest (2009). The effect size is defined as weak (0.51–0.65), medium (0.66–0.92), or large (0.93–1.0); see Almgren Bäck et al. (2024).

and Cavalier (2004), which both also score for length as a predictor for text quality in total number of words, find no difference between conditions. The only study available in German by Schüler (2021) investigates STT in Year 8 without focus on a specific group. This author compares students' text production by handwriting, typing and dictating conditions. The study findings regarding text length, assessed in the number of syllables, show a significant effect between these three modes, with F (2, 90) = 13.5, $p < .001$. Post hoc analyses result in significant differences between handwriting and STT, with $t = -5.09$, $p < .001$, and $d = .73$, and between typing and STT, with $t = -2.36$, $p = .02$, and $d = .37$. It is important to note that this author did not determine a significant difference between handwriting and typing, with $t = -1.9$, and $p = .06$.

An additional variable investigated by Almgren Bäck et al. (2024) in students with dyslexia is word length, i. e., words with seven letters and more, as an indicator of text complexity. These authors find that the proportion of words with seven and more letters increased in all eight students, with NAP = .65-1.00. This finding supports prior research suggesting that STT can encourage the usage of complex and diverse vocabulary. One explanation for this could be that STT encourages the writer to dictate relevant and semantically accurate words to the text that would be deselected in handwriting or typing because of the spelling complexity (Kraft et al., 2019). Another plausible explanation is that writers, in the case of these studies of secondary and post-secondary students, on purpose use orthographically more complex words because they realise that STT recognises those better than less complex words (Higgins & Raskind, 1995; Schüler, 2020).

As presented above, Quinlan (2004) finds a different effect in text length between middle school students with and without difficulties in writing fluency. It is thus assumed that STT does not affect text length in the same manner among different groups of students. This is in line with MacArthur and Cavalier (2004) who also find a positive effect for struggling writers but none for fluent writers. The effect of STT is, thereafter, not equal for all subgroups. However, more research is needed regarding effects on different subgroups of students (Quinlan, 2004; Schüler, 2020).

In summary, the research findings presented above on STT, respecting vocabulary diversity, text length, and word length, indicate: 1) positive effects or no effects, with negative effects either not found or not published; and 2) differences between groups observed, confirming Svensson et al.'s (2019) assumption that assistive technologies are especially beneficial for students with severe reading and writing difficulties.

4.2.1.3 Writing strategies and revision

An important element of the higher-order writing processes is writing strategies (Philipp, 2015). Haug and Klein (2018) investigate the acquisition of writing strategies in students in Grade 5 from regular classes under two conditions: writing by

hand or with STT. Their results show large, statistically significant gains in text quality, word count, and variety of rhetorical argument moves over time, but no interaction between time and treatment. Thus, in respect to strategy acquisition, STT is not more effective compared to the handwriting condition. However, these authors find a significant transfer to the untrained modality. Two conclusions are drawn from this. Firstly, while transcription skills were discussed as mode-specific by Hayes and Olinghouse (2015), the results of Haug and Klein (2018) suggest that STT provides an alternative medium for learning strategies, because learning a strategy by the means of STT is possible. Furthermore, higher-order writing processes are not or not exclusively mode-specific. Secondly, this result emphasises the positive effect of specific writing instruction such as instruction and training on strategies.

Quinlan (2004) investigates text revision – assigned to the higher-order processes by A. Sturm et al. (2017) – and observes that text quality decreases when students apply more revisions. This author suggests that revision is a specific challenge in using STT.

4.2.2 Research findings on STT condition and lower-order writing performance

The higher-order writing skills were presented above as the overall goal of writing. However, the lower-order writing skills, as a means to the end of producing legible and understandable text, are also of great importance. The following sections present a discussion of research findings on the impacts of STT on spelling, investigated as surface errors or word errors, and writing fluency, assessed as value of accuracy and speed.

MacArthur and Cavalier (2004) score for 'word errors', including homonyms, grammatical errors, start of sentence capitalisation, and end of sentence punctuation. In the STT condition, errors differ from the handwriting condition, since the technology recognises real words only (Chapter 3, Section 3.2.1). Errors relevant to STT include word substitutions, insertions, omissions, and incorrect endings. Errors typically in handwriting include spelling and illegible handwriting. However, errors on a word level may occur due to inaccurate revisions. In their study, MacArthur and Cavalier (2004) find a statistically significant interaction effect of disability status x condition, with $F\,(1, 27) = 6.11$, and $p = .020$, $\eta^2 = .184$. LD students made significantly less errors in the STT-condition than in handwriting, with $t\,(20) = 3.94$, $p = .001$, and $d = .83$, but this was not the case for typically developing peers. Quinlan (2004) reports the same finding for younger students with and without fluency difficulties. Again, a significant interaction is found regarding 'surface errors', i. e., spelling and punctuation, of mode x group, with $F\,(1, 38) = 17.38$, and $p < .01$, indicating the difference that the STT-mode has on the groups of fluent and less fluent writers. Post hoc analyses comparing the two modes within less fluent writers show, in addition,

that the less fluent writers make significantly fewer surface errors by dictating than by handwriting, with t (19) = 4.55, and $p < .0125$.

Although the only available study in German by Schüler (2021) does not focus on students with dyslexia, the descriptive analyses of the comparisons of handwriting, typing and dictating are of interest for the present thesis. The author investigates writing fluency, defined as writing performance within a set time by different writing accuracy variables. The dictating modality outperforms the other modalities in terms of: 1) the greatest number of completely correct transcriptions on a syllable level; 2) dictated texts being more accurate than typed, the latter still being more accurate than handwritten texts; 3) dictated texts containing the highest percentage of correct word sequences (%CWS); and 4) the highest minimal percentage of correct word sequences. The investigated variables, accuracy on a syllable level and the %CWS, both show statistically significant effects: accuracy, of F (2 90) = 112.09, and $p < .001$; and %CWS, of F (2, 90) = 31.82, and $p < .001$. Post hoc analyses show significant differences between all three transcription modes.

The only available study specifically investigating word-level accuracy in STT among students with dyslexia, by Almgren Bäck et al. (2024), finds positive results for word-level accuracy. Seven students of the multiple-baseline single-case study improved their proportion of word-level accuracy, with NAP = .69-1.00, while one remained, with NAP = .51, at the same level. While word-level inaccuracies were nearly non-existent during the dictating process, some inaccuracies occurred while revising. Regarding the correct word sequences, four students in their study improved, with NAP = .82-1.00, three students remained stable, and one student declined, with NAP = .03.

In sum, the state of research on lower-order writing skills points in the direction of an advantage of STT over handwriting and typing, but the state of research is still scarce. On the one hand, studies have worked with limited numbers of participants; on the other hand, few studies have investigated specific groups such as students with dyslexia with a uniform definition of their disorder (Matre & Cameron, 2024; Svensson et al., 2019). Furthermore, the question raised by Svensson et al. (2019), about the effect of STT on learning to read and write in all modalities, remains unanswered because the above- presented studies investigated differences between handwriting and dictating, or effects over time within dictated texts. Except for Svensson et al. (2019) who investigated the transfer of learning with tools to various modalities but did not capture writing, the studies reviewed here have not investigated the remedial aspects of STT tools. However, the latter authors have investigated further aspects of writing such as writing motivation, as addressed in the following section.

4.2.3 Research findings on STT condition and writing-related skills and demands

4.2.3.1 Reading fluency

Interventions can advantageously combine writing and reading (Shanahan, 2019) because they are based on the same basic processes (Linnemann et al., 2022) and because writing involves re-reading when editing and revising the text (Hayes & Olinghouse, 2015). In addition, some research studies about STT, e. g., Almgren Bäck et al. (2024) and Svensson et al. (2019), also involve the TTS function, in other words using a tool that synthesises written text to oral language. These authors suggest that it may be beneficial for students with dyslexia to listen to their produced text using TTS alongside STT: firstly, because their struggle to decode could be a barrier to revising and improving the text (Almgren Bäck et al., 2024); and secondly, because research on TTS by itself shows positive results on reading skills of students with dyslexia (Keelor et al., 2023; Wood et al., 2018). However, to date, no clear evidence is available on the effect of STT on decoding or reading comprehension. According to Almgren Bäck et al. (2024), the students' choice of when to use TTS or to read unassisted varies. Nevertheless, these authors positively estimate that the students, to some degree, used TTS autonomously during the intervention.

4.2.3.2 Writing motivation, general school motivation

Svensson et al. (2019) are interested in both STT and TTS but also in general school motivation and motivation to write. Although they do not capture objective changes in writing skills, they do obtain information about motivation, through questions to participants and parents. According to these authors, students and parents generally agree on the increase in general school motivation and writing motivation due to using TTS and STT apps. Interestingly, the students with the most severe written language difficulties answered more positively questions about these technologies, such as, "Has your motivation for schoolwork increased after working with the reading apps?" (Svensson et al., 2019, p. 8). Matre and Cameron (2024) also consider STT to be a tool possibly promoting engagement in writing, but they assume that little is known about students' motivation derived from STT use as transferred to writing in other modes.

4.2.4 Participation, social inclusion, and well-being

Already in 2004, MacArthur and Cavalier describe the advantage of STT in terms of the independence that a student may gain by the technology, compared to dictation to a human scribe, which had outperformed STT at that time. Research findings about STT in multiple disability contexts show that STT and TTS my foster independence and participation in school activities (Berner & Alves, 2021; Liu et al.,

2019). Regarding students with dyslexia, Almgren Bäck et al. (2024) observe students' operational independence and their ability to use STT, which the students consistently maintained during the STT intervention in a single setting and in follow-up assessment. However, little is known about those students' long-term experiences of AT use in the school setting and their actual independence, social inclusion, and well-being with STT in school (Almgren Bäck et al., 2023). Almgren Bäck et al. (2023) describe, by means of an interview study, mixed user experiences and challenges encountered when using STT in the school setting. These authors depict three major themes that they estimate as essential to rendering AT interventions sustainable, which are developed in the following discussion: 'emotional responses in the learning environment'; 'contextual factors'; and 'developing meaningful strategies'.

'Emotional responses in the learning environment' refers to students' thoughts, feelings and behaviour when using STT or avoiding its application. Individuals express mixed feelings when using STT (Almgren Bäck et al., 2023). Almgren Bäck et al. (2023) suggest that these feelings stem either from (non)acceptance of dyslexia and/or AT, or from unwanted attention. Some adolescents who used STT in Svensson et al.'s (2019) intervention study, prior to the interviews by Almgren Bäck et al. (2023), experienced emotional difficulties such as embarrassment in the classroom, anxiety, or frustration with the technology. These students' feelings are linked to real or perceived stigmatisation in relation to dyslexia and/or AT. One aspect of unwanted attention from peers could be a discomfort because the tool is visible, and the speech is audible, to those around. As a result, STT use exhibits a difficulty that is usually, unlike a physical disability or severe cognitive impairment, less obvious (Dawson et al., 2019). Another such aspect is seen in being uncomfortable leaving the classroom to use STT, or even in feeling a discomfort about being prioritised by teachers in occupying a separate room that peers would also like to use. By contrast, other individuals describe using STT as being destigmatising because it allows them to participate in the classroom assignments and because STT is considered as mainstream technology accessible to everyone (Almgren Bäck et al., 2023). These positive feelings are in line with the assumption by Dawson et al. (2019) suggesting that students with dyslexia may adopt most those technologies that blur the boundaries between everyday technologies, such as STT, and assistive technologies. In sum, these findings and observations indicate that context is relevant for technology integration and the well-being of students.

According to Almgren Bäck et al. (2023), 'contextual factors' in the school and classroom can function as facilitators or barriers. The authors, conducting research in Sweden, suggest that these factors do not concern access to digital devices in the school environment but that organisational shortcomings can hinder the continued use of AT. Their interview study reveals three subthemes: '(un)aware access and transitions'; 'need for environmental adaptations'; and 'continuous or fragmentary support'. Accordingly, there are several organisational prerequisites, further discussed in Section 4.3, as conditions for successful STT use, that might undermine

schools' investments in digital devices, AT, and good intentions. To summarise, the use of STT requires support from teachers and specialists to promote students' independence with the tools and hence, to improve social inclusion and well-being.

In addition to the requirements within the context, students who use STT for writing need support in developing new writing strategies when using a tool, as further elaborated in Section 4.3.2 addressing the introduction of the tool in the present study. The constraints imposed by writing differ from those imposed by speaking (Olive, 2014). However, dictating also differs from speaking: although dictating is an oral form of text production, meaning that the medium is spoken language, dictating a narrative text, as an example, is based on a written concept (Koch & Oesterreicher, 1985; Oesterreicher, 2008). This shift from the written to the spoken modality is not the only challenge that the tool imposes on the user: researchers describe a certain number of further issues that are challenging in dictation and, it is assumed, especially for students with dyslexia.

4.2.5 Recognition issues, technical issues, and time consumption

Automated speech recognition technology, and therefore STT-tools, has become much more efficient in recent years. However, not only the meta-analysis by Matre and Cameron (2024) but also several recent studies (Almgren Bäck et al., 2024; Matre, 2022; Mossige, Arendal, et al., 2023) describe word recognition errors that hamper the use of STT. Matre (2022), who investigates STT for all students, with the idea of an inclusive approach, reports accuracy issues when dictating among typically developing lower secondary students. The problem of a technology that does not capture the pronunciation of students with dyslexia is, according to Almgren Bäck et al. (2023), Almgren Bäck et al. (2024), and Svensson et al. (2019), one of the main reasons for students to abandon the STT-tool and potentially leads to frustration and decreased productivity. Mossige, Arendal, et al. (2023) agree that this is a major challenge for a writer, and that this can lead to a high effort in revising and correcting mistakes (Matre & Cameron, 2024). Recalling the differences in word accuracy rates depending on the age and possible accent of the user (Chapter 3, Section 3.2.2), accuracy issues might therefore be especially important for younger students with dyslexia who, in addition, potentially have comorbidities such as DLD. The time needed for text production with STT in addition to costly revision is also an issue highlighted by Matre and Cameron (2024) because STT requires the student to be trained with the system. Furthermore, technical challenges such as connection or login problems are among the most frequent criticism reported by students and teachers.

In summary, the literature on the benefits and challenges of STT for students with dyslexia reveals a mixed picture of its effectiveness. Although numerous studies have investigated its effectiveness, there is currently no specific evidence about contra-indications, aside from potential discomfort with the tool among students

and the commitment and time it requires. However, the above elaborated challenges suggest that the dictating process itself and the use of STT within the school context bear challenges that must be considered despite the benefits of STT. Almgren Bäck et al. (2023, pp. 6–7) express this conclusion as follows: "The accumulated research suggests that it is essential to address these drawbacks to make AT interventions sustainable".

Therefore, the next section addresses the conditions for the successful use of STT.

4.3 Conditions for successful STT use

Perelmutter et al. (2017) and Svensson et al. (2019) claim that AT, and especially STT, must be customised to the individual in order to be effective. Not only the individual fit of a tool but also the adjustment to the contexts in school and teaching are important. The research on an implementation guide for AT by Bausch et al. (2015) illustrates the complexity involved in, and how many steps that must be considered before, implementing AT. The following discussion addresses the complexity of using a tool such as STT in school. To this end, this discussion provides an overview of conditions for the successful use of STT. Conditions are organised as: contextual factors on the user (Section 4.3.1); the importance of a thorough introduction and support (Section 4.3.2); and the combination of the STT application with effective writing instruction (Section 4.3.3). This thesis discusses these conditions with an understanding of schools as entities including teachers and specialists. Meanwhile, neither addressing the responsibilities within the school, nor taking a position on the professional disciplines responsible for AT interventions[35], are within the scope of the thesis. Regarding the attribution of STT, this thesis is in line with Dawson et al. (2019) who state that AT is not a replacement for effective intervention strategies in reading and writing. On the contrary, a comprehensive program must include both intervention strategies and AT implementation.

4.3.1 Contextual factors

The previous sections emphasised that STT can be both beneficial and challenging at the same time. Although access to the tool or a device is one prerequisite, this is only one component within the entire context that plays an important role in the technology integration in the classroom. According to Petko (2012, p. 1352):

> Regardless of the intensity of well-intentioned efforts at other levels (such as improvement of hardware, software and content, the adaptation of curriculum procedures and academic strategy papers, or the providing of continuing education, coaching, sup-

35 Professional disciplines named by Bausch et al. (2015) are AT specialists, occupational therapists, speech/language pathologists, physical therapists and special educators, administrators, and general educators.

> port and educational leadership), all remedial measures will ultimately prove futile unless teachers have a fundamentally positive attitude toward the potential benefits of ICT.

This statement emphasises the central role of teachers and/or specialists[36], as is illustrated in the 'will, skill, tool model' (Knezek & Christensen, 2016). Four variables – the will, the skill(s), the tool(s) and pedagogical practice – are at the core of technology integration (Petko, 2012). Knezek and Christensen (2016) and Petko (2012) show, in their research on the 'will, skill, tool model' enhanced by pedagogy, that these important contributors to the integration of technology explain up to 90 % of variance in the level of ICT integration in the classroom. In the following, their contributions are outlined and supplemented by examples of research findings about the application of STT by students with dyslexia.

The area named 'tool', i. e., the availability, accessibility and extent of technology use include not only hardware and software but also professional development for teachers (Knezek & Christensen, 2016) and specialists (Bausch et al., 2015). Regarding STT, Matre and Cameron (2024) state, in their scoping review on STT for secondary students with LD, that the acceptability and usability of STT is encouraging but that, at the same time, it is clear that students need sufficient preparation, time and ongoing support to use these technologies successfully. Thus, the teachers and specialists might need, first and foremost, knowledge about the tool and the specificity of its impact on writing, to provide the support needed, but they also need knowledge to obtain the resources for a tool if not yet available (Bausch et al., 2015).

'Skill' refers to the ability to use and experience the technology. Self-perceived confidence, i. e., self-efficacy, and readiness to use the technology, also belong to this skill variable (Knezek & Christensen, 2016). Regarding the use of STT technology, Almgren Bäck et al. (2023) name organisational prerequisites that may contribute to the students' mastery experiences. Such organisational prerequisites include the capability, based on experience with tools and with their use, to anticipate difficulties. These authors refer, on the one hand, to more general knowledge such as about transitions between devices, e. g. from a tablet to a laptop[37], which requires consideration from teachers and specialists. However, on the other hand, these professionals also need specific knowledge of the tool, for example knowing how the tool functions in the auditory environment, to make meaningful adaptations.

A positive attitude towards the use of technology in instruction is named 'will'. Matre and Cameron (2024) conclude that acceptability and perceived usability of the tool they studied is generally high among teachers, parents and students. This is

36 This model focuses on the teacher. Specialists are also mentioned in this paper, as they are also considered important in the present study and the field of research in Switzerland.

37 In Sweden, frequently occurring at the transition from primary school to secondary school; see Almgren Bäck et al. (2023).

in line with Nordström et al. (2019) who state that the general attitude among teachers and specialists in Sweden towards AT such as STT is positive, since they estimate such tools as being a great help for students. However, as discussed in Chapter 3, Section 3.3.3, the vision of digitalisation varies among professionals, which impacts their expectations and engagement (Kerres, 2018).

'Pedagogical practice' in this model is defined as "any conscious activity by one person designed to enhance learning in another" (Knezek & Christensen, 2016, p. 312). This definition, which is in line with the TPACK model presented in Chapter 3, Section 3.3.3, includes also the level of confidence that teachers have in their use of instructional strategies for technologies (Mishra & Koehler, 2006). Almgren Bäck et al. (2023) write that lesson organisation is an important aspect for the students in using a tool. Since the setup of the dictation tool or installation in the provided workspace may take more time than for traditional writing tools, it is important to consider when, where and for how much time STT can be used within a lesson. In addition, teachers and specialists also need knowledge about the specificities of text production with STT to provide systematic training (Perelmutter et al., 2017).

However, MacArthur (2009) and Arcon et al. (2017) also take the role of the student into account. They emphasise the importance of motivation and will on the student's side, as well as their consequent use of STT. At this point, the importance of the role of the student is not questioned; and it seems obvious that, on the one hand, avoidance strategies must be prevented, while on the other hand, the importance of the user's participation in decision-making around AT, as well as the requirement of mutual agreement, are therefore emphasised as essential to maintain and develop individual AT strategies in the classroom. One example of this is the agreement between the teacher and the student about when and where to use STT (Almgren Bäck et al., 2023).

4.3.2 Software introduction and support

The role of the teacher and/or specialist has been emphasised as substantial for the successful implementation of the technology. One reason for this is that learners in general, and students with dyslexia in particular, must learn to use the writing technologies (MacArthur, 2009). Although the students today grow up surrounded by digital technologies, they do not necessarily have the appropriate strategies to handle a technology, or they do not know how to employ appropriate strategies, when they are left to themselves to manage their learning environment (Kirschner & van Merriënboer, 2013). This is why the students benefit when teachers and/or specialists understand how AT may mediate their writing and reading challenges, when the professionals acquire basic technical knowledge themselves, and when they know how to support students using AT adequately (Dawson et al., 2019). Regarding such support for the use of STT, research findings suggest that systematic support throughout school is necessary (Almgren Bäck et al., 2023). Ok et al. (2022) name,

in this regard: 1) environmental support; 2) device support; and 3) instructional support. These authors understand, by environmental support, the provision of a quiet place, a comfortable environment for speaking out loud, and a stable internet connection. The device support includes the appropriate hardware and software that provides a high speech recognition accuracy. The instructional support contains the opportunity to practice verbal skills and learning to edit, but it also provides scaffolds such as checklists and prompts to support the writing and editing process. The latter process is emphasised by several authors. MacArthur and Cavalier describe (2004) that STT technology imposes additional cognitive burdens that must be addressed, because "the transcription part of the text production process radically changes from encoding and typing to speaking" (Mossige, Almgren Bäck, et al., 2023, p. 100). MacArthur (2009, p. 102) states that the main question is not whether the writing technologies work, "but how to design effective instruction using the technology". In the MacArthur and Cavalier (2004) study, a research assistant therefore explained the system, demonstrated how to dictate, and guided the student through the initial training on the system. However, the users must learn numerous peculiarities of dictation. They must articulate carefully and avoid extraneous vocalisations. At the same time, they must dictate punctuation and format the text, such as starting a new paragraph. The dictating condition changes also the revision process. While the writer should recognise spelling errors in handwritten texts, the writer must, for example, learn to recognise homophones such as 'by' and 'buy' or hallucination errors where STT output does not correspond to the intended input (MacArthur, 2009). Mossige, Almgren Bäck, et al. (2023, p. 107) address this challenge by a structured working process for text production that is "based on years of experience at the Competence Center for Reading in Aarhus". This foresees three main phases, each with detailed steps: the preparation of the hardware and software; the dictation procedure; and the revision procedure.

In detail, the authors advise on these steps as follows (Mossige, Almgren Bäck, et al., 2023, p. 108):

1. Preparation: a. Prepare the computer and the task. b. Open an empty document and save it. c. Check the microphone and sound.
2. Dictation: a. Say the sentence aloud without recording it (or think the sentence). b. Record the sentence with speech-to-text. c. Say or type a full stop.
3. Revision: a. Listen to the sentence with text-to-speech. b. Assess the sentence. c. Make necessary changes.

A further difference between the handwriting and dictating conditions might lie in the size of segments treated at once. The model by Olive (2014) illustrates the parallel and cascading processing of segments by hand (Chapter 2, Section 2.1.1.3). Regarding dictation, the software might impose a different procedure because the software interprets a word input based on surrounding words. The accuracy therefore improves when using sentences as recognition units (MacArthur & Cavalier,

2004). The step 'a' of the dictation phase in Mossige, Almgren Bäck, et al. (2023) just mentioned is supposed to counter this feature of the software, through adaptation of the size of the segment. These authors assume that the oral rehearsal can improve sentence planning, structuring and pronunciation (Myhill & Jones, 2009), particularly beneficial for students with dyslexia and additional speech disorders (Almgren Bäck et al., 2024).

In summary, introduction of STT confirms the high demands placed on teachers and specialists already emphasised in the presentation of the two models, the TPACK-model, and the will, skill, tool model also including pedagogy, in Chapter 3, 3.3.3 and Section 4.3.1. However, this requires not only knowledge but also time. According to Matre and Cameron (2024), findings show that the success of STT introduction in schools depends, among other factors, also on the school and staff resources. The high level of resources necessary might, to some extent, explain why the National Assistive Technology Research Institute in the US found that, although schools tend to consider AT during meetings of the 'individualized education program' (IEP), it is less common to follow through on all the recommended steps: assessment, selection, implementation, and evaluation. To increase success of AT, one key point is thus providing instruction to in-service and pre-service teachers and specialists (Bausch et al., 2015).

4.3.3 Combination with effective writing instruction

Firstly, the compensation of lower-order writing processes by the introduction and application of STT does not replace intervention (Dawson et al., 2019). Secondly, the application of the tool must be combined with writing instruction (MacArthur & Cavalier, 2004). Research findings support the general principle that technological tools by themselves have little impact on learning, as elaborated in Chapter 3, Section 3.3.3 about media in education. Learning depends, instead, on a combination of the technology and the instruction, which is especially true for struggling learners. Therefore, instruction should be designed to help students benefit from the capabilities of the technology (MacArthur, 2009). The research findings by MacArthur and Cavalier (2004, pp. 45–46) illustrate this mutual influence: "Neither instruction in a planning strategy nor dictation alone had a statistically significant effect on the quality of their persuasive essays, but the combination resulted in superior essays". These authors assume planning as particularly important when writers compose via dictation; but also, further evidence-based strategies such as self-regulation strategies, including goal setting, organising, self-monitoring and self-evaluation are mentioned in STT research (Yeung, 2022). At this point, it is estimated that the content and features for evidence-based writing instruction elaborated in Chapter 2, Section 2.1.2 also apply to text production via dictation. Besides this, as mentioned in in the same chapter, S. Graham and Harris (2020) assume, on the one hand, that not only students in regular classrooms but also students with LD receive little

writing instruction and spend very little time writing. On the other hand, although teachers report using evidence-based writing practices, these procedures are applied infrequently. In respect to teaching, MacArthur (2009) recognises the high demand on professionals to provide both sufficient time on task to train in writing with STT, as well as to offer the necessary instruction to support students to take advantage of the STT technology and to provide specific and evidence-based writing instruction. Therefore, he pursues the "idea and reasoning that teachers would need support to design writing instruction …" (MacArthur, 2009, p. 94). In respect to research, Matre and Cameron (2024) conclude, from their scoping review on writing with STT in secondary students with LD, that research on STT requires consideration of the willingness of teachers and students to adopt STT. Furthermore, research should investigate the application of STT in the educational context, including how any potential benefits may occur and under what circumstances this happens.

In summary, to date there is a lack of high-quality research due, on the one hand, to rapid technological changes and, on the other hand, to the difficulty of investigating the topic of STT with randomised controlled trials within subgroups. As a result, questions such as, 'Who can benefit from STT?', 'In what contexts is the application of STT successful?', and 'At what age is the application of STT reasonable?', are, to date, not yet, or only partially, answered. The present thesis focuses on the specific group of students with dyslexia – although this group involves also a certain diversity – and asks about the challenges of implementation in the writing classroom.

5. Intervention Study – Methodological Approach

Being able to write is essential for academic and professional careers and everyday life, as argued in Chapter 2, Section 2.1. The strands of theory and the current state of research discussed in Chapter 2 to 4 have revealed that STT has the potential to be a useful tool for students with dyslexia to participate in society through writing. Nevertheless, there are also prerequisites to be met and challenges to consider. This thesis seeks answers to the question of the effectiveness and compatibility of this technology for students with dyslexia. Chapter 5 presents the methodological approach of the thesis research, while Chapter 6 presents its findings.

After two preliminary notes on the methodological approach in Section 5.1, Section 5.2 derives the research questions (RQs) from the theory and literature part and formulates the corresponding hypotheses. The overall research design, with its two strands, the quantitative in Section 5.3.3 and the qualitative in Section 5.3.5, each are followed by their respective sample description. Section 5.4 presents the selection of appropriate assessment instruments and the preliminary work to fit them to the study requirements. Section 5.5 presents the study equipment, i. e., hardware and software, and the introduction and teaching materials developed for the present, quasi-experimental study. An overview of the survey instrument is presented in Section 5.6.

5.1 Preliminary notes on the present methodological approach

5.1.1 Three study groups and their functions

Due to the conditions elaborated in Chapter 4, Section 4.3.3 for the use of STT, the present methodological approach contains a particularity which must be highlighted for a common understanding of the following RQs. Because writing instruction is an important condition for the benefits of STT (MacArthur, 2009), the present intervention study provides writing instruction by instructional videos. Since the provision of writing instruction goes beyond the application of a tool and might influence the results, it is considered as a second intervention and must be controlled, which is done by an additional group. Therefore, this study comprises three groups:

- Experimental group 1 (EG 1) uses the STT tool, receives writing instruction by videos, and treatment as usual (TAU)
- Experimental group 2 (EG 2) receives writing instruction by videos, and TAU
- Control group (CG) receives conventional writing instruction by the teacher, and TAU – 'conventional' as opposed to the specific writing instruction based on instructional videos provided in this study.

5.1.2 Investigating the effect of a writing tool by writing assessment

One major challenge of this study is the aim to determine whether the intervention of a writing tool is effective on writing outcomes. Writing outcomes, however, are influenced by the writing tool. This means that a standardised, minimally biased procedure of comparison must be found. Comparing dictated and handwritten texts between groups is not considered, as this would not allow the research to distinguish between differences due to the intervention and the modality. Furthermore, in EG 1 it would not be possible to distinguish how much of the change between T1 and T2 is caused by using the tool and how much is caused by changes in writing competence. Comparing all groups in both modalities, handwriting and dictating, is not possible because only EG 1 is trained in the dictating modality. Therefore, three deductions are made as follows.

1) To examine changes in writing and writing related skills in a standardised way, the students must write with the same tool. Handwriting is the modality known to all participants. This procedure can be supported by the research finding that writing strategies can be transferred from one modality to another (Haug & Klein, 2018). It has also the advantage that no specific assessment method for dictated texts needs to be developed for this comparison, as there are none, according to Svensson et al. (2019).
2) However, in contrast to the transfer of writing strategies, the development of writing fluency is mode-specific, meaning that it cannot easily be transferred from one modality to another (Hayes & Olinghouse, 2015). The EG 1 might therefore be disadvantaged regarding writing fluency, because the students practice dictating and write less by hand during the intervention period. This needs to be considered when testing hypotheses.
3) Thus, a separate research question addresses differences between fluency in the two modalities, i. e., handwriting fluency and dictating fluency, within EG 1 only.

5.2 Research questions and hypotheses

5.2.1 Research question 1

5.2.1.1 Effects of STT on text quality – RQ 1A

The main part of the study investigates the overarching question of whether the application of speech recognition technology is an effective method to improve writing among the group of students with dyslexia. The idea of STT for students with dyslexia it that students need to write to pursue their academic and professional career, but the ability to write texts is hindered for this particular group due to their difficulty with lower-order writing processes. In Chapter 2, Section 2.1, producing text was argued to be the ultimate goal of writing. Therefore, RQ 1A addresses text production with text quality as an overall indicator of text production skills. So far,

alterations in text quality due to the application of STT have not yet been examined for the group of students with dyslexia, and most studies have compared text quality output produced in different modalities rather than the effect of STT on text production in other modalities.

***1A** Does text quality increase significantly more strongly in students of EG 1 who used speech recognition technology between pre- and post-test, than in comparable students of EG 2 and CG?*

From a theoretical point of view, STT is assumed to free resources in the working memory which can then be used for higher-order processes. Therefore, text quality is expected to increase in the EG 1. If writing motivation increases due to STT use (see RQ 4 below), writing motivation could have an additional positive impact on text quantity, which in turn correlates strongly with text quality (Quinlan, 2004). The current state of research indicates that STT has positive effects on text quality and text length when compared with the handwriting modality. Despite concerns about its use in German and numerous challenges, a one-sided hypothesis is formulated:

$H1_{1A}$: The text quality of students' texts of EG 1 increases more strongly between pre- and post-test than the text quality of comparable students' texts of EG 2 and CG.
$\rightarrow H1_{1A} : \mu_{EG\ 1\ text\ quality} > \mu_{EG\ 2\ \&\ CG\ text\ quality}$

5.2.1.2 Effects of provided writing instruction by videos on text quality – RQ 1B

For the reasons mentioned in Chapter 4, Section 4.2, this study combines two types of intervention. Question 1A investigates the effect of STT, which is the main focus of the study. Meanwhile, the writing instruction by videos – with writing understood as composing, not transcription – provided for EG 1 and EG 2, might have an effect on text production too. It is recalled that the provided writing instruction has the following two goals: 1) writing classes for students with and without STT should be comparable; and 2) a certain amount of specific writing instruction and evidence-based writing instruction needs to be assured to accomplish one main prerequisite. The aim of this part of the investigation and RQ 1B is, therefore, to distinguish between the effects of STT and the effects of the provided writing instruction.

***1B** Does text quality increase significantly more strongly between pre- and post-test in students of EG 1 & EG 2 who receive the provided writing instruction, than in comparable students of CG who do not receive the same instruction?*

According to the current state of research, in recent years, a considerable share of students has not achieved basic skills for their age group in writing. One assumption is that time on task for writing is currently too low to develop this complex competence properly (Applebee & Langer, 2009). Another assumption is that teach-

ers do not dispose of specific knowledge to instruct on writing strategies, which is why mostly lower-order writing skills are taught (S. Graham, 2019). In the German-speaking part of Switzerland, the knowledge about the quantity and the content of writing instruction is considered limited. Therefore, consistently providing one hour of pre-prepared writing instruction for 18 weeks could be more in terms of amount, and more specific, than the usual writing instruction. The related hypothesis, therefore formulated in a directed way, is as follows:

$H1_{1B}$: The text quality of students' texts of EG 1 and EG 2 increases more strongly between pre- and post-test than the text quality of comparable students' texts of CG.
$\rightarrow H1_{1B} : \mu_{EG\ 1\ \&\ EG\ 2\ text\ quality} > \mu_{CG\ text\ quality}$

This study compares writing outcomes of the CG that does not receive writing instruction by videos to both experimental groups that receive writing instruction by instructional videos (see Section 5.5.3). The results will allow this study to differentiate between effects of STT and effects of writing instruction, but it will not allow any statement about the content of the writing instruction in the CG, since no data is collected about the writing instruction that the CG receives during the intervention.

5.2.1.3 Effects of STT on text length – RQ 1C

Text quality and the amount of text are strongly correlated (Quinlan, 2004). This is why the question of the development of text length is of interest and, in this thesis, associated to the text quality.

***1C** Does text length increase significantly more strongly in students of EG 1 who use speech recognition technology between pre- and post-test, than in comparable students of EG 2 and CG?*

On the one hand, the students who dictate their texts in the writing lessons during the intervention period write less by hand. Therefore, they have less training time in handwriting (see RQ 2 below). Since writing fluency is, according to Hayes and Olinghouse (2015), mode-specific, the same amount of training in dictating might not affect handwriting fluency. On the other hand, the use of a tool with potentially freed resources in working memory and an assumed increased writing motivation could affect the amount of text students write. Since the state of research is, due to the two opposite rationales, estimated to be not sufficiently solid to formulate a directed hypothesis (Bortz & Schuster, 2010), the following hypothesis about text length is formulated in a non-directional way:

$H1_{1C}$: The text length of students' texts of EG 1 develops differently between pre- and post-test than the text length of comparable students' texts of EG 2 and CG.
$\rightarrow H1_{1C} : \mu_{EG\ 1\ text\ length} \neq \mu_{EG\ 2\ \&\ CG\ text\ length}$

5.2.1.4 Effects of provided writing instruction by videos on text length – RQ 1D

Again, effects of STT and the provided writing instruction by videos must be distinguished.

***1D** Does text length increase significantly more strongly between pre- and post-test in students of EG 1 & EG 2 who receive the provided writing instruction than in comparable students of CG who do not receive the same instruction?*

While little is known about the amount and the specificity of writing instruction in German-speaking Switzerland, it was argued in RQ 1B that the provided instruction by videos for EG 1 and EG 2 contains specific writing instruction and fosters the provision of one hour per week. Therefore, the following hypothesis is formulated in a directed way:

$H1_{1D}$: The text length of students' texts of EG 1 and EG 2 increases more strongly between pre- and post-test than the text length of comparable students' texts of CG.
$\rightarrow H1_{1D} : \mu_{EG\ 1\ \&\ EG\ 2\ text\ length} > \mu_{CG\ text\ length}$

For reasons of comparability, all students write their texts at pre- and post-test by hand. Since the EG 1 students write less by hand during writing lessons, this could be seen as a disadvantage for this group. However, as the EG 1 still uses handwriting in other lessons and takes notes during writing lessons, this disadvantage is not estimated as problematic and thus accepted as a more conservative way of testing H1.

5.2.2 Research question 2

5.2.2.1 Effects of STT on handwriting fluency – RQ 2A

This section addresses the research questions' effects of STT on further writing and writing-related skills. Writing fluency is understood as the fluency in handwritten texts, unless otherwise stated as such in dictated texts, addressed in Section 5.2.3 Concerning writing fluency, defined by the quantity (e. g. number of words written) and correctness of the text (e. g. spelling), there are reasons to assume both positive and negative effects. On the one hand, one can argue that STT bears the risk that students no longer progress in their writing fluency skills, since these mode-specific (Hayes & Olinghouse, 2015) lower-order processes are taken over by the software. For example, the motor process of handwriting is substituted by oral text production. However, such a substituted process may be necessary for the acquisition of correct spelling (M.-L. Bosse et al., 2014). As a result, writing fluency could be negatively affected by STT. On the other hand, the use of STT might have a positive impact on spelling, because STT produces correctly spelled words – which does not guarantee the accurate recognition of the input and does not rule out the possibility

of students intervening with the keyboard. Thus, students with dyslexia, tending per definition to make spelling mistakes, are more likely to see correctly spelled words on the screen when dictating, than when writing by hand or typing. The repeated exposure to correct spellings, compared to seeing various incorrect spellings, could have a positive effect. Therefore, the RQs 2A and 2B ask whether the application of STT for students with dyslexia impacts writing fluency skills.

***2A** Does writing fluency differ significantly in students in EG 1 who used STT between pre- and post-test in comparison to students in EG 2 and the CG?*

As stated above, there are reasons to assume either that EG 1 does or does not improve writing fluency skills. Therefore, the following hypothesis regarding writing and reading fluency is formulated in a non-directional way:

$H1_{2A}$: Writing fluency skills develop differently between pre- and post-test in students of EG 1 and in students of EG 2 and CG.
→ $H1_{2A} : \mu_{EG\ 1\ writing\ fluency\ skills} \neq \mu_{EG\ 2\ \&\ CG\ writing\ fluency\ skills}$

Equally to the first research question, EG 1 might be disadvantaged in writing fluency in the handwriting assessment condition, due to less time being spent on handwriting during the intervention period.

5.2.2.2 Effects of STT on reading fluency – RQ 2B

A further effect is conceivable on reading skills, which Matre and Cameron (2024) describe as 'writing related skills'. Writing and reading are correlated and share common characteristics in acquisition and in instruction. Linnemann et al. (2022) write that they are based on comparable cognitive skills and common knowledge resources. In addition, the basic reading and writing processes must be automated to free resources for higher-order processes. The authors summarise literature assuming that promoting one, i. e., reading or writing, might have a positive effect on the other. Therefore, this thesis study asks whether text production with STT and revising on a computer affect reading fluency. The use of STT could, for example, impact reading due to the time spent on reading (Taylor et al., 1990) while revising. While STT is treating text, the user might spontaneously read what is being written. In addition, text written on the screen is more readable than handwritten text. Thus, it is conceivable that students are more likely to re-read text on the screen than on paper. Another example of STT impacting reading could lie in the revision process in this thesis study. Students of EG 1 are held to revise their dictated text by the TTS function of MS Word, called "Read Aloud" (see Section 5.5.2).

***2B** Does reading fluency increase significantly more strongly in students in EG 1 who used STT between pre- and post-test than in comparison to students in EG 2 and the CG?*

For the reasons discussed above, students of EG 1 might improve more strongly in their reading fluency than the EG 2 and the CG. Consequently, the following hypothesis to this question is formulated in a directed way:

$H1_{2B}$: Basic reading skills show a greater increase in students of EG 1 between pre- and post-test than in students of EG 2 and CG.
→ $H1_{2B}$: $\mu_{EG\ 1\ reading\ fluency} > \mu_{EG\ 2\ \&\ CG\ reading\ fluency}$

The writing instruction by videos provided to EG 1 and EG 2 focuses on higher-order writing skills, e. g. planning or structuring text. This instruction does not target or primarily foster writing and reading fluency. Therefore, controlling for effects of this instruction on reading and writing fluency does not appear necessary.

5.2.3 Research question 3

5.2.3.1 Dictating fluency versus handwriting fluency – RQ 3A

The investigation of STT implying a specific writing modality, and comparing groups using different modalities, harbours the challenge of distinguishing the effect of the tool from the impact of the modality. Therefore, RQs 1 and 2 focus the effect of the intervention on the only comparable modality which is handwriting. By contrast, RQ 3 investigates the effect of the intervention and the modality. To this end, RQ 3 investigates differences between texts produced in the two modalities, i. e., handwriting and dictating with STT (3A), and the effect of the intervention on dictated texts, i. e., the development over time (3B).

***3A** Are dictated texts from students with dyslexia significantly better in terms of fluency measures than their handwritten texts produced under comparable conditions?*

The speech recognition software takes over the transcription of spoken text. Chapter 4, Section 4.1 showed that this immediately leads to an increase in the amount of writing for adults. As discussed in the literature review, for children with dyslexia, it is unclear whether the tool leads to an immediate increase in the amount of text and the correctness. Two assumptions are made based on theory and the current state of research. Firstly, it is hypothesised that more text can be dictated during the same time period than written by hand, as the speaking speed is higher than the handwriting speed. Secondly, it is hypothesised that the correctness in dictated texts is higher than in handwritten texts, as the technology is developed in a way that transcribes speech sounds into existing words. This leads to two, one-sided alternative hypotheses as follows:

$H1_{3A1}$: Texts dictated by EG 1 at pre-test (T1) are longer than handwritten texts at pre-test.
→ $H1_{3A1}$: $\mu_{length\ dictated\ texts\ T1} > \mu_{length\ handwritten\ text\ T1}$

$H1_{3A2}$: Text dictated by EG 1 at pre-test contain significantly more correct spelling than texts written by hand at pre-test.
→ $H1_{3A2}$: $\mu_{\text{correctness dictated texts T1}} > \mu_{\text{correctness handwritten text T1}}$

5.2.3.2 Development of dictating fluency – RQ 3B

Question 3B, which examines changes over time in the dictating modality, compares dictated texts from EG 1 students at pre- and post-test.

***3B** Does dictating fluency in the group of students with dyslexia increase significantly between pre-test and post-test?*

As shown in literature, the application of a tool needs to be learned. EG 1 composes texts during writing classes with the assistance of STT. It is therefore assumed that a training effect of the use of STT impacts length and correctness of dictated texts, in the following hypotheses:

$H1_{3BI}$: Texts dictated by EG 1 at post-test (T2) are significantly longer than texts dictated at pre-test (T1).
→ $H1_{3BI}$: $\mu_{\text{length dictated texts T2}} > \mu_{\text{length dictated texts T1}}$

$H1_{3B2}$: Texts dictated by EG 1 at post-test contain significantly more correct spelling than texts dictated at pre-test.
→ $H1_{3B2}$: $\mu_{\text{correctness dictated texts T2}} > \mu_{\text{correctness dictated texts T1}}$

The answers to both RQs 3A and B should not be confounded with the question of the effectiveness of STT on the modality-independent acquisition of written language. The present answers may, however, contribute to the understanding of possibilities and limitations of STT to assist students with dyslexia in their modality-specific text production.

5.2.4 Research question 4

5.2.4.1 Effects of STT on writing motivation –RQ 4A

Chapter 3, Section 3.3.2 highlighted the importance of writing motivation as a predictor for writing achievement, especially for struggling writers (e. g. S. Graham et al., 2018). While some authors believe that writing motivation decreases over time, others think that the application of digital tools, such as STT, and the writing environment may impact writing motivation in a positive way (e. g. Ekholm et al., 2018). Therefore, writing motivation and possible alterations due to the STT-intervention will be investigated by RQs 4A. RQ 4A addresses the writing motivation.

***4A** Does the writing motivation of students in EG 1 show a significantly greater increase between pre- and post-test than the writing motivation of students in EG 2 and CG?*

Theory and literature allow the assumption that writing motivation increases when an impaired function is circumvented by a digital tool, which is why the present thesis investigates the following, one-sided alternative hypothesis:

$H1_{4A}$: The overall writing motivation increases more strongly between pre- and post-test in students in EG 1 than in students in EG 2 and CG.
→ $H1_{4A} : \mu_{EG\ 1\ writing\ motivation} > \mu_{EG\ 2\ \&\ CG\ writing\ motivation}$

5.2.4.2 Effects of provided writing instruction on writing motivation – RQ 4B

Chapter 3, Section 3.3.2 discussed that writing instruction and writing motivation correlate. At least some aspects of writing motivation appear to be malleable and sensitive to various types of writing interventions (Ekholm et al., 2018). It is conceivable that the writing instruction provided by videos also influences writing motivation more so than the STT condition. RQ 4B investigates alterations due to the provided writing instruction on the writing motivation.

***4B** Does the writing motivation of students in EG 1 & EG 2 show a significantly greater increase between pre- and post-test than the writing motivation of students in CG?*

Chapter 3, Section 3.3.2 argues that not enough is known about effects of writing intervention. For an unambiguous theoretical foundation such as this one, a two-sided hypothesis, as follows, must be formulated (Bortz & Schuster, 2010):

$H1_{4B}$: The writing motivation develops differently between pre- and post-test in students in EG 1 and EG 2 than in students CG.
→ $H1_{4B}: \mu_{EG\ 1\ \&\ EG\ 2\ overall\ writing\ motivation} \neq \mu_{CG\ overall\ writing\ motivation}$

5.2.5 Research question 5

The last research question addresses possible challenges that the implementation of an audible tool, i. e., STT, harbours. The literature review emphasised the possible challenges and high demands on professionals. Knowing these challenges is essential, firstly, to encounter them in practice and, secondly, to deepen the understanding of this thesis research findings, leading to RQ 5.

***5** Which challenges accompany the implementation of STT in the classroom?*

While RQs 1 to 4 are investigated by quantitative methods of inferential statistics, RQ 5 is explored through a qualitative approach. Based on theory and research lit-

erature (Chapter 4, Section 4.2), some assumptions about challenges can be made. These assumptions influence the exploratory approach to this question, which is discussed in Section 5.3.5 However, these assumptions are not tested in the form of statistical hypotheses.

The study design requires the development or adaptation and procurement of several study materials. Section 5.4, and the elements in green and in blue in Figure 5.6, present all assessment tools necessary for the present research, while Section 5.5 shows the developed instructional materials for the software introduction and writing instruction, and the acquisition of hardware and software.

5.3 Research design

5.3.1 Overall research design – quantitatively-driven, mixed methods design

Chapter 4, Section 4.1 stated that robust research in the domain of STT in the educational setting and its effect on the development of writing is much needed. This section addresses this need and elaborates the appropriate study design to investigate STT for students with dyslexia in writing classes.

The RQs and hypotheses presented in detail in Section 5.2 comprise two types of RQs: four about the effectiveness of a method; and one about difficulties met during the study and potentially met in the implementation of STT in the classroom. It would be difficult or even impossible to answer both types of questions by operating solely within either the deductive-objective-general approach of quantitative research methods or by an inductive-subjective-contextual approach that characterises qualitative research methods (Morgan, 2014). RQs 1–4 are based on theories and they test hypotheses. They require the research to evaluate observations which rely on the collection and analyses of data in an explicit and predetermined design. RQ 5 aims to capture and to discover patterns or themes, beginning with observations and potentially leading to the creation of theory or the generation of new hypotheses through flexible, emergent research designs (Morgan, 2014). The present study is linked to both rationales, of quantitative and qualitative research methods. Bringing both approaches together is the general goal of mixed methods research.

Glogowska (2011), naming mixed methods as a 'third way' of pragmatism, concludes that mixed methods have already contributed to research in the field of speech and language therapy (SLT). Glogowska estimates that mixed methods offer a wealth of possibilities for SLT because of the multifaceted and multifactorial nature of human communication and its disorders. Nevertheless, besides the strengths of mixed methods, Glogowska also names challenges of mixed methods. One problem is that there are only a few established and widely accepted procedures within the two traditional strands (Morgan, 2014). Glogowska (2011) states that mixed methods bear considerable challenges at each stage, from conceptualising over operationalis-

ing, to defining how the methods are used in conjunction with each other and how rigour is ensured. A further issue of bringing qualitative and quantitative methods together is, according to Moran-Ellis et al. (2006), how to integrate the results of such different data and of different formats. These authors also state that there are different reasons and varying ways to incorporate mixed methods into research designs and advise caution regarding the terms used.

To meet these challenges, this thesis follows the suggestion of Morgan (2014) who argues for a compact typology following a purpose-driven perspective. This requires an assignment of different levels of priority to each study component within a project that uses different methods. The next subsections allocate, firstly, the priority of the components. The resulting overall research design, as illustrated in Figure 5.1. Section 5.3.2, thereafter, determines when and how the data are integrated. It also assigns the weight as core and supplement components.

Priority level

The qualitative part of this present thesis research is largely planned according to the state of research and is not based on the quantitative research results[38]. Therefore, this mixed methods design is not a sequential contribution but a 'parallel studies design for additional coverage', as Morgan (2014, p. 36) names it, symbolized as "QUANT + qual". Morse and Niehaus (2009) call the equivalent design, 'quantitatively-driven mixed method designs', and specify the moment of data integration. As shown in Figure 5.1 below, the present research collects the qual data simultaneously, analyses them separately, and integrates the results with the QUANT findings.

Based on Morgan (2014), the quantitative method in this research project is designated as the 'core component' (symbolised as QUANT). This part of the study could stand on its own and answers the majority of the RQs. The qualitative part (symbolised as qual) in the present study might not stand alone due to the depth of the analysis and the criterion of saturation that is not sought (see Section 5.3.5). Therefore, the interview study is the 'supplementary component' within the project.

According to Morse and Niehaus (2009, p. 119) "the theoretical drive for quantitatively-driven mixed methods design is deductive, and the overall thinking about the project is deductive, regardless of the minor supplemental qualitative components". The qualitative component must be appropriate in terms of its implementation, and interface with the quantitative core; but within the qualitative component, the research follows the inductive paradigm for a qualitative study.

However, regarding this thesis project, it is considered that the two components, conducted in parallel by one researcher, may influence one another. The strategy of

38 Possible influences of the intervention study prior and during the intervention period on the qualitative part will be discussed in Section 5.2.5.

the data integration and how the two components could possibly interact with each other is outlined in Section 5.3.2.

5.3.2 Integration strategy of quantitative and qualitative data

The Figure 5.1 indicates by arrows the strategy for the data integration as well as possible interactions between the QUANT and the qual components. Interactions and integration are presented in the following paragraph from the top to the bottom, i. e., from number one to four, which corresponds to the timeline of the project.



The arrow nr. 1 illustrates the preparation of the study, including the development and evaluation of the study material. The first version of the interview guide is tested with pilot teachers and subsequently adapted. Feedback and collected QUANT data from the pilot studies might have influenced the interview guide.



The arrow nr. 2 stands for any possible impact of the QUANT component during the intervention period. The researcher was available for questions and the contact person for any issues during the introduction of the software and the 18-week intervention period. The researcher was made aware of problems met with the tool and the application of writing instruction by videos. This awareness is assumed to have had an impact on how challenges were addressed during the interviews.



The arrow nr. 3 stands for the first conscious integration of the data. The conduction of the interviews and the first cycle of qualitative data analyses were run before the data analysis of the QUANT data. This first analysis in MAXQDA was consciously used for the selection of the characteristics of the students integrated in the QUANT data analysis later (see Chapter 6, Section 6.1.1).



The last arrow, nr. 4, illustrates the actual goal of the integration methods, which is to foster understanding of the QUANT data and to extend the depth and breadth of the implementation of the technology in school. Therefore, the QUANT results and the qual results are presented separately, but the qual results are integrated into the discussion in Chapter 7.

5.3.3 Quantitative core – quasi-experimental study

This section elaborates the research design and the elements of the quantitative core component, including the rationale for the experimental study and the necessary transformation to a quasi-experimental study. This section also illustrates the timeline and the design of this quantitative component and explains the aim of each

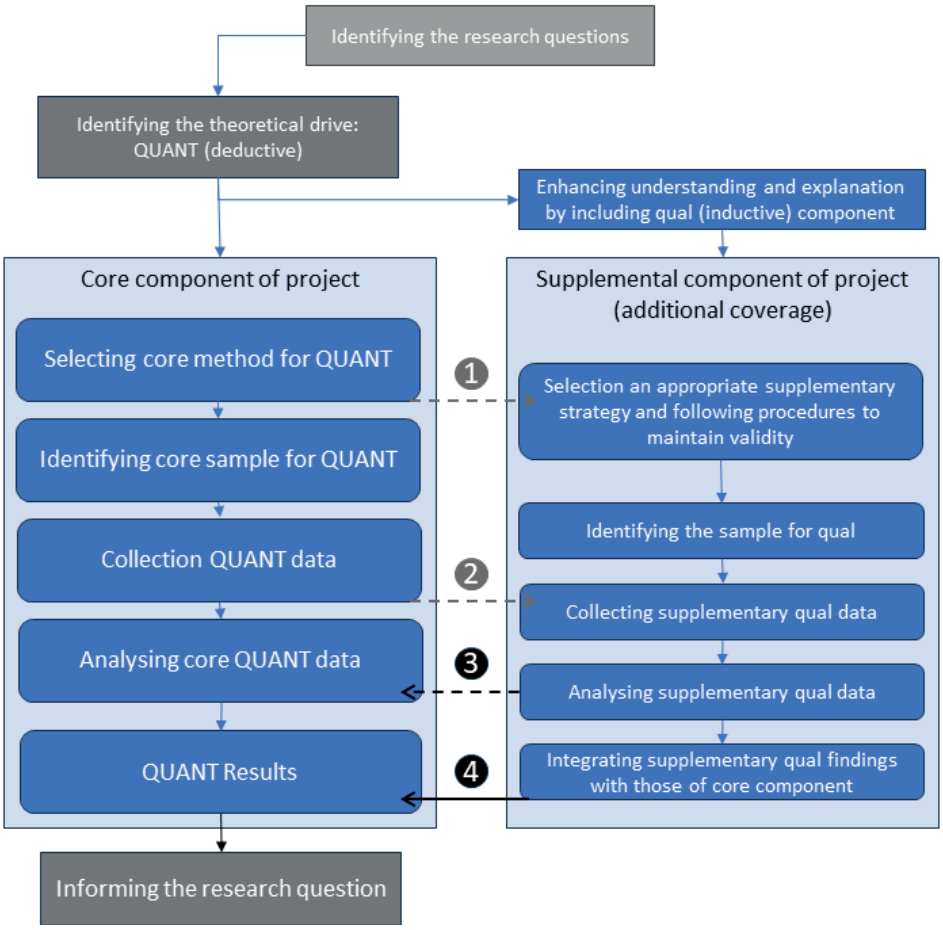


Figure 5.1: Illustration of design according to Morse and Niehaus (2009, p.118)

study group. Section 5.3.3.1 develops the foreseen statistical model for the data analysis, i. e., the general linear model (GLM), and names an alternative for data not allowing the calculation of a GLM.

5.3.3.1 Quantitative design

RQs 1 to 4 are based on theoretically derived hypotheses. These hypotheses are tested and answered in a deductive process. By the means of experiments, it is possible to experimentally manipulate causes that might affect outcome variables. Experimental research in the social sciences usually takes the form of program interventions (Morgan, 2014). The S3-guideline for diagnosis and treatment of students with dyslexia in German (Deutsche Gesellschaft für Kinder und Jugendpsychiatrie, Psychosomatik und Psychotherapie e.V [DGKJP], 2015) (now outdated) calls for high-quality intervention studies with randomised group assignment (RCT). Matre

and Cameron (2024, p. 1) state, "there is a great need for more robust research on the use of speech-to-text technology (STT) in educational settings, especially on its effect on writing skills". Bhide et al. (2018, p. 380) write, "the randomized controlled trial is the most rigorous and robust research method of determining whether a cause-effect relation exists between an intervention and an outcome". In the case of the present study, the main manipulation, or the so-called treatment condition, is the modality in which students produce text – dictating or handwriting. The outcome variables, several writing skills and writing motivation, which were specified in the research questions are used to investigate the effects of the intervention.

On the one hand, the current state of research, presented in Chapter 4, Section 4.2.2, has revealed that effective writing instruction is a condition for the successful application of an STT-tool. Thus, it is derived from this state of research that writing instruction should be included in the study design. On the other hand, the increased workload and the work intensification of teachers do not allow unconditionally for additional time investment in research participation (Creagh et al., 2023). As further elaborated in Section 5.5.2.4, video instructions are applied to reduce workload in terms of preparation and to enhance the comparability of writing instruction for all classes. However, the development of the instructional materials in this study has led to the growing presumption that the writing instruction by videos should be considered as an intervention on its own. This line of thought was reinforced by the later feedback from the pilot-class teachers. For this reason, a third study group, called 'control group' (CG), was installed to control for any effect of this writing instruction. This study comprises, therefore, three groups, described below, and illustrated in Figure 5.2 in black, red and blue.

Study groups and provided interventions

The CG, black in Figure 5.2, receives no intervention except 'treatment as usual' (TAU), which means that the participants received the treatment or intervention by the specialist, i. e., speech pathologist or special needs teacher[39], as planned independently from the study participation. These participants also received conventional writing instruction by the class teacher. The participants were assessed at pre- and post-test by the specialists. The 'experimental group 2' (EG 2)[40], red in Figure

39 Due to different policies in the various cantons, the individual setting can be with a speech pathologist or a special educational needs teacher.

40 The term 'experimental group 2' could lead to confusion, because the purpose of this group is to control for the STT intervention. This term was selected over a possible alternative, e. g. control group 2, for two reasons: firstly, because this group does receive an intervention; and secondly, because, in the communication in German with teachers, specialists and authorities, it was always referred to as an experimental group (in German: Interventionsgruppe 2), which is why this is maintained as EG 2 in the present thesis.

5.2, received the above-mentioned writing instruction by videos in the writing classroom. The participants also received the 'treatment as usual' (TAU) with the specialist, which was planned independently from the study. This group was partly assessed by the specialists and partly within the writing classroom. The 'experimental group 1' (EG 1), coloured in blue in the figure, received both interventions, the assistance for text production by the STT-tool and the writing instruction by videos. As also illustrated in Figure 5.2 in light blue, the EG 1 received a software introduction prior to the intervention. Depending on the weekly time of speech therapy or support by the special needs teacher, the treatment time in the individual setting was entirely or partially used for the software instruction. During the intervention period, the participants of EG 1 also received TAU. In the case of issues with the STT-tool, the specialists functioned as contact persons for the classroom teacher and could use the single setting to provide support if necessary. EG 1 was also assessed partly by the specialists and partly within the writing classroom.

Timeline of the study

The intervention study took place during the school year 2022/2023. The introduction of the STT-software in EG 1 started after the summer holidays 2022, in the single setting of speech therapy or special needs education. The pre-test was administered after the autumn holidays, in October and November 2022, followed by the start of the intervention period in the writing classroom. The intervention lasted for 18 school weeks. Due to varying school holidays, and exceptional weeks, e. g. ski camps, the post-tests data collection took place between March and May 2023. Depending on the group, the testing took place in the single- and/or classroom settings (Table 5.11). The total assessment time per student was approx. 50 minutes for EG 2 and CG and approx. 70 minutes for EG 1. As mentioned above, another adaptation in the study design was made to increase the probability of a sufficiently large sample.

With one exception, all data were collected at pre- and post-test times. The exception was the assessment of writing motivation in EG 1. Since the software introduction could have altered the writing motivation, this part of the pre-test data was collected before the software introduction, to define the baseline of writing motivation, thus approximately six school weeks earlier than EG 2 and CG. The assumption was that the software equipment and possible beliefs in a valid solution or first-encountered difficulties could potentially impact the writing motivation in EG 1 during the software instruction. Natural changes during six school weeks were considered to be less important than the impact of the software instruction. The preliminary work, illustrated in light green in Figure 5.2, such as adaptations of assessment tools, and development of instructional videos or pilot studies, are mentioned, respectively, in Sections 5.4 and 5.5, presenting the single elements.

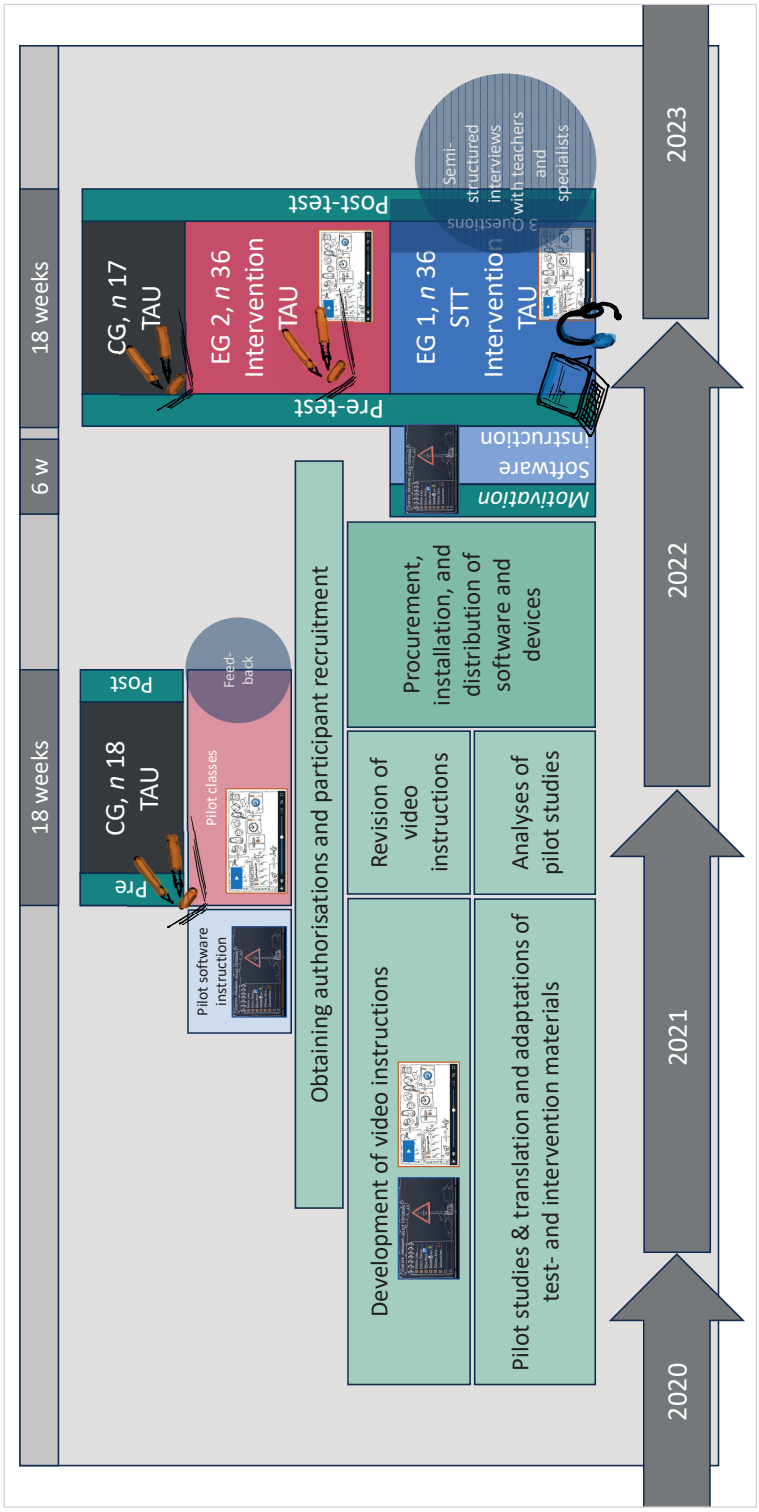


Figure 5.2: Timeline of quasi-experimental study with mixed methods design

Abandonment of randomisation

Experimental designs are based on the idea of the removal of irrelevant sources of variability that could reduce the precision of sample results and therefore reduce the statistical power of tests (Cohen, 1988). Randomisation is the usual way to control sources of bias that could result in undiscovered differences between the groups, which could impact an outcome variable and been mistaken as treatment effect. However, the participant recruitment, shown in the centre of Figure 5.2, proved challenging from the start, including recruiting participants for the pilot studies. For the following three reasons, the initially planned randomisation was abandoned, which was equal to a transformation of the initial plan of an experimental study, a so-called randomised controlled trial (RCT), to a quasi-experimental study. The following explains the three reasons for this important change in the design and explains how the risk of loss of reliability and power is encountered:

1) Firstly, the sample recruitment was anticipated to take longer than expected, which could have been problematic for timely and randomised group allocation. This was, in particular, due to the fact that the equipment for EG 1 had to be ordered in advance, set up, and distributed to the various regions of German-speaking Switzerland before the start of the software introduction, by the end of August 2022.
2) Secondly, the resulting self-assignment to the groups enabled more flexible participation options. For example, participation in EG 2 was possible if a teacher wanted to take part in the study, but the specialist only agreed to administer the pre- or post-test but not the software instruction. A participation in the CG was, meanwhile, possible if the specialist was willing to participate but not the class teacher. In addition, the first part of the control group could be surveyed in the 2021/2022-school year, as the test materials were ready by then.
3) Thirdly, several specialists stated therapeutic motives for participating in one group or another. Without randomisation, it was possible to take these motives into account, which seems ethically sensible, because a waiting-control-group design was not possible due to the duration of the study, lasting over ¾ of a school year, and in some cases the limited resources of remediation in individual settings for Grade 5 students.

To summarise, it is argued that the abandonment of randomisation has advantages for this study and is considered essential for the recruitment of the required *N*. However, the major disadvantages of this decision, namely the risk of the influence of known and especially unknown factors, need thorough consideration. The consequence of the self-assignment, according to Pildal et al. (2007), is that benefits of an experimental intervention are overestimated. In the present study, two measures are taken to counteract the possible bias of self-assignment. Firstly, further information of the participants was collected among specialists and parents to control for the comparability of the groups (see Section 5.3.4.6). Secondly, the study groups are

compared by statistical analysis on the even distribution of these characteristics and on the skills level at pre-test. The sample and the comparability of the groups are described in detail in Section 5.3.4.

The data collection in the CG took place over the two periods presented above. In autumn 2021, approximatively half of the CG was recruited, which is illustrated in the upper centre of Figure 5.2. In other words, pre- and post-tests were administered to 18 participants one year prior to the main data collection. An additional advantage of this procedure was the confirmation that all assessment tools and the instructions by videos for the assessment were handled without problems and worked out well.

The recruitment of the main sample for all groups confirmed the expected challenge, and the omission of the randomisation was subsequently considered as a suitable measure. This is because: on the one hand, the search for participants was delayed by waiting times for authorisations in certain cantons; while, on the other hand, recruiting the required number of participants needed several attempts via various channels and many individual contacts. The recruitment of participants for the CG and EG 2 was extended until autumn holidays 2022, just before the intervention started.

Scientific quality criteria

Important indicators for the quality of any experimental interventions are the scientific quality standards. Objectivity is a key feature of experimental methods with well-conducted experiments, and an essential standard of replicability (Morgan, 2014). Validity, another quality standard, stands for experiments and/or instruments that measure what they claim to measure. They are also intended to be reliable, which means that they produce similar results across time and across measures (Glogowska, 2011). To this end, the outcome variables must, for example, be well-defined.

In addition to the precise description of the procedure and the already mentioned preliminary comparison of the groups, this thesis addresses the quality standards at various points, but especially in Section 5.4 about the assessment tools. In an intervention study, however, not only the assessment tools are keys for quality standards, but also the standardisation and replicability of an intervention. A particularity of this study, and important for the quality standards, are the many video instructions that were provided for both the assessment and the writing instruction for the EG 1 and EG 2. The advantage of these videos is that they are the same for all participants and over time, i. e., at pre- and post-test. Since they are developed in advance by the researcher, the instructions most likely follow more precisely the intended instructions than an oral, simultaneous instruction. An additional advantage of the videos is that they could be used in a study replication.

In the present study, the assessments of the outcome variables were carried out by the specialists and partially by the class teachers. These were, therefore, not blinded, as appraised by Galuschka et al. (2014) for experimental studies, since the treatment

condition was known to the teachers and specialists. However, all test documents were coded with digitally randomised, five-digit numbers to hide the group affiliation. This was equivalent to blinding the scoring and analysis of all outcome data.

5.3.3.2 Statistical analyses – GLM

This thesis research requires statistical analyses to answer the RQs but also, as already mentioned, for the comparison of the study groups and for the adaptation and validation of the assessment instruments. This sub-section describes the main statistical model applied to answer RQs 1 to 4. It then presents the necessary sample size with its effect sizes and test power, followed by the potentially problematic self-assignment to the study groups and the resulting question of comparability of the study groups, due to the abandoned randomisation. It also addresses the assumptions required for these analyses. The preliminary investigations of the sample are presented in Section 5.3.4. All statistical analyses that are required for the adaptation and validation of assessment tools are included in Section 5.4.

All analyses to answer the research questions were run in IBM SPSS statistics, version 28.01.1.0 (IBM). The sample size was planned with the free software G*Power, version 3.1.9.7 (Buchner et al., 2014), to ensure the appropriate interpretation of the study results (Rasch et al., 2014).

RQs 1 to 4 investigate the effectiveness of one or two treatment conditions for writing skills and writing motivation. The two treatment conditions are STT and writing instruction by videos. Three aspects for the choice of the statistical model are important. Firstly, the data of the same persons must be analysed at two times, requiring a repeated measurement with dependent data. Secondly, Hypotheses A and B of RQs 1, 2, and 4 should each be tested for differences among the three groups with a single analysis, due to the two grouping variables: STT and/or writing instruction by videos. Thirdly, second-order effects are sought to be investigated and therefore integrated into the model, in order to detect specific characteristics of participants that potentially could amplify or attenuate effects.

The linear model framework, also known as the general linear model (GLM), allows this study to analyse the outcome of more than two groups at repeated measurement times (Field, 2018). The 'one-way repeated measures ANOVA', one of the many special cases of the GLM (Knapp, 1978), tests whether there is a statistically significant difference in the outcome of a variable at two or more measurement times (Bortz & Schuster, 2010; DATAtab Team, 2024). Such ANOVAs are run to answer RQs 1, 2 and 4, whereas *T*-tests, another case of the GLM, and ANOVAs are run to answer RQ 3.

The 'one-way repeated measures ANOVA' tests whether there is a difference or a main effect over time, in the present study between pre- and post-test measures. As shown in Equation 5.1, the total variance (SS_T) consists of the variability between the persons and of the variability within persons. The variability between persons is

the variance explained by the model, the so-called model sum of squares (SS_M), also known as the systematic variance. The variability within the persons consists of the residual variance (SS_R), which is the unexplained remaining variance (Rasch et al., 2014). The effect of the model is the repeated measure (Field, 2018), in the present case called *Z*, for 'Zeit' in German:

$$SS_T = SS_M + SS_Z + SS_R \tag{5.1}$$

To investigate the main effect of the treatment condition, the group factor (*G*) must be included into the model. This is possible by a 'two-way repeated measures ANOVA' (with repeated measurements on one factor). This model allows the research to answer three questions: 1) whether the repetition of the measurement *Z* has an impact on the outcome; 2) whether factor *G*, i. e., the group or treatment condition, has an impact on the outcome; and 3) whether there is an interaction effect between the factor *G* and *Z* (DATAtab Team, 2024), also known as first-order interaction (Rasch et al., 2014).

The model of the 'two-way repeated measures ANOVA' can be derived from the regression path diagram (J. M. Graham, 2008). It attributes the variance of the dependent variable to the two main effects, the interaction, and an error component (Bortz & Schuster, 2010). The model is based on Equation 5.2:

$$SS_T = SSM + SS_{factorG} + SS_{factorZ} + SS_{interactionG\,Z} + SS_R \tag{5.2}$$

(Field, 2018; Rasch et al., 2014).

The present study has a 2×3 plan with two independent variables (two measurement times and three groups, i. e., treatment conditions). Factor *Z* is coded with x_1, factor *G* is coded with x_2 and x_3, and the interaction $Z \cdot G$ is coded with x_4 ($x_1 x_2$) and x_5 ($x_2 x_3$). The Regression Equation 5.3 for this linear model with a first-order interaction shows the assumption that an estimated dependent variable ($\widehat{y}$) vy is predicted by the constant (b_0), also called intercept, the indicator variables (b), in the present example b_1 to b_5.

$$\hat{y} = b_0 + b_1 Z + b_2 G_1 + b_3 G_2 + b_4 Z \cdot G_1 + b_5 Z \cdot G_2 \tag{5.3}$$

(Jaccard & Turrisi, 2003).

The 'two-way repeated measures ANOVA' is calculated in software such as SPSS. The calculation tests the overall fit of the linear model, elaborated above, to a set of observed data, by analysing whether using group means predict the outcome significantly better than using the mean of all scores. The two-way repeated measures ANOVA calculates the *F*-statistic. A large *F* indicates that group means differ, which allows the discrimination between cases from different groups. Meanwhile, a small *F* indicates that the ability to predict observed data will be poor, because group means are the same or similar (Field, 2018). The significance test and the effect size can be derived from *F*-statistics (Rasch et al., 2014), which is now automatically calculated

in software such as SPSS. SPSS transforms the effect size for ANOVAs with repeated measures into partial eta square η_p^2 (). This *F*-test does not identify differences between specific means; however, the model parameters (the *b*-values) do (Field, 2018). To identify these differences, post hoc tests or visual inspection of outcome graphs are necessary.

As elaborated above, 'two-way repeated measures ANOVA' can explain some of the differences in the measured values by the 'systematic variance' or 'effect variance', due to the differences in the experimental manipulation (Rasch et al., 2014). However, human behaviour is often determined by more than one factor. In the present study, several additional factors, such as co-morbidities or gender, potentially have an effect on the outcome. Of particular interest also are interactions between the main factors and these additional factors. If there is a systematic effect of an additional factor, this explains part of the residual variance. Therefore, by reducing the unexplained variance, a possible effect of the main factor becomes either more pronounced (Rasch et al., 2014), or a so-far suppressed effect becomes visible.

Three-factorial designs are possible with a repeated measures ANOVA if, in addition to the main effects and the first-order interactions, the second-order interaction is also coded. Correspondingly, in the two-factorial ANOVA, the additional indicator variables are generated by multiplying the indicator variables that are involved with main effects in the second-order interaction (Bortz & Schuster, 2010). In other words, factor *C* – in the present analysis, for 'students' characteristic' – is added to and multiplied with the factors *Z* and *G*, resulting in main effects *Z*, *G*, and *C* as well as the interaction effects *ZG*, *ZC*, *GC* and *ZGC*. The regression equation for this model is based on the Equation 5.3 but extended by the factor *C* and all interactions.

In the present study, the GLM with repeated measures is considered adequate to answer the RQs. This allows the research to analyse: 1) effects over time between pre- and post-test; 2) effects of the three different treatment conditions over time; and 3) potentially suppressing effects of additional, non-repeated factors. By contrast, no repeated measures were necessary to calculate differences between writing modalities for RQ 3A. This could, therefore, be done by *T*-tests or one-way *F*-tests.

Prior to the data analysis, the questions about the statistical power and the sample size planning, the assumptions for the GLM, and the question about one- or two-sided testing, must be considered. These questions are discussed in the following subsection.

Power analysis and sample size planning

The power of a test is the probability that a given test will find an effect if there is one in the population. This relates to the Type II error, which is to miss an effect that genuinely exists, and which is the β-level. "The power of a test can be expressed as $1 - \beta$" (Field, 2018, p. 84). The power analysis builds the core of a rational basis for deciding on the sample size to be used in an investigation (Cohen, 1965). G*Power

software (Buchner et al., 2014) is used for its a priori computation. The following paragraph describes the setting and the reasoning for the latter.

The sample size planning for a two-way repeated measures ANOVA depends in addition to the three parameters of statistical interference i. e. power, significance criterion, and effect size, and also on the mean correlation of the factor 'measurement occasions', i. e., the correlation between the outcomes at pre- and post-tests, which is not known before running the experiment. Since no other studies could provide guidance, the correlation between repeated measures is set to $r = 0.5$. According to Rasch et al. (2014), the assumption of a correlation is justified when the repeatedly measured characteristics can be regarded as relatively stable within test subjects but differ between test subjects. This is the case for intellectual skills or personality features (Rasch et al., 2014). In the present work, this stability is assumed for writing skills and most components of writing motivation. Self-efficacy was discussed as to some degree dependent on the task and might therefore be less stable. However, because the tasks in pre- and post-tests are identical, the stability is also presumed for self-efficacy.

Power, significance criterion, effect size and sample size are related to each other. Therefore, one is a function of the three others and influence each other. For the present study, the anticipated effect size f is set to .25 which is a medium effect size according to Cohen (1988). The test power is set to $1 - \beta = .90$, together with an α-level of .05. While power on .90 is rather conservative, the α-level is not set stricter to avoid a higher probability of Type II error (Field, 2018). The number of groups for the within-between interactions is calculated for three treatment conditions which were analysed on up to two further non-repeated variables (3 x 2 x 2 = 12). The addition of factors has advantages and disadvantages. They may explain further residual variance, but at the same time they may decrease the power (Rasch et al., 2014). Therefore, during the analyses, attention must be paid to whether a factor has scientific value.

The power analysis results in a sample requirement of $N = 96$. The repeated measures ANOVA therefore requires a minimum group sample size of $n = 32$.

Repeated testing and inflated error rates

The investigation of students' characteristics that might amplify or attenuate an effect of the treatment condition results in the repetition of statistical analyses on the same data. This increases the risk of an α-error accumulation, also called Type I error or family-wise error, which must be considered. According to Field (2018), there is not only one way to counteract this build-up of errors. It is possible to adjust the level of significance for individual tests so that the overall α-error rate across all comparisons remains at .05. The most popular and easiest way is to divide α by the number of comparisons k, known as the Bonferroni correction, shown in Equation 5.4. In the present work, the significance level therefore will be adjusted according to

the number of repeated analyses, nevertheless keeping in mind that the Bonferroni's test has the disadvantage that it is conservative with an impact on the power:

$$p_{\text{crit}} = \frac{\alpha}{k} \tag{5.4}$$

(Field, 2018).

As elaborated above, the repeated measures ANOVA computes the analyses of the main effects and the interaction separately. As presented below, up to seven students' characteristics are checked on second-order effects and interactions. Equation 5.5 results in a p level of .005. Therefore, only statistical results with a significance of $p \leq .005$ will be interpreted and discussed further:

$$p_{\text{crit}} = 0{,}005 = \frac{0.05}{10} \tag{5.5}$$

Assumptions for repeated measures ANOVAs

The assumptions for repeated measures ANOVAs are: 1) the interval scaling of the outcome measures; 2) the normal distribution of the outcome measures within groups; and 3) the homogeneity of the variances within groups – instead of the independency of the data, which is, per se, not given in the repeated measures ANOVA, an assumption about the type of dependency of the data of the different measurement times applies to methods with repeated measurements; and 4) the correlations between the individual levels of the repeated measurement factor must be homogeneous (Rasch et al., 2014).

In the present data, the first assumption of interval scaled outcome measures is met as further elaborated in Section 5.4. For the second assumption, different possibilities to deal with the assumption and violation of normal distribution exist. While Field (2018) does not question the importance of meeting the assumptions for the accuracy of the test statistics and p-values, he expresses concerns about significance testing for normality, because they depend much on the sample size. The present sample sizes, between $n = 34$ and $n = 36$ in each group, lie above the widely accepted 'large enough value of 30', but there is no simple answer because this also depends on the population distribution (Field, 2018). The size of the present sample could have been considered as being only on the edge of ignoring normality, according to this indication. However, Bortz and Schuster (2010) write that $n < 10$ is a small sample size and assume that ANOVAs are relatively robust above this size, when the groups are equal in size. The group sizes of n 34 to n 36 were considered as equal. According to both indications, the sample size in the present study is assumed to be large enough to accept the second assumption for the analysis of the main effects and the main interactions. However, boxplots and Q-Q plots are visually checked for outliers. Unless many outliers are found, these are excluded from the analysis because repeated measures ANOVAs are sensitive to outliers. In case of a larger number of outliers, a non-parametric alternative can be considered (Hemmerich, 2024b). The

analyses of second-order effects and interactions reduce the sample size through the further division into subgroups. For these analyses, the scattering of the data by graphs were visualised and significant tests of normality were run. However, significant tests have to be treated with caution, according to Field (2018, p. 248), who writes, for small samples, that one must "pay attention if your significance tests are significant but resist being lulled into a false sense of security if they are not". The aim of these observations, therefore, is to interpret significant second-order effects with caution. As for the third assumption about homogeneity of variance, Field (2018) discusses concerns of Levene's test. However, on the one hand, repeated measures ANOVAs are rather robust for minor violations of these assumptions with equal sized groups (Methodenberatung, 2024). On the other hand, in the present research, the common interpretation of Greenhouse-Geisser and Huynh-Feldt adjustments are no alternative since only two measurement points are analysed (Girden, 1992). Therefore, Levene's tests are checked for major violations.

Field (2018) names, in addition, uncorrelated predictor variables as an assumption for statistical analyses with more than one predictor variable. These predictor variables should not correlate too highly. At the same time, predictors should also be uncorrelated with 'external variables' because they, otherwise, could lead to unreliable conclusions drawn from a model (Field, 2018). As elaborated above in Section 5.3.3.1, further information of participants is collected due to the abandoned randomisation. Theoretically, some variables describing participants' characteristics could be correlated with each other; e. g. time of therapy, severity, or type of school may be correlated, since the enrolment in a specialised school or the amount of therapy time depends in general on severity or comorbidities. The risk of these correlations is counteracted in three ways. Firstly, Spearman's rank correlations among these characteristics are analysed (see Table 5.2 in Section 5.3.4.5). Secondly, potentially moderating effects of these variables were analysed by separate models; and because of the repetition of analyses, measures are taken to avoid inflated error rates. Thirdly, significant correlations must be considered in the interpretation of the results.

One-sided or two-sided significance testing

Some of the hypotheses are formulated as directional. This could be an argument to perform one-sided significance tests because of their greater power. Although there are occasional circumstances where the investigator does not need to know about effects in the opposite direction (Cohen, 1965), in the present thesis two-sided tests appear, for the majority of research questions, more reasonable because the state of research about STT for students with dyslexia is not considered solid enough to take the risk of leaving out important information.

The *F*-statistic only informs about the presence or not of significant differences between the groups but does not inform about how the study groups differ. In case of significant results, the post hoc analyses indicate how the treatment conditions

differ (Rasch et al., 2014), which is why post hoc analyses are run for comparisons between more than two groups.

Post hoc power analyses are run to investigate RQ 3 because the respective analyses are only run on one study group and therefore on a smaller sample. The reduction in size of the sample decreases the power and increases the risk of a type II error (Cohen, 1988). According to the resulting power, the results therefore must be interpreted with caution.

5.3.4 Sample – Quasi-experimental study

5.3.4.1 Participant recruitment

The power analysis, discussed above, indicates the necessary number of 96 participants, 32 in each group, to conduct the intended analyses. However, a higher number of participants is desirable, especially for EG 1, because students dropping out of a study, with a setting involving several people and implementing a challenging technology, is likely.

The main recruitment of participants took place between February 2022 and September 2022. Half of the CG was recruited before, see Section 5.3.3.1. 18 participants were recruited via existing contacts with speech pathologists and special educational needs teachers. Geographically, the recruitment took place in 21 of 26 Swiss cantons where German is either the only official language or one of the official languages. To the end of having an effective participant recruitment, an advertisement video, lasting for six minutes, in the style of the instructional materials for the writing instruction, was created. The video contained the key information on the study participation.

In the first step, the video was sent by email to speech pathologists of the school sector through the association of Swiss German speech pathologists (DLV, Deutschschweizer Logopädinnen und Logopädenverband). At the same time, the specialised schools for language (Sprachheilschulen) in Fribourg, Bern, Zurich and St. Gallen were contacted personally. In a second step, personal contacts were established with teachers, principals and stakeholders like school services, responsible for special needs education and SLT. The response rate allowed respondents to complete EG 1 by the end of June 2022, which was the latest date to equip this group and distribute the devices. By the end of June 2022, the recruitment of the CG and EG 2 was not yet completed. In addition, there was an imbalance between the number of participants from mainstream and specialised schools. The positive response rate from specialised schools in Freiburg and St. Gallen, interested specifically in participation in EG 1, was equalised by contacting the not yet contacted specialised schools running Grade 5 classes, in a third and final step. Besides that, the last canton's authorisation received by the end of June 2022 allowed the researcher to address fur-

ther people not yet contacted. The recruitment lasted until the autumn holidays, i. e. in October 2022, just before the start of the study after the holidays.

5.3.4.2 Inclusion and exclusion criteria

As emphasised in Section 5.3.4.1, this study aims to investigate a representative sample of the population with dyslexia. Therefore, the criteria for inclusion or exclusion are kept to a minimum. Two criteria are mandatory for the inclusion in the study:

1. All students must be enrolled in Grade 5 in German-speaking Switzerland.
2. All students must have received a diagnosis of dyslexia prior to the study, according to the cantons' policies.

Furthermore, the dictation should make sense for the students eligible for the study. The specialists of all participants were thus asked whether they could imagine the student to use STT. If, for example, a comorbidity such as AD(H)D or dyspraxia was so severe that the specialist could not imagine the student handling the device, the student was not included in the study. Regarding IQ, only students with an intellectual disability (IQ < 70) were excluded.

Both the students and the parents had to agree to participate. The parents gave their consent by signing the declaration of consent (Appendix A). All involved teachers and specialists also had to agree to participation in the study with its provided procedures. Grade 5 of primary school, in Switzerland also called '7 HarmoS' was selected for multiple reasons. Regarding the terms, this thesis research refers to the international literature and thus mentions the primary school years. Firstly, Svensson et al. (2019) suggest that compensational measures are installed for a student with persistent problems if there is no sufficient progress after three to four years of intensive support of traditional techniques. Secondly, regarding younger primary school students, the state of research on compensation for those students was, at this point in time, estimated to be weak. In addition, there are also reasons for promoting handwriting in the early school years (e. g. Hurschler Lichtsteiner, 2020; Santangelo & Graham, 2016). Thirdly, the literature review by Schröder et al. (2022), on digital support of students with special educational needs for narrative writing competencies, shows that most of the support by digital technologies is only implemented from Grade 5 onwards. It was therefore concluded that the acceptance rate for a study with younger kids might be too low. Fourthly, although the state of research indicates that compensation with older students might be more successful (Matre, 2022; Schüler, 2021), the following practical reasons led to the choice of Grade 5 for this study: 1) Grade 6 was excluded because of the transition from primary to secondary school, taking an important place in Switzerland in this respective school year; and 2) secondary school was excluded because of the limited number of students from Grade 7 onwards that are still regularly supported by speech pathologists

and/or special needs teachers, which would have greatly reduced the probability of recruiting a sufficiently large sample.

5.3.4.3 Sample description – Quasi-experimental study

A total of $N = 118$ students were recruited for the study. These participants were found in the following 10 out of 21 eligible cantons: Aargau, Basel-Land, Bern, Freiburg, Graubünden, St. Gallen, Schaffhausen, Thurgau, Zurich, and Zug. A total of 42 speech pathologists and special needs teachers participated with one or more than one student in the study. While some of these specialists only participated in the assessment of the students of the CG and/or EG 2 at pre- and post-tests, others also provided the software introduction to the students of EG 1 and EG 2. The students of EG 1 and EG 2 were enrolled in 36 classrooms, involving 35 teachers, since one teacher taught two classes in writing.

Sample attrition

Eleven students had to be excluded for diverse reasons: family moved away (2), lost data (2), execution of intervention obviously not according to indications (2), and problems with dictating and too much typing (4). One student was excluded because of Cochlear implants and therefore questionable diagnosis of dyslexia according to the definition in Chapter 2, Section 2.2.1. The eleven dropouts were equally distributed to EG 1 (6) and EG 2 (5). 107 students completed the study. Two students, one of the CG and one of EG 1, refused to write one of the two texts. For the student of the EG 1, this was the pre-test text, whereas the student of the CG refused to write the post-test text. One student of the EG 1 only started to write the text of the pre-test after five minutes. Therefore, for this participant there is no CBM measurement available for the pre-test.

Demographic information of the sample

Table 5.1 provides demographic information on the sample and students' characteristics according to the treatment condition.

5.3.4.4 Participant characteristics and comparability of the study groups

The aim of investigating the efficacy of STT in the field with diverse students with dyslexia led to the decision to recruit students with a minimum of exclusion criteria, as stated in Section 5.3.4.2. Therefore, the participants may vary considerably due to factors such as comorbidities, type of schooling, or multilingualism. In addition, as elaborated in Section 5.3.3.1, the omission of the randomisation of the group assignment, in favour of a sufficiently large sample, carries the risk that the groups differ

Table 5.1: Sample and students' characteristics

Variable		Condition (study group)		
		CG: Control (TAU) *n* = 35	*EG 1: STT & video writing instruction (TAU)* *n* = 36	*EG 2: video writing instruction only (TAU)* *n* = 36
Age (month)				
M		136.77	135.36	134.64
SD		7.97	6.89	5.45
Gender				
Male		26	22	25
Female		9	14	11
Language(s)				
German		15	24	16
German + other(s)		20	12	20
Type of school				
Regular		23	21	23
Special language needs		12	15	13
Severity of Dyslexia				
Severe		15	22	14
Moderate		17	10	20
Mild		1	0	1
Comorbidities				
DLD	no	14	22	21
	yes	21	13	14
Phonetic-phonologic dis.	no	28	33	34
	yes	7	2	1
AD(H)D	no	27	24	26
	yes	8	11	9
Dyspraxia	no	33	33	34
	yes	2	2	1
LD+ [1]	no	29	25	27
	yes	6	6	6
1 or more co-morbidities[2]	No	6	9	8
	Yes	29	26	27
Time of therapy/ind. setting				
M (min/week)		60.68	65.52	50.58
SD (min/week)		37.24	27.12	36.03
Parents' highest degree				
None		0	0	0
Compulsory education (CH)		7	7	11
Graduation or apprenticeship dipl.		12	17	13
Higher degree (BA, MA …)		8	12	8

1 Inconsistent information by the specialists, e. g. no answer for LD, but indication of reinforced measures in certain subjects, was interpreted as presence of LD+.
2 One missing in each group, i. e., EG 1 and EG 2.

regarding certain factors and that the results therefore lose validity. To address both issues, further information about the participants (see Appendix B) and parents (see Appendix C) was collected. The following paragraphs list the information collected and the rationale for its selection. They also explain why IQ is not considered necessary to assess. Possible correlations between participants' characteristics are examined and presented in Table 5.2 at the end of this section.

Age: This study investigates the effectiveness of STT only among Grade 5 students. The age range is therefore limited. However, age might play a role because the school careers of students with dyslexia are not straightforward (Schulte-Körne & Galuschka, 2019) and can thus vary.

Gender: The research literature describes gender differences in written language development (Reilly et al., 2019) and in motivation (Camacho et al., 2021). Therefore, gender is included both to compare the groups and as a possible variable with second-order effects. Gender is assessed as a binary variable (male/female) to be in line with the binary standardisation of the SLS 2–9 (Wimmer & Mayringer, 2014) applied in this study.

Language(s): German as a second language (L2) can be assumed as a factor that can impact the use of STT due to language skills such as pronunciation or morpho-syntactic skills. Usually, a diagnosis of dyslexia and subsequent speech therapy is only given after a certain time period of language contact with the school language. All L2 learners are therefore expected to speak German. In this study, data are collected about whether a pupil speaks 'German only' or 'German and one or more other languages' but not about the specific languages.

Type of school: In Switzerland, students with dyslexia are enrolled in regular schools or in specialised schools for language disorders, in German, the so called 'Sprachheilschulen'. While dyslexia by itself is not a sufficient factor for enrolment in a specialised school, according to the cantons' policies, it is commonly present in these specialised schools due to co-occurrence of dyslexia with DLD and other impairments (e. g. Kanton Zürich, Bildungsdirektion, 2022). Specialised schools traditionally accommodate students with severe language disorders only. Today, certain schools, e. g. the specialised school of the canton of Freiburg, also accommodate students with high needs for socio-pedagogical support (Institut St. Joseph du Guintzet, 2021). Important for this thesis study is that the groups are equally distributed concerning the type of schooling and comorbidities (see below).

Severity of dyslexia: In Switzerland, no official and nation-wide policy on diagnosis of dyslexia is available. According to Snowling et al. (2020), the severity of dyslexia plays an important role in the development and impact of the disorder. Therefore, information about severity appears important for this intervention study. However, as there is no standardised cut-off for severity, this information is collected as a subjective estimation by the specialist, assessed as severe, moderate or mild dyslexia. This study does not differentiate between students with global and isolated reading or spelling difficulties.

To obtain a notion about the validity of the subjective scoring, Spearman's rank correlations are calculated between the severity estimations by the specialists, on the one hand, and the total reading fluency score and the CWS produced in 5 minutes, on the other hand: Spearman's $\rho = .620$, $p < .001$ for the reading fluency, $\rho = .417$, $p < .001$ for writing fluency. According to Cohen (1988), the severity estimation in the present findings correlates strongly with reading and moderately with the writing score.

Comorbidities: As elaborated above, comorbidities are frequent and important features in combination with dyslexia (McArthur et al., 2000). Information about the presence or not of DLD, phonetic-phonological disorder, dyspraxia, learning disability (LD) and AD(H)D are collected. No difference is made between attention-deficit disorders with or without hyperactivity. As discussed in Chapter 2, Section 2.2.1, there is disagreement on LD as exclusion criteria from dyslexia. The participants for the present study are recruited from several Swiss cantons with different policies according to this question. Therefore, it is possible that participants with LD are in the sample, which is why this information is collected. Specialists were asked to tick the presence of LD, but the information sheet also contained space for further diagnostical information. This part of the data collection resulted in some ambivalent information regarding LD and information on support provided in other subjects. Thus, all statements on LD and ambivalent information of educational support were summarised as LD+. For example, if LD was not ticked but an indication of support needs in other subjects was given, this was recorded under LD+.

Time of therapy: All study groups received treatment as usual (TAU) during the intervention period. TAU is understood as support in the individual setting given by a speech pathologist or special needs teacher. The amount of time in individual settings, therefore, could impact the study results and must be comparable in all groups. In general, the specialised schools provide at least two speech therapy sessions per week, whereas two sessions per week in regular schools are less common despite severity, due to the limited resources available. Time of therapy is counted in minutes per week. Specialists were held to give an average over the school year in case of irregular therapies.

As discussed in Section 5.3.3.1, time of therapy is analysed on correlations with other characteristics of students. Four characteristics correlate significantly with time of therapy (see Table 5.2). Type of school with ρ .825, and p <.001, severity of dyslexia with ρ .336, and p <.001, and DLD with ρ .492, and p <.001, are all strongly correlated with time of therapy. LD+ correlates significantly, but with weak effect size, with time of therapy, with ρ -.238, $p = .020$ (Cohen, 1992). Participants with LD+ receive less time in the individual setting than the others. A possible explanation for this might be that these participants receive additional support by other specialists.

These results led to a subsequent measure in the data analysis: under the conditions that the groups are comparable in terms of time of therapy initially, this variable is primarily only analysed in this context. This avoids strongly overlapping

predictors in the model. Nevertheless, these correlations must be kept in mind for the interpretation of results, especially concerning type of school. If the school type shows a significant effect between groups and over time, time of therapy might be secondarily added to the model to avoid an unreliable conclusion due to time of therapy as an external factor.

Parents' highest degree of education: Access and support to use a digital tool could depend on the socioeconomic environment in which a student grows up. Therefore, in this study with self-assignment to one of the three groups, this could be a bias. It is imaginable that parents would promote or prohibit the participation in a specific group due to differently informed reflexions. Therefore, the highest degree of education from either one of the parents as an indicator of socioeconomic status is asked for. To avoid embarrassment, parents could return the answer sealed. The scale contains the following:

- *no compulsory schooling completed*
- *ompulsory schooling completed (i. e., nine school years)*
- *econdary II completed, i. e., graduation of high school or apprenticeship*
- *egree of higher education, i. e., bachelor's degree or higher*

The highest degree of the parents is found to be correlated with the type of school, language, and LD+. This is consistent with the prevalence of dyslexia and comorbidities, described, respectively, in Chapter 2, Section 2.2.3.2 and Section 2.2.3.3 about environmental factors and comorbidities.

Intelligence quotient (IQ): The lack of common policies and definition of dyslexia in terms of IQ, discussed in Chapter 2, Section 2.2.1, results in an inconsistent handling among Swiss cantonal educational systems. Consequently, in Switzerland the IQ is not assessed for every student with dyslexia, and therefore there is no available information. For two reasons, it is argued that the exact IQ is not a necessity for this thesis study. Firstly, extreme outliers in terms of IQ on both sides of the IQ curve are unlikely. Students with an IQ below 70, i. e., with intellectual disabilities, are not included in the study. Teachers and specialists dispose of this information. Meanwhile, concerning highly gifted children, i. e., > 130, the literature states the difficulty of the detection of dyslexia. This is due to the fact that highly gifted students with difficulties in literacy often manage to compensate for difficulties by their high intellectual competences, until advanced school years. As a result, these students are difficult to be recognised as 'dyslexic' because the assessment tools are not performant enough for its detection (Chusseau, 2018). Secondly, it is argued, in line with Chapter 2, Section 2.2.1, that interventions do not primarily depend on the IQ. Nevertheless, as presented above, for the students with LD, defined as IQ between 70 to 85, the information is collected as part of the comorbidities and used to control for second-order effects. For this group, in contrast to the counterpart on the upper side of the IQ curve, i. e., IQ 115–129, this information might be important to understand potential difficulties of students when learning to deal with a new tool and new procedures.

Table 5.2: Spearman's rank correlation coefficients between features and characteristics of participants

	Time	Age	Gender	Lang.	School	Severity	DLD	Phon.Dis	AD(H)D	LD+
Time	–									
Age	.172	–								
Gender	-.013	.027	–							
Language	-.012	.017	.099	–						
School	**.825*** **	**.197***	-.071	.099	–					
Severity	**-.336*** **	-.166	.104	.117	**-.299****	–				
DLD	**.492*** **	.184	-.086	**.235***	**.679*** **	**-.207***	–			
Phon. Dis.	.083	-.017	**-.220***	.015	.086	-.031	.093	–		
AD(H)D	-.036	**-.209***	-.176	**-.230***	-.107	-.040	**-.251****	-.049	–	
LD+	**-.238***	.044	-.018	.164	**-.292****	-.071	**-.224***	-.082	-.008	–
Degree	-.160	.133	-.091	**-.291****	**-.294****	.115	-.153	.031	.057	**-.216***

* $p < .05$; ** $p < .01$; *** $p < .001$

The three groups are compared on each of these characteristics. Chi² tests are run for the group comparisons of the categorial variables: Gender, language(s), type of school, severity, comorbidities, and parents' highest educational degree. One-way ANOVAs are applied for the metric variables of age and time of therapy.

An additional possibility to assure group comparability is to compare the outcome measures of interest at the first measurement time. This is done systematically for all outcome measures before running the intended model, as presented at the beginning of each results section in Chapter 6, Section 6.1. In case of significant differences between the groups regarding pre-test scores, an alternative statistical model, a univariate ANOVA (UNIANOVA), should be run. To this end, the change over time is calculated by the difference between pre- and post-test, serving as the outcome variable. The baseline score for the pre-test becomes then a control variable.

The results of the comparisons (Table 5.3) show that the groups differ in their distribution significantly on a single characteristic, namely the phonetic-phonological disorder. The number of students affected in each group, i. e., CG 7, EG 1, 2, EG 2 1 differs with $p = .033$. Since this characteristic could mostly impact the treatment condition of EG 1, i. e., dictating, and because the significant difference is not to the disadvantage of EG 1, this difference between the groups is considered not to impact the study in a significant way. While phonetic-phonological disorders can impact the writing process in a negative way, it is argued at this point that this problem would already be captured by the severity of dyslexia. This difference thus will only be considered in the case of unexpected results. All other analyses regarding these characteristics did not result in significant differences between the groups according to these characteristics. Thus, the planned statistical model should therefore not be biased due to such differences between the groups.

Table 5.3: Comparison of groups according to participant characteristics

Variable	Comparison between groups
Age	$F(2,104) = .89, p = .413$
Gender	$X^2(2, N = 107) = 1.46, p = .482$
Language	$X^2(2, N = 107) = 5.08, p = .079$
Type of school	$X^2(2, N = 107) = .45, p = .798$
Severity	$X^2(2, N = 107) = 6.55, p = .162$
Co-morbidities	
DLD	$X^2(2, N = 105) = 4.38, p = .112$
Phonetic-phonological disorder	$X^2(2, N = 105) = 6.85, \mathbf{p = .033}$*
AD(H)D	$X^2(2, N = 105) = .68, p = .711$
Dyspraxia	$X^2(2, N = 105) = .42, p = .811$
LD+	$X^2(2, N = 105) = .37, p = .833$
Highest degree	$X^2(2, N = 105) = 2.00, p = .735$
Time of therapy	$F(2,96) = 1.65, p = .197$

5.3.5 Qualitative supplemental component – Semi-structured interviews

RQ 5, about challenges of implementing STT in school, requires a qualitative approach (see Section 5.3.1). The design for this qualitative part, the supplemental component, is presented in this section. Firstly, the research interest pursued by this component is recalled. Decisions regarding the research methods, the tools applied, the criteria for the selection of interviewees, and their demographics are presented.

The intention of a qualitative supplement (qual) to a research project with a quantitative core (QUAN) could be characterised as achieving a broader picture by "identifying relevant and important factors in need of exploration which were not previously salient to the researcher" (Glogowska, 2011, p. 254), or as a compensation for inadequacies in meaning or detail that may occur with a quantitative core (Morse & Niehaus, 2009). Both reasons may apply to a research project in the field of speech language therapy (SLT) because complexity is an inherent part of many interventions conducted in the field of SLT. Reducing this complexity to outcomes showing whether the intervention appears to be effective or not may leave many important questions unanswered (Glogowska, 2011). Therefore, on the one hand, the qualitative supplemental component could inform about possible difficulties met during the intervention that impact the generated data of the quantitative core component and thus increase the understanding of those results. On the other hand, by the exploration of potential difficulties of the tool-application in the classroom, this component could increase depth or breadth of knowledge about the technology implementation with its benefits and challenges or raise more specific questions for future research (Greene et al., 1989; Morse & Niehaus, 2009).

5.3.5.1 Design of qualitative supplemental component

As stated above, the qualitative supplement aims to gather and structure further information and content about challenges and benefits that teachers and specialists observe during the study, comprising the software instruction and the intervention period in the classroom. Although, for example, the attitude of professionals in respect to digital technologies in school is of great importance for the implementation of a tool, the focus in the frame of the supplemental component is not to uncover unconscious or hidden patterns and meaning. However, additional interests and insights that this qualitative part brings to light are approved. Thus, the present approach seeks a method including both, i. e., to gather information about the topic of interest and to allow additional information. The semi-structured interview meets this demand (Morgan, 2014). A semi-structured interview is based on an interview guide, with a certain flexibility (Helfferich, 2011).

Regarding the group of people estimated the most suitable to explore RQ 5, the following considerations, based on the manual for conducting qualitative inter-

views[41] by Helfferich (2011), led to the selection of teachers and specialists of EG 1 as interviewees. Firstly, only EG 1 received the software introduction and used the tool during the intervention period. Secondly, the people most actively involved in the intervention study were the students, the specialists and the teachers. It was assumed at this point that the teachers had the broadest view of the challenges of implementing a tool in the classroom. However, thirdly, the professionals such as speech pathologists or special needs teachers who provide the instruction to use the tool are also close by and may bring in an additional outside perspective. This leads to a further specification of the interview type for this thesis. Since the interviewees are experts and the research interest is primarily focussed on factual information, the type of interview can be described as an expert interview. This type of interview tolerates more control and structuring interventions by the interview guide. The information is assumed to be 'true', rather than a construction the meaning of which must be decoded – in contrast to a narrative perspective, where information is context-bound and subjective truths according to (Helfferich, 2011, p. 76).

Within the group of teachers and specialists, a heterogeneous sample is aimed for. However, the group composition is not taken into account in the case of a dropout. As the aim of this research question is to identify the difficulties, at least one professional from each dropout is asked for an interview. To gather a maximum of information about both the software introduction and the intervention, the interviews are conducted towards the end of the quantitative core component, as illustrated in Figure 5.3.

At the beginning of the study, the professionals of EG 1 received notice of a possible later request for an interview. The selected interviewees were contacted during the intervention period.

Regarding the role of the students of EG 1, their opinion and how they felt about the tool could be pivotal for appropriate understanding of the results. Therefore, all students of EG 1, including the students dropping out during the intervention, answered three additional questions at the end of the intervention, also as illustrated in Figure 5.3. However, the students were not considered for the exploration of RQ 5 since their perspective on the implementation of the technology in school was estimated to be narrower.

The qualitative part of this thesis research is a supplement and is, therefore, not expected to stand alone, as elaborated in Section 5.3.1. For this reason, theoretical saturation[42] was not sought in the interviews and no additional interviews conduct-

41 "Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews", see Helfferich (2011).

42 'Theoretical saturation' is a term from Grounded Theory methodology that stands for a (provisional) end point of the analysis, meaning that new data do not provide any substantial increase in knowledge for the generated theory; see Mey and Mruck (2011).

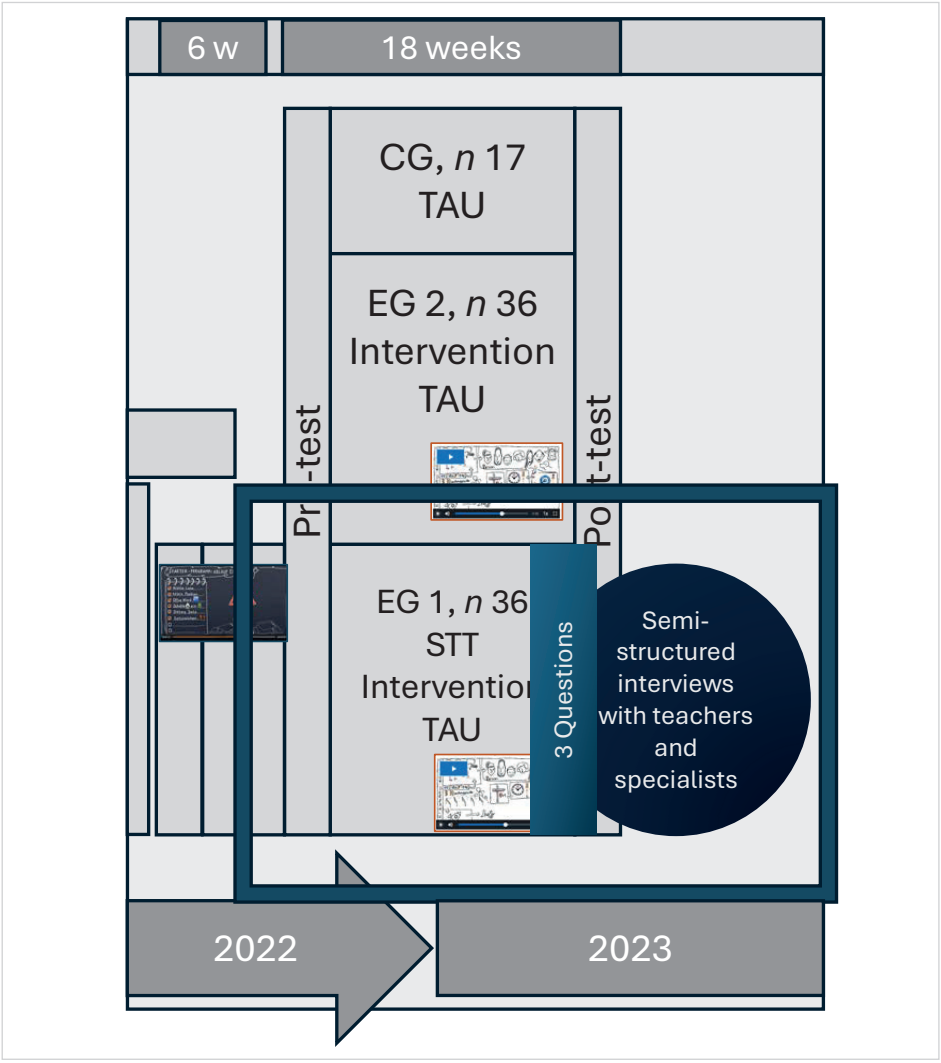


Figure 5.3: Timeline and position of the qual. supplemental component within the study

ed during the analysis. The number of interviews[43] was determined, on the one hand, by the number of dropouts, and, on the other hand, by the aim of heterogeneity of the interviewees. However, Helfferich (2011) indicates, in her manual, between 6 and 30 interviews as a 'medium' sample size, estimating six as the minimum for a hermeneutic interpretation. The aim for the present research component was to reach at least a medium sample size.

The semi-structured interview requires an interview guide (Helfferich, 2011). Its construction is presented in detail within the research instruments in Section 5.4.5. The interviews were audio recorded and subsequently prepared for analysis. Therefore, the audio recordings were transcribed to text with focus on verbatim transcription, meaning that pauses, fillers etc. are not transcribed. This process is a reduction of data the researcher must be aware of (Richards & Morse, 2013). The interviews in Swiss German were transcribed by the researcher, while Transkripto.de (2024) was mandated for the transcription for interviews conducted in Standard High German. The transcription rules correspond to the rules for verbatim transcription by Transkripto.de (2024).

A further decisive step in the design of an interview analysis is the decision on the evaluation strategy (Helfferich, 2011). The aim of the present research supplement is, rather, a detailed description and understanding of the data but no development of a theory such as targeted by the Grounded Theory method (Rädiker & Kuckartz, 2019). The qualitative content analysis allows the present study to analyse and structure linguistic material systematically through figures of semantic units. The analyses seek to filter out themes, content and relevant facets in order to summarise the elaborated paraphrases within subcategories and categories (Mayring, 2022). To this end, MAXQDA, a qualitative data analysis software (QDAS), has been employed. Although MAXQDA supports all phases of the analysis in a project, the central function of MAXQDA employed in this part of the present study is the assignment of codes (categories) to selected parts of the data, in the present research, words or passages of a text, and the inductive generation of additional and refined codes (Rädiker & Kuckartz, 2019). The code system, also presented in Section 5.4.5, contains a first set of codes deduced from the state of research, i.e, a concept-driven approach, subsequently refined and supplemented by an inductive, data-driven process.

5.3.6 Sample – Interview study

As stated above, the sample is designed to include specialists and/or teachers of students who dropped out, and heterogeneous interviewees, purposively selected regarding: all disciplines; varying amount of professional experience; accompanying

43 The maximal number of possible interviews in the present study is limited by the number of participants in EG 1, i. e., teachers and specialists of $N = 42$.

one and more than one participant; different gender; professionals from mainstream and specialised schools; initiating the study participation; and not, with or without experience with STT.

In the course of the intervention, five students dropped out of EG 1, and one additional student was excluded because of an uncertain inclusion criterion. In one of these cases, an interview with the teacher or the speech pathologist was not possible. A total of eleven interviews with 12 professionals was conducted, with one interview, on the request of two teachers, conducted with them both together. Table 5.4 contains the demographic information of the interviewees.

In six cases, the interviews took place face-to-face at the workplace of the interviewee, and in one case in a public space close to the workplace. Three interviews were conducted via Microsoft Teams due to organisational issues, while one interview was conducted in a meeting room of the Zurich University of Teacher Education. The interviews lasted between 20 minutes and 1 ½ hours.

Table 5.4: Demographics of the interviewees

Profession	Less than 10 years of experience	More than 10 years of experience	Initiation of participation	Number of students assessed
Teacher in regular school	G. ♂			3
Special education teacher		TRR. ♀ 711[1]	✓	2
Speech and language therapist (regular school)		P. ♂ 72[2]	✓	1
	S. ♀			3
		L. ♀		1–1 Dropout
		D. ♀ 73[3]	✓	1
		J. ♀	✓	1
		MC. ♀	✓	1–1 Dropout
Teacher in specialized school for language impairment	MM. ♀	K. ♀		6 3–2 Dropouts
		M. ♂		9–1 excluded
Speech and language therapist (specialized school)	TR. ♀			*4–2 Dropouts*74[4]

1 Specialist that intended to use STT independently of the study.
2 Specialist with own experience in the use of STT (especially Nuance Dragon®).
3 Specialist who had been a colleague of the interviewer for many years.
4 Dropouts figuring at two places in the table.

5.4 Selection, adaptation, and fit of assessment instruments

The study design requires the development or adaptation and procurement of several study materials. Section 5.4 and the elements in green and in blue in Figure 5.4 present all assessment tools necessary for the present research, while Section

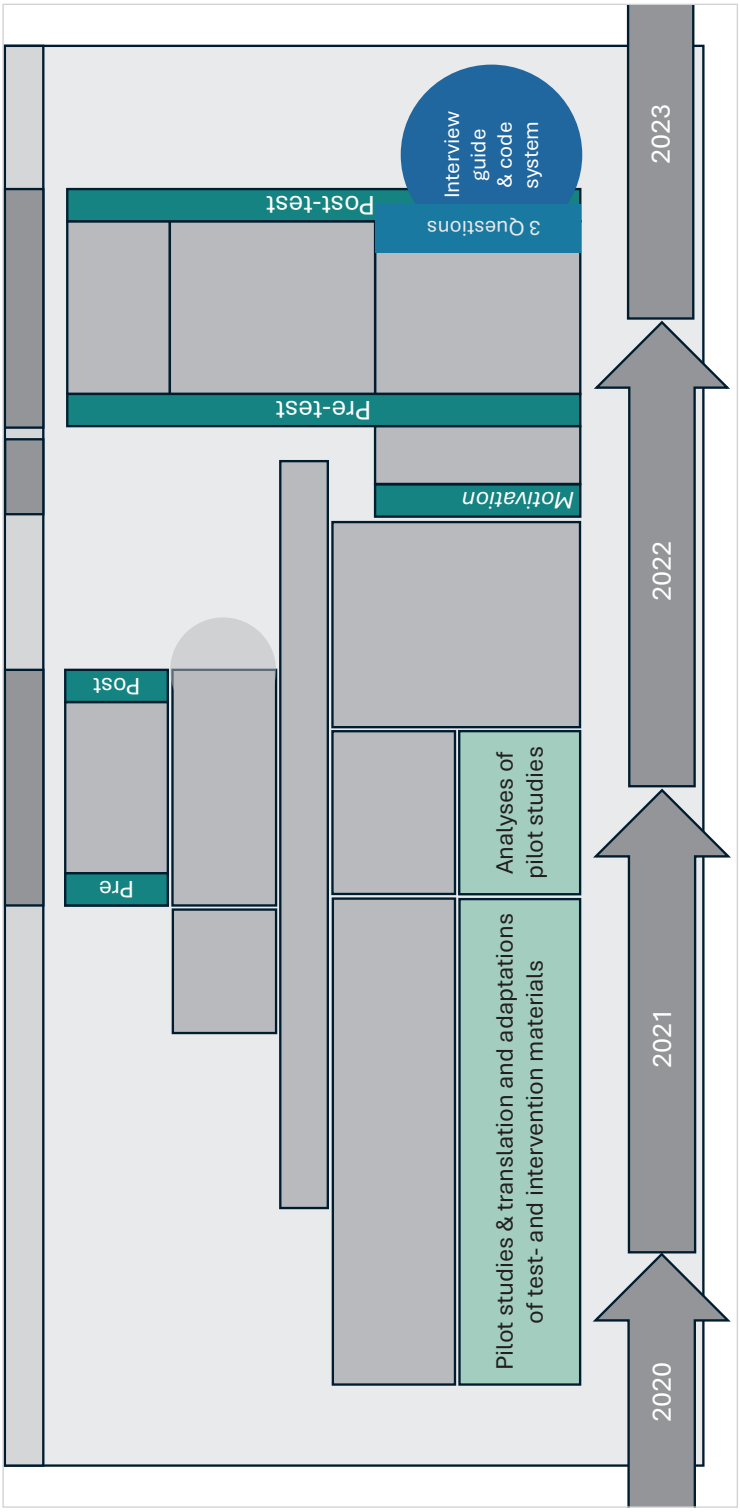


Figure 5.4: Development and application of selected and/or adapted study materials

5.5 shows the developed instructional materials for the software introduction and writing instruction, and the acquisition of hardware and software.

RQs 1 and 2 investigate the effect of STT on text quality and possible impact on reading and writing fluency. Therefore, an assessment tool or method for higher-order writing skills (see Section 5.4.1) and lower-order writing skills (see Section 5.4.2), as well as a tool for measuring reading fluency, must be selected or adapted to the study purposes. The effects on writing fluency are, by RQ 3, also investigated within dictated texts and between the handwritten and dictated texts of EG 1. RQ 4, asking about effects on writing motivation, requires a survey instrument to capture writing motivation and changes in writing motivation over time (see Section 5.4.4). It is desirable that these assessment tools are reliable, valid, and sensitive to changes. RQ 5, exploring challenges of the implementation of STT in the writing classroom, requires the development of an interview guide and a code system (see Section 5.4.5).

Several points must be kept in mind when selecting assessment methods and tools. Firstly, the quality criteria of the tests must be met. Where these are only partially satisfactory, this shall be considered in the interpretation of the results. A challenge of the study design, with a geographically dispersed sample and collaboration with many teachers and specialists, is met by using standardised assessment instruction by the means of the video instructions. The instructions contain the exact oral instructions with visual support where necessary, and the exact monitoring of assessment time. This method aims to maximise the comparability of the assessment. Secondly, the selection and adaptation of tools must be determined by the intended use of the test score and should match the feasibility of data collection and analysis in terms of the available time and financial resources in the research project. Thirdly, the burden on participants and teachers and specialists should remain within reasonable limits. Lastly, possible causes for rater effects (see Chapter 2, Section 2.1.3) should be avoided from the start (Feenstra, 2014). Knowledge about the writer's identity or, in the present case, the study group affiliation might, for instance, influence the raters' judgment (S. Graham et al., 2011). Therefore, all study documents were randomly coded by a five-digit number, created by a random number generator (Matheretter, 2024).

5.4.1 Assessment of text quality – TOWL-4

The aim of RQ 1 is to investigate text quality. Text quality in this thesis is understood in the rather broad sense according to Neumann (2017), which includes content-related aspects of text production such as coherence and cohesion, but also linguistic and orthographic aspects.

To date, there is no procedure in German to assess writing performance beyond the word level in a standardised way (see Chapter 2, Section 2.1.3). Therefore, non-standardised procedures are the only options for research and practice,

which should not be a problem per se because, according to the meta-analysis by Galuschka et al. (2014), the inclusion of studies with non-standardised measures does not lead to an overestimation of the effect sizes. However, Winkes and Schaller (2022) and S. Graham et al. (2011) state that informal procedures are fraught with difficulties such as questionable reliability for research. An additional difficulty for writing assessment is that different text genres need appropriate assessment tools (Dockrell et al., 2015), and inferences based on text quality scores should be limited to genre-specific writing in order to make valid inferences (Bouwer et al., 2015). Ideally, writing assessments should not only consider one text genre and one sample (Bouwer et al., 2015; Feenstra, 2014; Y.-S. G. Kim & Schatschneider, 2017).

A holistic approach in the present study is not considered, for two reasons. Firstly, in the present sample, the texts of students with dyslexia are expected to be at the bottom of the spectrum. Therefore, the differentiation within this sample and the sensibility to changes over the intervention period by a holistic approach are considered questionable. Secondly, analytical scoring methods are suited to provide more extensive information on writing performance (Feenstra, 2014). Therefore, an analytical assessment is considered to give more precise information about changes over time than a holistic approach.

Selection TOWL-4 and test description

Two subtests of the Test of Written Language – Fourth Edition (TOWL-4) have previously been translated to German and applied in research studies in Freiburg (CH), for example by Winkes and Schaller (2022). This translated version is considered as the most suitable and meaningful for this study. The present study aims to assess narrative texts, such as the TOWL-4 by Hammill and Larsen (2009). Subtest 6 includes contextual conventions such as the use of punctuation marks, capitalisations, or sentence structure. Subtest 7 focuses on story composition, e. g. plot (storyline), story sequence, or writing style, but also comprises items about vocabulary (Hammill & Larsen, 2009). The latter contains mostly information matching the above-determined definition of text quality. Therefore, for RQ 1, the score of Subtest 7 is used, which has the necessary construct validity based on theory. By contrast, Subtest 6 mainly provides information about lower-order writing processes, which is why this subtest does not appear meaningful for answering RQ 1. However, this subtest might be of interest in the investigation of RQ 2. Thus, the quality criteria for Subtest 6 are mentioned in this section, while the use of it for data analysis is addressed in Section 5.4.2.

The TOWL-4, published by PRO-ED, states the following purposes: measuring and identifying students in need of support or intervention; identifying strengths and weaknesses in students' writing abilities; documenting progress resulting from written language intervention; and providing measurement in written language research. It contains seven subtests. It is designed for 9;0 to 17;11-year-olds. The story

test, which is the focus of the present study, provides standardised instructions on the task and on elements that are essential for a good narrative text. It also gives an example story for a one picture input. After the instruction, the students receive the picture card (A or B). Each picture contains a main action, i. e., a main event and secondary scenes, which should stimulate writing. The students receive 5 minutes to plan, during which they are allowed to take notes. Then, they write for 15 minutes. The produced story provides the content assessed by Subtest 6, Contextual Conventions, and Subtest 7, Story Composition (Hammill & Larsen, 2009; McCrimmon & Climie, 2011). The scoring parameters vary in the number of points distributed per item: 0 or 1 points; 0, 1 or 2 points; and 0 to 3 points. Primarily, the scoring is based on yes/no qualifications or quantifiable data, such as 0 = not evident, 1 = 1–2 items evident, and 2 = 3+ items evident (McCrimmon & Climie, 2011). Subtest 7 should be scored according to adult standards for composition rather than age or grade standards. A Supplemental Practice Scoring Booklet contains ten sample stories and corresponding scoring results for scoring training (Hammill & Larsen, 2009).

Quality criteria of TOWL-4 and the translated version in German

The available German translation of TOWL-4 contains the translated instruction with the example story and the scoring items. The scoring items were partly translated and partly adapted to specific language features in German. Although this translation is not standardised for German, it offers the advantage of having a solid basis in English. The American version, according to McCrimmon and Climie (2011), meets acceptable standards regarding test construction and the quality criteria. These authors assume that the reliability estimates are consistently within the acceptable range. They also estimate that the validity is acceptable. Data for quality criteria are not available for German nor specifically for students with dyslexia. Meanwhile, Winkes and Schaller (2022) analyse the criterion validity in German of several measures of curriculum-based measurement writing (CBM-W) with the translated version of TOWL-4 scores. Although TOWL-4 and the CBM-W scores do not measure the exact same features of written language, they correlate in Grades 4 to 6 with >.50 – commonly accepted as the criterion validity threshold in English CBM-W-literature (McMaster & Campbell, 2008). To date, no standardised norms are available for German (Winkes & Schaller, 2022).

Writing researchers (e. g. Bouwer et al., 2015; Feenstra, 2014) suggest that the writing task is important to make valid interferences about the writing ability. By using the German version of TOWL-4, the task is in pre- and post-tests alike; but in the present study, a task-specific assessment according to the writing instruction is not estimated to be of advantage since writing instruction varies between the CG and both experimental groups.

A further important aspect for the validity of writing assessment is the number of raters for the scoring of the written product (Brouwer, 2012). The resources of

the present study do not allow several raters for each text, but the main researcher and the two additional raters are trained in a circular approach of discussing the items, clarifying the item description if necessary, rating ten texts, and calculating the interrater reliability. A total of 30 texts were rated within three common rating sessions. While the TOWL-4-manual indicates an interrater reliability of .80 (Hammill & Larsen, 2009), the interrater reliability in the present study shows a substantial agreement for Subtest 6, according to Landis and Koch (1977), while the agreement for Subtest 7 remains moderate (Table 5.5). The second session of adaptation to the item description resulted in an inferior interrater reliability. As a result, most of the newly added precisions were removed. After the third session, the scoring of the text was done by one rater only. The reason for this lower agreement in the present study, compared to the original test manual, is considered to lie in the difficulty of scoring low performance texts compared to adult standards.

Table 5.5: Interrater reliability of translated TOWL-4 version

Subtest	**Fleiss' κ**	**Fleiss' κ**	**Fleiss' κ**
	texts 1–10	*texts 11–20*	*texts 21–30*
6	.75	.66	.74
7	.50	.49	.55

Subsequently, two measures were taken to reduce the risk of low validity due to the limited agreement between raters. Firstly, the main researcher and both raters decided together on all ambiguous items encountered during the rating process. Secondly, the two raters were to rate the same part of the sample from pre- and post-tests, meaning that a participant's texts were rated both times by the same rater to work around bias due to the rater. The reliability within raters (intrarater reliability) after 5 months showed substantial agreement for both raters (Table 5.6), varying between Subtests 6 and 7.

Table 5.6: Intrarater reliability of translated TOWL-4 version

Rater	**Subtest**	**Cohen's κ**
1	6	.61
	7	.76
2	6	.78
	7	.61

The internal consistency of the German version of the TOWL-4, calculated for the sample of students with dyslexia, resulted in Cronbach's $\alpha = .83$ for the 12 items of Subtest 7 (TOWL-4 manual: .74/.72). This is considered a good value and the scale therefore reliable.

The German version of the TOWL-4 (Winkes & Schaller, 2022) was adapted to the video instruction mode. The instructional videos were drawn and animated in

MS Power Point (Microsoft, 2024) and edited in Camtasia (TechSmith, 2020). The practicability of the videos was tested within the pilot classes one year prior to the main study. The practicability of the assessment by videos was approved by all pilot teachers. The output of this preliminary work gave no reason to doubt the assessment method by video.

In addition to the original TOWL-4, the data of the total words written (TWW) were assessed to obtain the text length, called 'written composition productivity' by Yeung (2022), since the number of words produced in a text has been found to be a reliable and valid measure of overall writing ability (Troia et al., 2019).

Limitations

Two limitations seem notable at this point. Firstly, the German version of TOLW-4 has not yet been standardised. Therefore, neither a comparison of the sample with a norm sample is possible nor a statement about the intervention effect relative to a norm. Secondly, as illustrated above, the interrater reliability is only moderate. Despite the measures taken, these need to be considered.

5.4.2 Assessment of writing fluency – CBM-W

RQ 2A asks whether writing fluency measures differ significantly in students in EG 1 compared to EG 2/CG, whereas RQs 3A and 3B investigate possible changes in writing fluency within handwritten and dictated texts from EG 1.

The broader definition of writing fluency, discussed in Chapter 2, Section 2.1.3, is a more comprehensive measure of the transcription skills, i. e., lower-order writing skills. This refers to the amount of text written legibly and correctly in terms of spelling, grammar and semantics, within a certain period of time (Malecki & Jewell, 2003). The Subtest 6 of the TOWL-4 (see Section 5.4.1) comprises items in relation to lower-order writing skills, including items such as handwriting, spelling, punctuation or sentence construction (Hammill & Larsen, 2009). However, several measures of curriculum-based measurements – writing (CBM-W) also inform about different aspects of lower-order writing skills (Malecki & Jewell, 2003). Therefore, the most suitable variables to assess the RQs of interest must be determined. The following paragraphs delve into this question.

While the total words written (TWW) rates the productivity of writing but not the accuracy, it does not count as a value of the broad definition of writing fluency by itself. The words spelled correctly (WSC), the correct word sequences (CWS), and the incorrect word sequences (IWS), all inform about the accuracy analysed, according to different reference values. These measures always give an absolute measure which is dependent on the quantity written (Malecki & Jewell, 2003). Since WSC only considers orthography, CWS and IWS are considered more meaningful to answer RQ 2A because they provide further information such as about semantics or

punctuation. In addition, WSC is a measure that does not fit the analysis of dictated texts because the speech recognition technology only transforms spoken language into existing words, meaning that the STT output contains only correctly spelled words, unless one interferes by keystrokes. A measure that counts the correctness of a word on a word level only would, therefore, result in a meaningless outcome.

CBM-W also provides measures which are independent of the amount of text produced. The percentage of CWS (%CWS) or IWS (%IWS) from TWW provides such a relative measure, independent of the amount of text (Malecki & Jewell, 2003). Therefore, this value can give a wrong impression because the variable does not depict the amount of text produced; when, for instance, a student writes 10 out of 20 words correctly in the pre-test but 20 out of 40 in the post-test (Winkes & Schaller, 2022). For this thesis research, a relative measure is estimated of interest for RQ 3A, namely, to compare texts written in two different modalities which have an impact on the amount produced. However, for RQ 2A, the loss of the aspect of quantity, pointed out by A. Sturm and Schneider (2019) and Winkes and Schaller (2022), is considered a disadvantage for answering this RQ. One further measure explained by Malecki and Jewell (2003) is the correct minus incorrect word sequences, abbreviated as CIWS or CMIWS. This measure results by subtracting the IWS from the CWS and gives an accuracy measure that relates primarily to exact or correct writing.

According to the study conducted on German by Winkes and Schaller (2022), the more complex measures, such as CWS or CIWS, produce a higher alternate-form reliability from Grade 4 onwards than the simple measures, TWW and WSC. Using the median of three samples results in a few exceptions for these Grades in an alternate form reliability of $r > .80$. The criterion validity, calculated in these authors' study by the subtests of the translated TOWL-4 and a global teacher rating, matches the recommended Anglo-American threshold of at least moderate correlation (between $r = .40$ and $r = .70$) for the more complex measures from Grade 4. The criterion validity of CBM-W ratings in studies in English was acceptable, i. e., $r \geq .50$, for CWS and CIWS. Nevertheless, this comparison in German is based on informal assessment tools and must, therefore, be treated with caution. The scoring methods of TWW and WSC are not sufficiently valid for all analysed grades from Grades 3 to 6 (Winkes & Schaller, 2022).

Sensitivity to development is an important criterion for an assessment that is designed to measure change over time, especially for weak writers who progress more slowly than others (Winkes & Schaller, 2022). Despite the lack of standardised norms for writing fluency in German, Winkes and Schaller (2022) assume that CBM-W is a valuable assessment method to observe changes over time, which in their study is measured with two data collections in autumn and the following spring.

On the one hand, studies about these measures were not particularly conducted regarding students with dyslexia. On the other hand, there are advantages and disadvantages of certain measures. For this reason, this thesis research additionally examines the correlation of the collected data to select the most informative measure

for the data analyses. Therefore, in a first step, all available data about lower-order writing skills are explored in the written and dictated texts to take an informed decision. Available data stem from curriculum-based measurement – writing (CBM-W) and Subtest 6 from TOWL-4. In a second step, the extent to which the measures correlate with each other is calculated for handwritten texts (Table 5.7) and for dictated texts (Table 5.8).

Table 5.7: Correlation matrix (Person's r), spelling and writing fluency measures in handwritten texts

	TWW	WSC	CWS	IWS	CIWS	%CWS	%IWS	Sub_6
TWW (CBM)	—							
WSC	.99***	—						
CWS	**.97*** **	.99***	—					
IWS	**.9*** **	**.97*** **	**.95*** **	—				
CIWS	**-.40*** **	-.10	**.38*** **	**-.80*** **	—			
%CWS	-.03	**.23***	**.65*** **	**-.43*** **	**.82*** **	—		
%IWS	.08	-.16	**-.54*** **	**.54*** **	**-.85*** **	**-.86*** **	—	
Sub_6 (TOWL)	**.56*** **	.58**	**.59*** **	**.55*** **	**.44*** **	**.60*** **	**-.54*** **	—

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

Both values, the CWS and the Subtest 6 of TOWL-4, correlate significantly in the present study with all other values. CWS correlates particularly strongly with TWW ($r = .97$), with the WSC value ($r = .99$), and with IWS ($r = .95$). It also correlates strongly with the %CWS of TWW ($r = .65$), the %IWS, and the subtest 6 of TOWL-4. The correlation with the CIWS is still significant on a $p < .001$ level, with a moderate correlation. As discussed above, TWW did not produce a high enough criterion validity in the study by Winkes and Schaller (2022). For this reason and because it does not inform about writing fluency, TWW is not considered despite the high correlation. In addition, TWW of the TOWL-4 is already considered in writing productivity (see Section 5.4.1). WSC informs about spelling on a word level but does not comprise further information about morphosyntactic writing skills. Therefore, CWS is prioritised for the present study. The Subtest 6 of TOWL-4 likewise correlates significantly and with moderate to strong effect sizes with all other measures. Since the correlations are slightly lower, the CWS is selected as fluency measurement in this thesis research.

Svensson et al. (2019) write that specific assessment methods for dictated text need to be developed to compare dictated texts. However, CBM-W values are considered to allow rating dictating fluency and comparing different modalities. However, two points must be considered. Firstly, WSC would be meaningless, as discussed above. Secondly, the productivity and accuracy must be analysed separately, since the modality affects both speed and correctness. Again, the selection of the most appropriate measure results from the correlations between the measures (Table 5.8). The total words dictated (TWD) (blue in Table 5.8) correlate significantly and par-

ticularly strongly with CWS ($r = .95$) and IWS ($r = .89$) as well as significantly with a moderate effect with CIWS ($r = .44$). This value does not significantly correlate with the percentage values. Since the total number of words produced in a text has been found to be a reliable and valid measure of overall writing ability (Troia et al., 2019), this measure is included in the analysis of dictated texts.

Regarding the accuracy of the dictated texts without the amount, %CWS and %IWS correlate in an equally significant and strong way in the present study. %CWS (red in Table 5.8) is considered as having a more positive connotation and therefore included in the further analysis. However, %CWS is considered as an appropriate variable to compare texts produced in different modalities. Nevertheless, the disadvantage of a productivity-dependent variable is that the score does not illustrate variations of productivity. Furthermore, some previous studies only provide productivity-dependent variables such as CWS (green in Table 5.8). Thus, to obtain a broader picture of the development of accuracy in dictated texts, the three variables TWW, %CWS and CWS, are analysed.

Table 5.8: Correlation matrix (Person's r), writing fluency measures in dictated texts

	TWD	CWS	IWS	CIWS	%CWS	%IWS
TWD (CBM)	—					
CWS	**.95*** **	—				
IWS	**.89*** **	**.70*** **	—			
CIWS	**.44** **	.70***	-.02	—		
%CWS	-.04	.21	**-.37*** **	**.66*** **	—	
%IWS	.02	-.23	.37*	-.68***	**-.97*** **	—

Note: * p < .05, ** p < .01, *** p < .001

Procedure, instruction, and rating

The present research aims at keeping the strain for student participants, teachers and specialists as low as possible. To answer RQ 2A, no additional handwritten text was collected, with the CBM-W scores instead obtained from the same texts, i. e., TOWL-4 assessment. Therefore, the instruction for TOWL-4 Subtest 7 contained the additional element of marking the point where the students were after five minutes of writing. Only the first five minutes are analysed by the CBM-W rating. Meanwhile, for the dictating assessment, the students received a story starter such as, "This morning I woke up and ….". The students received one minute to think about this and five minutes to produce text following this beginning. The instruction (see Appendix D) was translated from Hosp et al. (2016) and incorporated in an instructional video. In sum, in both cases, the goal has been to obtain a sample of five minutes of continuous writing.

Malecki and Jewell (2003) report interrater reliability >.98 in ratings of research studies. Essential for the reliability of the rating are guidelines with descriptions or

examples (S. Graham et al., 2011) to maintain consistency over the whole analysing process and between pre- and post-test. Therefore, the rules for CBM-W-analysis from Hosp et al. (2016) were translated into German and adapted to language-specific features. Prior to the study, these rules for the analysis were tested on ten student texts and further specified for each ambiguous item (see Appendix E). One person only rated the CWS in this thesis research. Therefore, no interrater reliability is calculated.

Limitations

CBM-W proves to be more reliable and valid when using the median of three writing samples than when using a single writing sample (Winkes & Schaller, 2022). In the present study, only one sample of each modality at pre- and post-test was collected with regard to the workload for students, teachers and specialists. However, the single sample might diminish the reliability, limiting the quality of the fit of the assessment method.

The different assessment settings of CBM-W in the dictated and handwritten task might be a limitation for the comparison of dictated and handwritten texts. However, since both settings support five minutes of continuous writing and because the content is not assessed by CBM-W rating, this difference is considered acceptable to answer the RQs. Furthermore, the %CWS in the comparison does not comprise the amount of text and therefore should not affect the results heavily. In addition, the development of the productivity over time compares the modality specific assessed by the same setting.

5.4.3 Assessment of reading fluency – SLS 2–9

Firstly, it was hypothesised, in Section 5.3, RQ 2B, that the intervention could affect writing competences as well as the related reading competences (Linnemann et al., 2022). Secondly, it was argued that the time on task could impact the basic reading skills. Therefore, students possibly re-read more on the screen compared to their re-reading on paper, because the text is more legible and the introduced dictating procedure demands revision using TTS. Reading competence contains several subcomponents: reading speed, reading comprehension, word reading and pseudoword reading (Winkes, 2014). RQ 2B focusses on reading fluency, i. e., accuracy and speed. Consequently, this thesis study requires an assessment tool measuring reading fluency. The following discussion, in Section 5.4.3.1, presents first the selected assessment tool, considered meaningful for recognising any changes and to be fitting for the study design. Then, the advantages and limitations of the tool for this study are discussed. Section 5.4.3.2 outlines the standardised assessment by instructional videos.

5.4.3.1 Salzburger Lesescreening 2–9 (SLS 2–9)

The 'Salzburger Lesescreening' 2–9 (SLS 2–9), according to the authors Wimmer and Mayringer (2014), tests the basic reading skills in German, providing a parallel test version. The students are asked to read a sequence of relatively simple sentences as quickly as possible, and to mark at the end of each line whether the statement in the sentence is correct or incorrect. The entire test, including instructions and preparation of the materials, takes less than 15 minutes in an individual setting. The actual reading time is limited to 30 seconds' reading of exemplary sentences, followed by 3 minutes of effective test reading time.

According to the test construction, the aim of the test items is to ensure that the knowledge requirements for judging the sentences as correct or incorrect should not cause any problems. In addition, the wording is chosen with care to ensure that the judgement is as clear as possible. The most significant variation in the sentences in the present study was the length. Care was also taken in the formulation of the sentences to be judged as false, to ensure that the critical information was only given towards the end of the sentence. The test has been standardised from the second to the ninth school year. Wimmer and Mayringer (2014) recognise objections that the assessment of sentences could involve more than word processing, for example linguistic sentence processing. However, based on the test construction and several studies on the quality criteria, these authors assume that the SLS 2–9 reflects basic reading skills and is closely related to reading speed. Due to differences between boys and girls in the norm sample of the test, the SLS 2–9 manual provides separated norms for both genders. The SLS 2–9 test results are assessed using a reading quotient. Like an intelligence quotient, the reading quotient expresses how far the measured reading skill deviates from the average of the norm sample. According to Wimmer and Mayringer (2014), the reading quotient has two advantages compared to a percentage rank. Firstly, it is less affected by the composition of the respective sample at the edges of the performance distribution (e. g. with very low values). Secondly, it is more sensitive for differentiating between very poor test performances. The SLS 2–9 has a parallel test reliability of .87 to .95, depending on the grade level. The correlations between the SLS 2–9, form A (to be applied in the present thesis research) and word reading speed are high, with .89 for Grade 2 and moderate with .55 for Grade 8. This decrease in the course of school years is explained by the articulation in the read aloud testing which limits speed more strongly in older and more competent readers. By the authors, reliability and validity were judged to be satisfactory. The test is standardised in Austria for each Grade from Grades 2 to 9. The applicability of this standardisation in the entire German-speaking region is supported by the fact that only minor differences were found between Austria, Germany and Switzerland in repeated comparative international studies on reading literacy, PISA 2000–2012 (Wimmer & Mayringer, 2014).

Three main advantages are decisive for the choice of this tool for the present study. Firstly, the SLS 2–9 is an economical method for assessing basic reading skills, which is important in research in order to minimise the constraints for participants and professionals. Secondly, the tool is sensitive to differences without floor and ceiling effects. While not all students with dyslexia are weak readers, many of the results of the present sample are expected to be in the lower part of the range. A test without floor effect is therefore important. Thirdly, the test can be repeated at relatively short intervals due to the available parallel forms (Wimmer & Mayringer, 2014), which is required for the pre-/post-test design of the present study.

Since the other above presented assessment tools do not provide norm values in German, the norms were considered to be another advantage. However, this advantage proved to be in vain, since some raw values of the present sample were below the norms provided. Therefore, the raw values were used in the statistical data analyses (see Chapter 6, Section 6.1.3.3).

Limitation

A disadvantage of the SLS 2–9 might be the lack of research literature found on the applicability of the STL 2–9 to detect progress in intervention studies.

5.4.3.2 Standardised assessment by instructional video

As mentioned above and already elaborated for the administration of the TOWL-4, a video instruction was also developed for the SLS 2–9. The same instructional video was applied at both pre- and post-test. It lasted for eleven minutes. The video presents all the necessary materials, i. e., the test booklet and the two pens, and leads step by step through the instructions for Grades 5 to 9 (Wimmer & Mayringer, 2014, pp. 29–32). The 30-second exemplary sentences and the 3-minute test reading time are timed within the video. Thus, the specialists only had to run the video. Therefore, all participants received the identical instruction.

The speech pathologists and the special needs teachers are familiar with the application of diagnostical tests and aware that the test instructions must be precisely respected. The link was sent to the specialists by e-mail. No specialist reported any problems running the video.

5.4.4 Assessment of writing motivation – Translation & adaptation of SWAS

The importance of writing motivation and the relationship to writing achievement has been elaborated in Chapter 3, Section 3.3.2. Ekholm et al. (2018) and Camacho et al. (2021) recommend, for research and assessment of writing motivation, firstly, to give an explicit definition of the theoretical background, and secondly, to clarify

assumed concepts and base them on previous empirical studies and relevant theories of motivation. As elaborated in Section 3.3.2, the present thesis aligns with the model of writing motivation by Wright et al. (2019) and elaborates here the selection (see Sections 5.4.4.1 and 5.4.4.2), translation and adaptation (see Section 5.4.4.3), and fit of an assessment tool (see Sections 5.4.4.4). Section 5.4.4.5 gives limitations of the applied tool to answer RQ 4; and Section 5.4.4.6 explains how the impact of the STT tool on writing motivation, is addressed among students of EG 1.

5.4.4.1 Considerations for the selection of an assessment tool

Writing motivation, like motivation in general, is domain- and task-specific (Bandura, 1997; Bruning & Horn, 2000), and a tool should therefore be adjusted to the study and its RQs. Ekholm et al. (2018) estimate that research related to specific genres could lead to better predictive utility of writing motivation measures than more domain-general studies. According to Troia et al. (2012), the assessment of writing motivation should include multiple measures of motivation that represent discrete constructs within this multi-faceted construct (Troia et al., 2012). Furthermore, the context of writing influences writing motivation (S. Graham, 2018). Therefore, writing modality and the use of STT should be captured.

Various methods, ranging from self-report scales to interviews, document analysis and focus groups, allow research to explore writing motivation (Camacho et al., 2021). Steiner and Benesch (2021) write that questionnaires are often used to explore opinions, attitudes and positions on issues or facts. Camacho et al. (2021) specify, for writing motivation, that studies rely heavily on self-reported or teacher-reported scales. For this thesis study, instruments for a self-reported survey are therefore assumed to be more likely to be available and appear to be, for practical reasons of the global research design, most adapted to the present research goal. However, the issues and possible biases of a self-report survey must be considered.

At the time of the preliminary work on the study, an adequate instrument in German for assessing writing motivation did not exist (Hennes et al., 2023). Wright et al. (2019) studied the literature on existing English writing motivation assessment tools for adolescent students and concludes that existing instruments contain important limitations for measuring writing motivation, due to the lack of multidimensionality, low reliability, or validity scores. Thus, these authors developed, in a multistep approach, the 'Self-Beliefs, Writing-Beliefs, and Attitude Survey', abbreviated as 'SWAS'.

5.4.4.2 Fit and selection of SWAS to assess writing motivation

Wright et al. (2019) based their approach to assess writing motivation on the above-mentioned review of existing questionnaires on both writing motivation theory and knowledge of reading motivation. Then, they tested the questionnaire,

firstly, with an EFA on a sufficiently large sample. In the next step, they confirmed the resulting factor structure with a confirmatory factor analysis on a second, sufficiently large sample. In sum, the resulting SWAS consists of a multidimensional questionnaire developed in a multi-stage process. In the study conducted by Wright et al. (2019), the overall reliability coefficients of the two factor analyses (α .936 and .943) indicate that the SWAS produces reliable scores for the various factors and the students' writing motivation. The reliability scores for the isolated factors range from Cronbach's α 0.697 to 0.885. The lowest score for self-efficacy is explained by the fewest number of items and is estimated to be equal or higher than previous research (Wright et al., 2019). External validity, established by the correlation of observed measures of writing motivation, i. e., teachers' perceptions, and direct measures of writing skills, i. e., TOWL-4, shows statistically significant but modest correlations, explained by the small validation sample ($n = 53$).

The validated questionnaire disposes of two main advantages for the present study. Firstly, the questionnaire has a theoretical foundation that defines the construct of writing motivation. Secondly, it is empirically established by analysing and communicating the quality criteria of reliability and validity. Further advantages of the SWAS might be that it also aims to investigate writing motivation of academic writing; and, it was validated on the text genre of narrative texts, and also the TOWL-4 which is also selected for the present study. However, for the present research, there are shortcomings in the use of the SWAS that must be considered:

1) The English survey needs to be translated into German and adapted to a sample in the present study that tends to show lower reading skills, due to the younger age and the diagnosis of dyslexia. Therefore, the resulting survey instrument must be assessed on whether it measures what it claims to measure (Steiner & Benesch, 2021).
2) This questionnaire does not provide any insight into the influence of the writing tool on writing motivation, as discussed in Chapter 3, Section 3.3.2.
3) S. Graham et al. (2023, p. 3) criticise that the SWAS "did not address students' motives for writing". Causal clauses, such as 'I write because it makes me feel better', inquire about students' writing motives.

Nonetheless, for the purpose of this thesis study, the SWAS was considered the best option in the context of this research and therefore selected.

5.4.4.3 Process of translation and adaptation of the survey

The aim of the writing motivation assessment is to measure changes due to the STT intervention. The study sample of Grade 5 students with dyslexia is the target group of this survey. However, the questionnaire could possibly be used for a broader target group in terms of age or for educational purposes. According to Bühner (2021), a strategy based on the above presented aim is necessary for the construction of a

questionnaire. For the present thesis research with an existing questionnaire, this does not concern the construction but the translation and adaptation of the SWAS.

Translation and adaptation strategy

Literature recommends a forward-backward translation process (Beaton et al., 2000). However, the aim of the present study is not for the translation to be as close as possible but a translation with an adaptation in terms of language complexity to the target group. The target group is younger than the sample in the study by Wright et al. (2019) and has been diagnosed with dyslexia earlier. Therefore, the present group is considered to have lower reading skills than the original, Grade 8 sample. Instead of the forward-backward translation, an expert panel of speech pathologists with extensive experience of working with children with dyslexia was installed. In a first step, the SWAS was translated by a speech pathologist with a focus on language difficulties. Regarding sentence structure, care was taken to ensure questions were as short as possible and with as few subordinate clauses as possible. Meanwhile, high-frequency vocabulary was chosen where possible, and specific vocabulary kept to a minimum. The first version was then responded to by one primary school student of the age group without dyslexia. First adaptations were made (e. g. 'klug' was replaced by 'intelligent'). The questionnaire was then discussed with a second speech pathologist to focus on linguistic difficulties, with the aim of running a pilot, as discussed next, a necessary step for all questionnaires (Fink, 2013).

Pilot version and adaptation according to observation of responding students

For the present research, five children with dyslexia and five children without dyslexia, aged between 10 and 12, were selected as the relevant pilot sample, as presented in Table 5.9.

Table 5.9: Demographics of questionnaire pilot sample

dyslexia	gender	grade-level	multilingual
Yes 5	Female 4	End of 4th grade: 5	Yes 7
No 5	Male 6	End of 5th grade: 4	No 3
		End of 6th grade: 1	

The children's answers, questions and comments were collected and used to optimise the questionnaire. Due to translation and comprehension difficulties, one question was dropped completely: *I am confident in writing for many purposes (persuade, inform, entertain, or express)*. Minor adjustments were made to some further questions. The translated and adapted questionnaire consisted of 29 questions which are scored on a 4-point scale, as in the original survey by Wright et al. (2019). The version for the present study was finalised after formatting adjustments based on feedback from a speech pathologist (Appendix F).

Exploratory factor analyses and analyses of reliability

The translated and adapted questionnaire was administered to the present study sample before the software introduction for EG 1, at pre-test for EG 2 and CG. The timing of the motivation assessment in EG 1 is chosen to minimise bias. Both the introduction to the software and natural changes over time could have an influence on writing motivation. The introduction of the software is assumed to have had a greater bias than the approximately 6 school weeks, which is why EG 1 answered the questionnaire before the introduction of the software and not during the pretest like EG 2 and CG.

The data of the pre-test assessment were then used to confirm the factor structure of the translated and adapted scale. As a given factor structure was used as a basis, no overall EFA was carried out for this thesis research. Instead, it was examined whether the items for each of the four factors from the original study (Section 3.3.2) load on the same factors. The items grouped into one factor all showed significant Bartlett test results ($p < .001$), and all Kaiser-Meyer-Olkin coefficients were higher than the required .50. These conditions were therefore met. However, the condition of a sufficient sample size for exploratory factor models was not met in the sample of students with dyslexia ($N = 108$) (Bühner, 2021). Therefore, a larger sample of $N = 529$ was used. This larger sample, consisting of the study sample and a further 421 classmates without dyslexia, was used to run the analysis to achieve a more stable factor structure (Flütsch Keravec, forthcoming). With a single exception, the first EFAs showed a structure of two factors with eigenvalues greater than 1.0 Further analyses were run excluding one or two items per factor conducting EFAs – according to visual analysis of the SPSS output and assumptions on difficulties. Table 5.10 shows the results for both samples, presenting the explained variance and reliability scores after the exclusion of the items (see grey-framed).

Table 5.10: Values of variance and reliability of factor analyses

	N 108 with dyslexia			**N 529 (108 with & 421 without dyslexia)**		
	Excluded items	variance	Reliability Cronbach's α	Excluded items	variance	Reliability Cronbach's α
Self-efficacy	27 & 28	47.825	.633 (4)[1]	27	40.631	.620 (5)
Self-concept	18	44.196	.741 (6)	18	49.037	.784 (6)
Beliefs about writing	10	46.642	.765 (6)	-	41.720	.760 (7)
Attitudes tw. writing	2 & 6	47.610	.806 (7)	2 & 6	50.449	.824 (7)

1 number of items

Therefore, the result of the combined sample, highlighted by the grey background, was chosen. In respect to the reliability and the aim of excluding a minimum of

items, the model based on the larger sample is considered to fit better. The disadvantage of this procedure might be that the difficulties of the group with dyslexia with Items 10 and 28 are not considered. However, visual inspections of the scree plots in the smaller sample of *N* 108, in Figures 5.5 and 5.6, show that for both questionable motivation components 'self-efficacy' and 'beliefs about writing', the eigenvalues are close to eigenvalue 1.0. The model for the larger sample is therefore considered tolerable for the students with dyslexia.

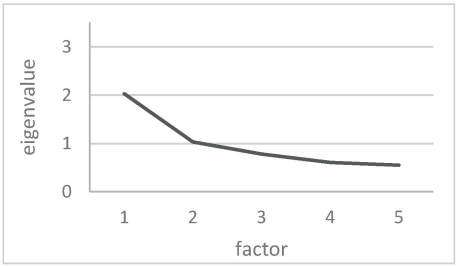


Figure 5.5: Self-efficacy

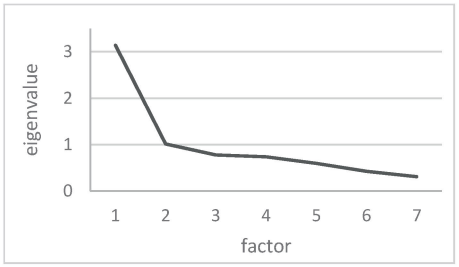


Figure 5.6: Beliefs about writing

Excluded items

The four excluded items are Items 2, 6, 18, and 27:

Item 2: Ich überarbeite meine Texte nicht gerne.
Item 6: Ich möchte in der Schule weniger schreiben.
Item 18: Ich bekomme keine guten Noten fürs Schreiben, weil ich einfach nicht intelligent genug bin.
Item 27: Wenn ich für einen Text keine gute Note bekomme, dann deshalb, weil ich Glück habe.

It is assumed at this point that the students had problems understanding: 1) negatively worded items; 2) the concept of marks[44] for texts; and 3) the subordinate clauses. Regarding the negatively worded items, this concerned all four excluded items.

5.4.4.4 Limitations of writing motivation assessment in this thesis research

Writing motivation, in Chapter 3, Section 3.3.2, presented as a multi-dimensional construct, is investigated by the RQ 4 as the sum of the single components according to the model of Wright et al. (2019). The investigation of the individual parts is beyond the scope of this thesis but is considered important to obtain a more accurate picture of writing motivation and its development among students with dyslexia.

Two limitations regarding the translated and adapted version of the SWAS must be discussed. The first limitation concerns Cronbach's α coefficients, informing

44 According to the specialists and teachers, some of the students in the sample have never received any marks for texts.

about the internal consistency. The coefficients range between .620 and .806 (Table 5.10). The value of .620 for the concept of 'self-efficacy' is rather low. However, this is in line with the original SWAS (Chapter 3, Section 3.3.2.2). Two reasons are considered in explaining the low consistency. Firstly, just as in the original study by Wright et al. (2019), the number of items is low and, due to the exclusion of one item, further reduced. Secondly, self-efficacy is the most task-specific construct of writing motivation. Regarding the present study, firstly, a highly task-specific assessment of writing motivation could be problematic, since the writing instruction varies between the CG and both experimental groups. Therefore, the assessment tool must, to some extent, assess general academic writing motivation. Secondly, the writing motivation survey assesses the writing motivation of all participants, using different writing tools. Therefore, the present tool cannot investigate STT as a specific writing medium unfamiliar to two groups.

The second limitation concerns the order of the questions, which was in the translated version inadvertently the same as the list of questions published in the article by Wright et al. (2019). However, the original SWAS, as presented to the students, had a different order. At this point this is not considered a problem. According to Bühner (2021), the order is relevant when ordered by difficulty, which does not apply here. However, in a further version, the order would be aligned with the original and the items for self-efficacy would be further elaborated.

5.4.4.5 Additional questions on tool predilections for the EG 1

The shortcoming of the present assessment design, which does not assess tool-specific motivation and/or emotions, must be compensated for. Therefore, the EG 1-students were asked to answer three additional questions about their feelings while using the STT tool. The three questions presented in Table 5.11 were asked using a MS Forms online survey:

Table 5.11: Questions asked to EG 1-students

	Original questions in German	**English translation**
1.	Welche Aussage passt am besten? *-Ich schreibe Texte lieber von Hand.* *-Ich diktiere Texte lieber.* *-Diktieren und von Hand schreiben mache ich etwa gleich gerne.*	Which statement fits best? *-I prefer to write texts by hand.* *-I prefer to dictate texts.* *-I like to dictate and write by hand about equally well*
2.	Ich denke, dass ich in Zukunft … *… viel diktieren werde.* *… manchmal diktieren werde.* *… selten diktieren werde.* *… nie diktieren werde.*	I think that in the future … … I will dictate a lot. … I will dictate sometimes. … I will rarely dictate. … I will never dictate.
3.	Wie fühlst du dich beim Diktieren?	How do you feel while dictating?
4.	Freiwilliger Zusatz: *Was möchtest du Silvana Flütsch noch sagen?*	Optional: *What else would you like to tell Silvana Flütsch?*

While Questions 1 and 2 are multiple-choices, Question 3 is open-ended. In the same online form, i. e., in Question 4, the participants had the option to leave a comment or remark to the researcher.

For reasons mentioned above, these questions and answers are not part of the writing motivation scale. Therefore, they are included in the qualitative analyses and triangulated with the quantitative results. The results of the multiple-choice questions are, on the one hand, used as factors to analyse RQ 3B; on the other hand, the results are analysed by the qualitative content analysis, using the closed-ended answers as factors to analyse the open-ended answers.

5.4.5 Preparation of qualitative data collection – interview guide and code system

5.4.5.1 Construction of interview guide

The interview guide is constructed on the basis of the manual for conducting qualitative interviews by Helfferich (2011) and her proposed four-step procedure[45] to elaborate questions: collecting, checking, sorting, and subsuming.

The 'collection of questions' means that the researcher gathers a maximum of questions and keywords in the context of the research topic of interest. All concerns regarding suitability are put aside (Helfferich, 2011). For the present study construction of the interview guide, the questions address teachers' and specialists' opinions and attitudes about benefits, challenges, and the feasibility of writing with STT in school for the group of students with dyslexia. Collected questions also intend to evocate details about characteristics and features in the students and the contexts that support or hinder the use of STT. As elaborated in Section 5.3.5.1, the interview guide seeks to give space to unexpected narratives and considerations.

The second step of 'checking the questions' includes a reduction and structuration of the questions. This can be done with review questions such as, "Is the question going to be answered automatically when one talks about the topic?" or "Is the question suitable for generating open answers and narratives?". These first two steps help to elaborate an inventory but also to reflect on the prior knowledge – possibly diffuse – and about assumptions that come from literature or one's own field experience. This procedure stimulates the awareness that the interviewees' narrative could deviate from the researcher's expectations (Helfferich, 2011, 182 cont.).

The interview guide should comprise between one and four bundles of questions. Therefore, the third step of 'sorting' allows the research to create these bundles according to content, depending on the research interest. Some questions may remain as individual questions, receiving a separate place in the guide later (Helfferich,

45 'SPSS bei der Leitfadenerstellung': Sammeln- Prüfen – Sortieren – Subsumieren; Helfferich (2011).

2011). The interview guide for the present study, formulated in German, comprises four bundles (Appendix G):

- Opening, collection of challenges and difficulties of STT in the classroom or in a single setting (speech/language therapy (SLT) or special education)
- Motives/attitude of the teachers and specialist
- Behaviour of the child and handling when using the device/software in the classroom/in SLT.
- Writing processes/writing quality and writing motivation when dictating from the teacher's or specialist's perspective

In the last step, 'subsuming', the aim is to find one narrative prompt for each bundle. This stimulus should be as simple as possible and contain as few presuppositions as possible. These prompts must be tested on their suitability for evoking as many aspects of interest as possible. The researcher enters the narrative prompts in the first column of the interview guide, organised as a table. The researcher then transforms the aspects and questions assigned to this bundle either into keywords, in the second column, or pre-formulates questions, in the third column. The second column serves for checking whether the content has already been addressed. The third column contains the questions that are mandatory and should be asked using the predefined formulation. The fourth column adds suggestions for maintenance questions or questions to motivate the interviewee to continue and/or to show interest and understanding in relatively short narrative passages (Helfferich, 2011).

The development of the interview guide by one researcher in this thesis study was then discussed and revised with a researcher with experience in conducting interviews. In a next step, the interview guide was tested with a speech pathologist by means of a fictional interview. Subsequently, the interview guide was applied within the preliminary studies with two speech pathologists and four class teachers. Each time, the suitability of the interview guide was reflected on and improved when shortcomings were recognised. Different versions of the interview guide were prepared, depending on the discipline of the interviewee and whether the interviewer and interviewee were on first-name terms or not (see Appendix G for an exemplary interview guide).

5.4.5.2 Development of code system

The analysis and interpretation of the interviews is based on the fundamentals and techniques by Mayring (2022). The flowchart in Figure 5.7 illustrates the main steps of the development, and the differentiation of the category system with its codes, which are a central element of the qualitative content analysis and presented in the following paragraphs.

The literature reviews in Chapters 2 to 4 suggest that the application of STT harbours certain difficulties. Therefore, the aim of RQ 5 is to explore possible challenges

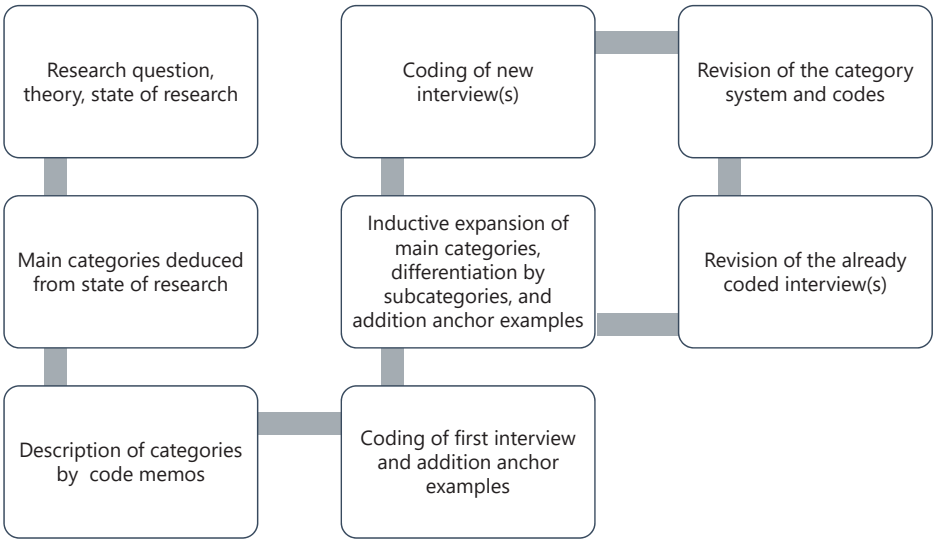


Figure 5.7: Evaluation model of the interview analysis on STT in the classroom (own representation)

and facilitators in the use of STT (see Section 5.3.5). This implies that there are assumptions about challenges and facilitators, although RQ 5 is primarily explorative. Therefore, a first set of categories is deduced, on the one hand, from the current state of the literature and, on the other hand, from the preliminary investigations and interviews. However, these first categories are neither exhaustive nor in depth. Therefore, the first deduced categories must be complemented and differentiated, which is done inductively from the original linguistic material of the interviews.

The procedure with an initial set of categories can be criticised, if an orientation towards categories can hinder the understanding of the material. However, for the present supplemental research component, working with a category system supports the comparability of the results and the assessment of the reliability of the analysis (Mayring, 2022).

The first set of main categories, deduced form the state of research and from spontaneous feedback received during the intervention study, are inspired by certain considerations that were already present in the development of the interview guide. Therefore, some similarities are apparent between the interview guide and the code system: 1) motives and attitude of the teacher and/or specialist; 2) challenges and facilitators in the writing classroom; 3) challenges and facilitators in relation to the software; 4) challenges and facilitators depending on the software introduction; and 5) challenges and facilitators in the handling and/or behaviour during dictation. The first interview was then analysed with these five basic categories. At this point, the approach was modified to an inductive process. Firstly, the categories were expanded with additional main categories, e. g. supportive adaptations, features and resources. Secondly, the categories were refined with up to two further levels of subcategories. Anchor examples were added to the code system. In a circular process, new inter-

views were then coded and necessary adjustments applied to the code system. The whole process was carried out by the thesis researcher. Changes to the code system and the scoring approach were recorded in the MAXQDA logbook.

The resulting code system (see Appendix H) contains seven main categories that differ significantly from the categories deduced at the beginning: 1) positive experiences; 2) challenges of STT in SLT or school; 3) supportive adaptations, features, and resources; 4) handling of dictating and STT; 5) motives and attitudes among teachers and specialists; 6) experiences with STT introduction in single setting; and 7) experiences with video writing instruction in writing classes.

5.4.6 Summarised overview of the survey instruments

Table 5.12 summarises the assessment tools used in this study and the scores collected for the data analysis.

Table 5.12: Overview of assessment tools, analysed scores, and settings

Component of interest	Assessment tool	Analysed score(s)	Assessment setting
Text quality	TOWL-4; Subtest 7	Score subtest 7	Classroom setting (EG 1 & EG 2), single setting CG
Text length	TOWL-4	Total words written (15 min)	
Writing fluency	TOWL-4 + CBM	CWS of 5 min TOWL-4	Classroom setting (EG 1 & EG 2), single setting CG
-accuracy		%CWS/TWW of CBM (dictated texts only)	Single setting for EG 1 only
-speed			
Reading fluency	SLS 2–9	Raw score (total sentences read minus errors and omissions)	Single setting all groups
Writing motivation	Translation of SWAS	Sum of means of all components -attitude towards writing -beliefs about writing -self-concept (-self-efficacy)	Classroom setting (EG 1 & EG 2), single setting CG
Challenges for application in school	Semi-structured interviews	-Narratives -Code system -Code frequency	Onsite and online interviews with teachers and specialists of EG 1
Additional questions to EG 1	Open ended and closed ended questions	-Percentages -Narratives	MS forms online questionnaire

5.5 Intervention – Equipment, instructional, and teaching materials

The present study, on the STT-intervention for EG 1, on the one hand, and the writing instruction by videos for EG 1 and EG 2, on the other hand, requires study equipment and instructional material. Instructional materials are needed for the software instruction, the writing instruction by videos, and the instructions for pre- and

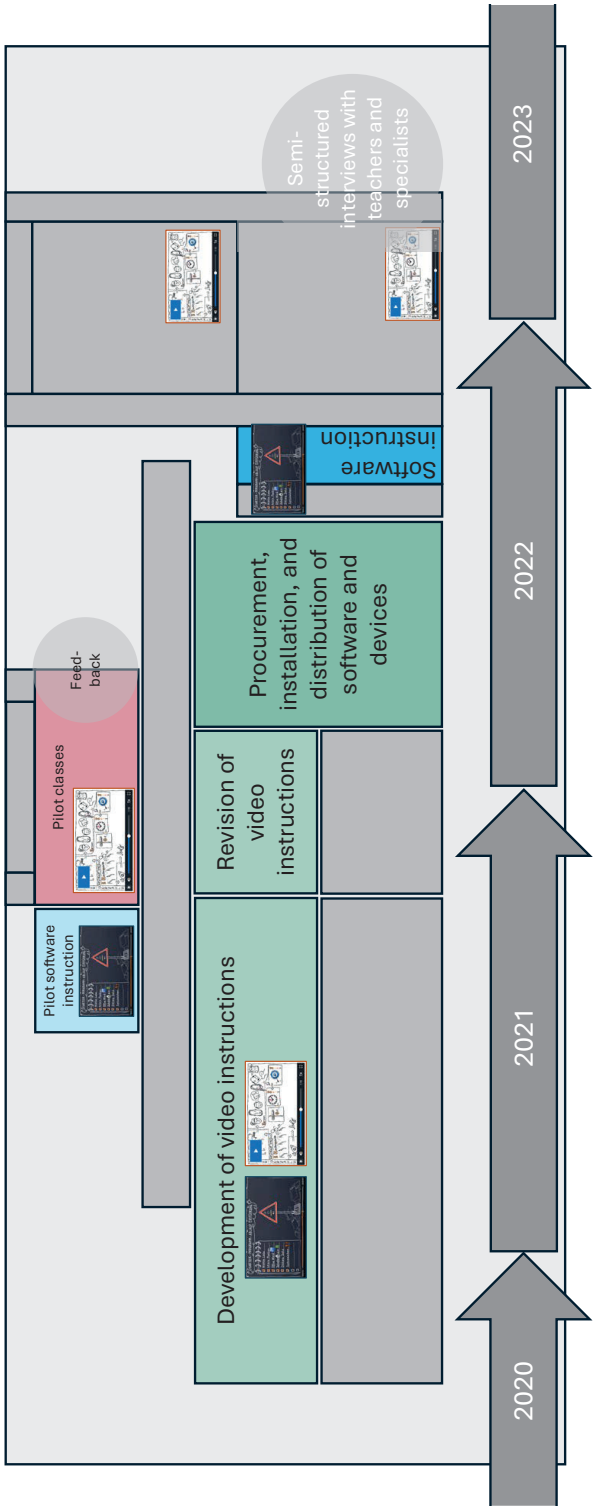


Figure 5.8: Development and procurement of study materials

post-test. Figure 5.8 illustrates the preliminary work in the study timeline. Section 5.4.1 presents all hardware and software used in this study. Section 5.4.2 outlines the main content of the software instruction. Section 5.4.3 presents the essentials of the writing lessons by instructional videos.

5.5.1 Selection of hardware and software

The hardware equipment consists of a Microsoft Surface Go tablet, a keyboard, and a headset.

The Microsoft Surface Go tablets – Model 2 for the pilot study and Model 3 for the main study – have 4GB random-access memory (RAM) and 64 GB internal storage (ROM, read- only memory). These models were selected for the pilot study due to their price-performance ratio, which matched the study budget. The models proved suitable for the pilot study and worked well with the tested software. These devices were therefore purchased for the main study.

Further study equipment includes the Microsoft Surface Go Type Cover keyboard and a Poly Blackwire 3215 headset. The keyboard is used for the purpose of revision, which is challenging with voice commands only, and provides the option of setting punctuation marks manually. Mono headsets were considered more adapted to the classroom situation after the pilot trials, because the student must still hear a teacher or specialist talking. These headsets are plugged into the 3.5 mm jack. An additional USB-C to USB-A adapter allows the storage or exchange of documents.

Prior to the study, the software Nuance Dragon Home 15[46] was purchased and installed on the tablet. This software was chosen for four reasons: 1) it runs without internet connection, which was considered important since certain school web accesses do not allow external devices; 2) the university's legal service authorised the use of this software; 3) this software is designed to adapt to the speaker over time; 4) the majority of studies reviewed by Liu et al. (2019) used an edition of Dragon Naturally Speaking software.

To avoid using any personal data of the participants, each tablet was equipped with a fictitious Office 365 account. These accounts were kindly provided for study purposes by Fritic[47]. 12 participants had access to a personal computer provided by their school and were therefore only equipped with the software, the headset, and the USB-C adaptor. In agreement with the persons responsible for ICT at the respective schools, these participants used their own Office 365 school account.

46 Dragon Home 15 was discontinued in 2024; see Nuance Communications (2024).

47 Fritic: Centre of Competence of the Canton of Fribourg, for all aspects of media as well as information and communication technologies (ICT) in the classroom; see (*Centre De Compétences Fritic*, 2024).

5.5.2 STT-instruction in individual setting for EG 1

The review of the current state of research presented in Chapter 4, Section 4.3.2 emphasised the importance of a thorough introduction to the software (Liu et al., 2019). MacArthur and Cavalier (2004) describe that a research assistant in their study explained the system, demonstrated dictation, and guided participants through the initial training. Liu et al. (2019) summarise that most studies include such an introduction but state that the number and duration of these sessions vary greatly. Although these latter authors do not indicate the time or quantity necessary and do not give specifications about the content, they recommend that students receive training in the use of their specific STT technology. The review in Chapter 4 also emphasised possible challenges due to the shift from written to oral text production which must be addressed. Mossige, Almgren Bäck, et al. presented, in 2023, a study protocol which indicates a precise sequence for the dictating procedure.

The present study introduces the use of the software in a comparable way to Mossige, Arendal, et al. (2023), although it was not possible to refer to the Scandinavian research team's publication when preparing the introduction in 2021–2022. In contrast to other studies, the present thesis research provides the software instructions by video instructions because the specialist for each participant executes the introduction in the field. This means that the principal researcher developed and distributed the videos and was available to support the specialist in case of questions and uncertainties. Therefore, the mentioned videos were intended to introduce the participant to the use of STT and, simultaneously, transmit the necessary knowledge to the executing specialist. The latter is important because, according to (Galuschka et al., 2014), interventions provided by a study author with solid knowledge about the topic and the interventions have been shown to be more effective than those by other conductors. It is assumed that this knowledge impacts the effectiveness of an intervention (Galuschka et al., 2014).

The revised introduction, after the preliminary trials, comprised two parts: 1) the set up and training of the system; and 2) the practice of dictating with a proposed dictating procedure. In total, six therapy sessions of 30 to 40 minutes were foreseen for this introduction, including dictation practice. The aim of this introduction was to be nearly autonomous in dictating texts in the writing classes at the end of the introduction period.

The first session introduced participants to the set-up and the initial training on Nuance Dragon Home 15, which is targeted at adult users. Therefore, the content was reduced to its most important and presented in a linguistically simplified form. Although the participants could use voice commands for punctuation, the introduction stipulated that punctuation marks are set using the keyboard. Voice commands for revision were not introduced. The following four introductory sessions were structured as follows:

- The second session aimed at getting used to starting the software, Dragon Home, and to open a Microsoft Word document, as well as getting used to the headset and switching the microphone on and off. The dictation was limited to the word level.
- The third session addressed the dictating procedure. The introduction advised the students to: 1) think about the sentence; 2) dictate the sentence and set the punctuation mark; and 3) listen to the sentence via TTS and revise the sentence. During this session, the students dictated predefined sentences.
- The fourth session aimed at automating the dictating and revision procedure introduced before. In contrast to the previous session, the students would try to invent new sentences.
- The fifth and sixth session still targeted the automatisation of the procedure but moved on to more extensive linguistic input. Therefore, story starters were provided to dictate consecutive sentences.
- During the sixth session, the pre-test assessment took place and was instructed on by a separate video. Therefore, a five-minute probe with a story starter was timed and collected.

The first version of the videos for the introduction was tested by three speech pathologists and their students with dyslexia during the school year 2021/2022. The participating students were in Grades 4 and 5. The specialists took notes and gave oral feedback to the researcher. According to the feedback, the researcher revised all videos. The revision addressed two major issues: firstly, the linguistic complexity was reduced through shorter sentences and reduced speaking pace; secondly, the reading of text passages in the original second session was removed, as the accuracy was too low.

In sum, the procedure is similar to Mossige, Almgren Bäck, et al.'s (2023) procedure. However, it does not include the oral rehearsal assumed to be helpful by those authors. This could possibly be a limitation of the instruction in the present study. Another limitation of the instruction could be that the word level at the beginning of the introduction might compromise accuracy, due to the speech recognition system that is generally more accurate in a larger context.

5.5.3 Writing intervention for EG 1 and EG 2

One challenge of this specific quasi-experimental research design for students with dyslexia is that they come from many different places and classes. It was argued, in Chapter 4, Section 4.3.3, that the provision of effective writing instruction is one condition for the effective use of STT. The participants are enrolled in cantons with the curriculum, "Lehrplan 21" (D-EDK, 2024), and used, up to the start of the intervention, one of the two most common teaching materials authorised for German language arts classes, i. e., "Die Sprachstarken" or "Sprachland" (Interkantonale

Lehrmittelzentrale, 2024). However, comprehensive empirical studies that provide an insight into the actual use of teaching materials in lessons and the effect on students' learning performance are rare, to date (Totter et al., 2024); nor are data available about what exactly is taught in the writing classroom or how much time is spent on specific writing instruction – with focus on text writing in distinction from spelling – in German-speaking Switzerland. Therefore, teaching materials for the present study were specifically developed with the aim of providing evidence-based writing instruction and assuring the comparability for all participating classes. The application of the provided materials during the 18-week intervention period was mandatory for the participation in both experimental groups. An additional aim was to keep the teacher's workload, of preparing writing instruction, to a minimum.

Various considerations outlined as follows resulted in the selection of the medium of video instruction, and in the selection of a teachers' manual serving as the basis for the instructional materials. This latter was, namely, the "Aufsatztraining für 4. bis 6. Klassen: Ein Lehrermanual mit Unterrichtsmaterialen zur Förderung von Schreibkompetenz und Arbeitsverhalten" by Glaser and Palm (2014), meaning "Narrative writing training: Teachers' manual for 4th to 6th grade fostering writing competence and working behaviour" (translated Flütsch Keravec).

The scope of the present research does not allow to thoroughly elaborate learning with instructional videos. The present thesis research considers that video instruction as a learning tool is sufficiently evidence-based (Brame, 2016), although the benefits for learning of a system-paced presentation mode with a predefined presentation speed (Koning et al., 2011) and therefore the issue of too high or too low mental demands (Kerres, 2018) could be questioned. However, the visually and audibly conveyed information must meet the implications for instructional design described by R. E. Mayer and Fiorella (2014, p. 307) such as the signalling principle, saying, "the multimedia message should contain cues that emphasize the core essential material". Therefore, the video instruction developed for this study, which is primarily a means to the end of providing the writing instruction, aimed at following the recommendations by Brame (2016, p. 5):

- Keep videos brief and targeted on learning goals.
- Use audio and visual elements to convey appropriate parts of an explanation; consider how to make these elements complementary rather than redundant.
- Use signalling to highlight important ideas or concepts.
- Use conversational, enthusiastic style to enhance engagement.
- Embed videos in a context of active learning by using guiding questions, interactive elements, or associated homework assignments.

While possible criticism is not questioned, the following discussion outlines advantages and limitations of instructional videos for the present research. Firstly, ex-

pert knowledge can be shared (see Sections 5.4 and 5.5.1), which appears especially important regarding the use of instructional videos in the writing classroom since teachers in English-speaking countries rate their own expert knowledge in the area of writing instruction as rather poor (S. Graham, 2019). Although little is known about this aspect in the specific region of the present study, this might also apply. Secondly, a further advantage of videos is considered in allowing the present study to implement the instructional strategies that were presented as evidence-based in Chapter 2, Section 2.1.1, i. e., explaining, modelling and illustrating scaffolding practices. Thirdly, the instruction by videos can be given several times, if necessary, e. g., if a student was absent or is working in a separate setting, or if a class needed the input more than once. This is considered as counterbalancing, to some degree, the potential disadvantage of an instructional video mentioned above.

The teacher's manual by Glaser and Palm (2014) was selected because it instructs specifically on the writing of narrative texts, as distinct from spelling instruction. The manual describes the ability to produce text that: 1) fits the writing occasion; 2) is comprehensible; and 3) is appealing to the reader. The program teaches a limited number of writing strategies including associated genre-specific skills, and introduces self-regulatory procedures to control, monitor and evaluate one's own strategic approach, such as setting writing goals and evaluating writing outcomes. The instructional techniques include, for example, modelling scaffolding tools through visualisation, and it provides learning aids (Glaser & Palm, 2014). Many of these contents were summarised as key points of evidence-based writing instruction in Chapter 2, Section 2.1.2.

The manual itself is based on the state of the art of evidence-based writing instruction and includes the approach of self-regulated strategy development (SRSD) (Harris et al., 2003). This approach, which has been shown to be especially efficient for struggling writers (S. Graham et al., 2005), can "assist students in developing knowledge about writing and powerful skills and strategies involved in the writing process, including planning, writing, revising, and editing" (Harris et al., 2003, p. 2). It also supports the students to monitor both to manage the own writing and writing development. The third aim in this manual is to promote the children's development of positive attitudes towards writing and themselves as writers.

The instructional videos were drawn, animated and edited as summarised earlier (see Section 5.4). The first version of the instructional videos was applied by three Grade 5 and two Grade 6 classes from autumn 2021 to spring 2022. Based on feedback from teachers and students, the videos were revised in terms of pace of speech and complexity of language. In addition, teachers unanimously reported that the 'Smiley Protocol', intended to encourage self-reflection on writing strategies, was too time-consuming to implement and did not elicit sufficiently differentiated reflections from students. Consequently, the protocol was removed. The latter change in particular must be considered when interpreting the results of RQ 1B, as the content

of the instructional videos no longer fully complies with the evidence-based procedure of the SRSD.

All videos and additional teaching material were provided via the Moodle platform of the University of Fribourg. A two to three sentence description, with the indication of time, labelled each video. An 18-week overview of units outlined the content. A first option with a faster pace for regular classes (see example Appendix I) and an alternative with a slower pace was provided.

6. Study Results

6.1 Results of quantitative data analysis

Text quality and text length, writing fluency and reading fluency, and writing motivation, were assessed using the instruments and statistical analyses described in Chapter 5. This chapter presents the results for RQs 1 to 4, in their respective order. Each section elaborates, firstly, whether the assumptions were met or how any violations were dealt with. From this, it is derived whether a general linear model (GLM) can be run. The second part of each section contains the descriptive data of interest to each respective RQ. Each section's third part presents the results of inferential statistics, according to the findings, in tables and/or graphs, followed by the interpretation of the results. For each RQ, the main effects and the first-order interaction are calculated, firstly, by one model. Further effects of students' characteristics are then explored by separate analyses to check for any second-order effects. Because of the repetition of the analyses and the risk of committing an α-error accumulation, i. e., familywise error, a Bonferroni correction is applied for these explorations, by adapting the p-level to $p \leq .005$ (see Chapter 5, Section 5.3.3.2). The most relevant findings of the data exploration for second-order effects are described and interpreted within each section. The further explorative data are presented in tables in the Appendix J where supplemental findings of interest are also highlighted by frames and bold font.

Since the initially planned sample size of N 96 was exceeded, the statistical power to analyse RQs 1, 2, and 4 was $1 - \beta > .90$. However, the models to answer RQs 3A and 3B, regarding dictating only, were run within the data of the EG 1 students. The statistical power of these tests is therefore reduced due to the smaller sample size, with minimum valuable data for $n = 34$. A post hoc power analysis is run with the G*Power 3.1 software (Buchner et al., 2014) to determine statistical power. The power for a two-sided t-test between two dependent means is $1 - \beta = .81$, calculated to detect moderate effect sizes. The probability of a type II error, in other words the risk of not rejecting a false null hypothesis, is therefore higher for RQs 3A and 3B.

On a subgroup level, the group 'mild severity' includes only two participants. With one exception, the exclusion of participants with mild dyslexia neither changed the significance nor the effect size of most analyses. Therefore, this subgroup is only excluded for the reading fluency analysis, as this group did affect the model's assumptions.

6.1.1 Selection of students' characteristics from theory and qualitative component

The software MAXQDA, used in the qualitative content analysis, allows the research to summarise and calculate frequencies of entries. This function in MAXQDA is used to obtain an overview of challenging characteristics mentioned by teachers and specialists with a high frequency. These characteristics are entered into the statistical model for second-order effects and interactions, to explore whether they suppress or reinforce effects.

All interview partners talked about 'oral language problems', and 'behaviour and working process issues' were mentioned frequently, meaning concentration, and structure in working processes, while dictating. DLD is considered to be a major reason for oral language problems, still in Grade 5. DLD is therefore systematically included in the explorative analyses of second-order effects. 41 of the 107 participants had received a diagnosis of DLD earlier in life. Oral language skills and/or dictating could also be affected by a bi- or multilingual background (Feng et al., 2024); but only four statements by interviewees mentioned bilingualism or multilingualism as an adverse factor. However, the group of students speaking more than one language comprises almost half of the sample, i. e., $n = 55$ speaking only German versus $n = 52$ speaking German and (an)other language(s). Due to the number concerned, bilingualism or multilingualism is therefore included in the models. Only ten participants had phonetic-phonological disorders, and these students were not equally distributed, as discussed in Chapter 5, Section 5.3.4.1. A phonetic-phonological disorder could impact dictating; but since only two participants in EG 1 are concerned in this respect, this comorbidity is not systematically included in the model. The above-mentioned problems of 'behaviour and working process issues' are considered as typical problems of students with AD(H)D. Therefore, AD(H)D is also explored systematically. 28 participants had an AD(H)D diagnosis. However, students without AD(H)D could possibly have problems with working behaviour. The available data do not allow the research to capture these problems.

Although not analysed in depth, teachers and specialists from specialised schools referred frequently to complex disorders. Thus, it is hypothesised that the type of school could have an impact and is therefore also systematically included in the exploration. Furthermore, theoretically derived variables to control for are gender and severity (see Chapter 2, Section 2.2.3). The review of current state of research shows differences between the groups and in the development of groups in terms of severity (e. g. Quinlan, 2004) and gender (e. g. Camacho et al., 2021), which led to the integration of these characteristics in the exploratory analyses. As discussed in Chapter 5, Section 5.3.3, the presence of learning disabilities and further support (LD+) are also systematically explored.

6.1.2 Research Question 1 – Text quality and text length

6.1.2.1 RQs 1 A and 1 B – Text quality

RQ 1 asks about significant effects on text quality between groups, due to the use of STT and/or the writing instruction by videos. The score of Subtest 7 of the TOWL-4 (Chapter 5, Section 5.4.1) serves to answer these RQs 1A and 1B.

Firstly, the pre-test score is used to check whether the assumptions to run repeated measures ANOVAs are met and whether the groups differ significantly from each other in the pre-test scores. Normal distribution is assumed to be given, although the Shapiro-Wilk test for EG 2 is significant at p .021, which would compromise the GLM as a robust analysis. On the one hand, all Kolmogorov-Smirnov tests are not significant; and on the other hand, all groups with comparable sizes have over 30 participants. There are no outliers. Homogeneity of variances is given according to Levene's test p >.216. Homogeneity of covariances is given according to the Box-test, but close to significance, p = .060. A one-way ANOVA for the group comparison and repeated measures ANOVAs can thus be run. As a precaution for robust testing, the critical normal distribution should be kept in mind in case of a significant result. As elaborated in Chapter 5, Section 5.3.3.2, the correction of the significance level $p \leq .005$ is applied for the two-way repeated measures ANOVA with two factors. The group comparison of the pre-test output using a one-way ANOVA shows no significant difference between the three groups: with F (2,103) = .043, and p = .958.

The data on N 105 students were valuable for the analysis. Table 6.1, presenting the means (M) and standard deviations (SD), show increasing M over all three treatment conditions.

Table 6.1: Descriptive statistics; M and SD for text quality measures

	Group	N	Mean (M)		Standard deviation (SD)	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
Score subtest 7	CG	34	6.62	6.94	3.47	3.11
	EG 1	35	6.77	6.97	3.81	3.50
	EG 2	36	6.81	7.44	3.58	4.14

Table 6.2 presents that one-way repeated measures ANOVA across all groups. The increase in the text quality scores between pre- and post-test is non-significant, despite the general increase of M from 6.73 points at pre-test to 7.12 points at post-test. Over the 18-week intervention period, this is equal to an increase of 0.39 points in Subtest 7 of TOWL-4, which has a maximum score of 21 points. Neither differences between the groups are found nor a significant interaction between the treatment condition over time. In other words, the treatment condition had no effect on the development of text quality during the intervention period. Figure 6.1 illustrates the development over time, pre-test as measurement occasion '1' and post-test as measurement occasion '2', among the three groups with different treatment conditions.

The three columns on the left visualise the unsystematic development at the individual level, while both the slopes in the column on the right and Figure 6.2 show the mean values of the development on a group level, which does not differ significantly.

Table 6.2: Main effects and interactions text quality

TOWL Subtest 7	Sum of squares (SST)	Mean of squares (M)	F (2,102)	p	η2
Main effects within (time)	7.88	7.88	1.465	.229	.014
*Interaction (time*group)*	1.82	.91	.169	.844	.003
Main effects between (group)	4.52	2.26	.108	.897	.002

To summarise, Hypothesis $H1_{1A}$ is rejected. The quality of texts from students of EG 1 does not show a stronger increase between pre- and post-test than the quality of texts from comparable students in EG 2 and CG. Thus, the use of the STT tool does not show a significant effect on text quality over the 18-week intervention period. $H1_{1B}$ is also rejected, because the quality of texts from students in both experimental groups with the writing instruction by videos does not show a significantly stronger increase than the quality of texts written by students from CG. The writing instruction by videos did not significantly increase text quality over the 18-week intervention period.

6.1.2.2 Exploration of effects of participants' characteristics on text quality

As elaborated above, it is possible that students' characteristics function in a suppressing manner and interact with main variables. Therefore, two-way repeated

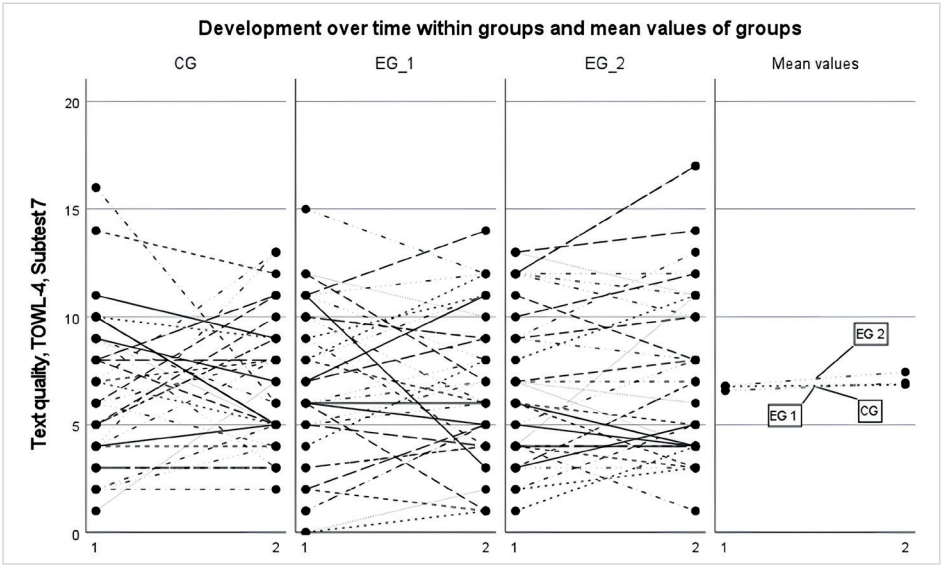


Figure 6.1: Development of text quality over time and group, presented by individuals and groups

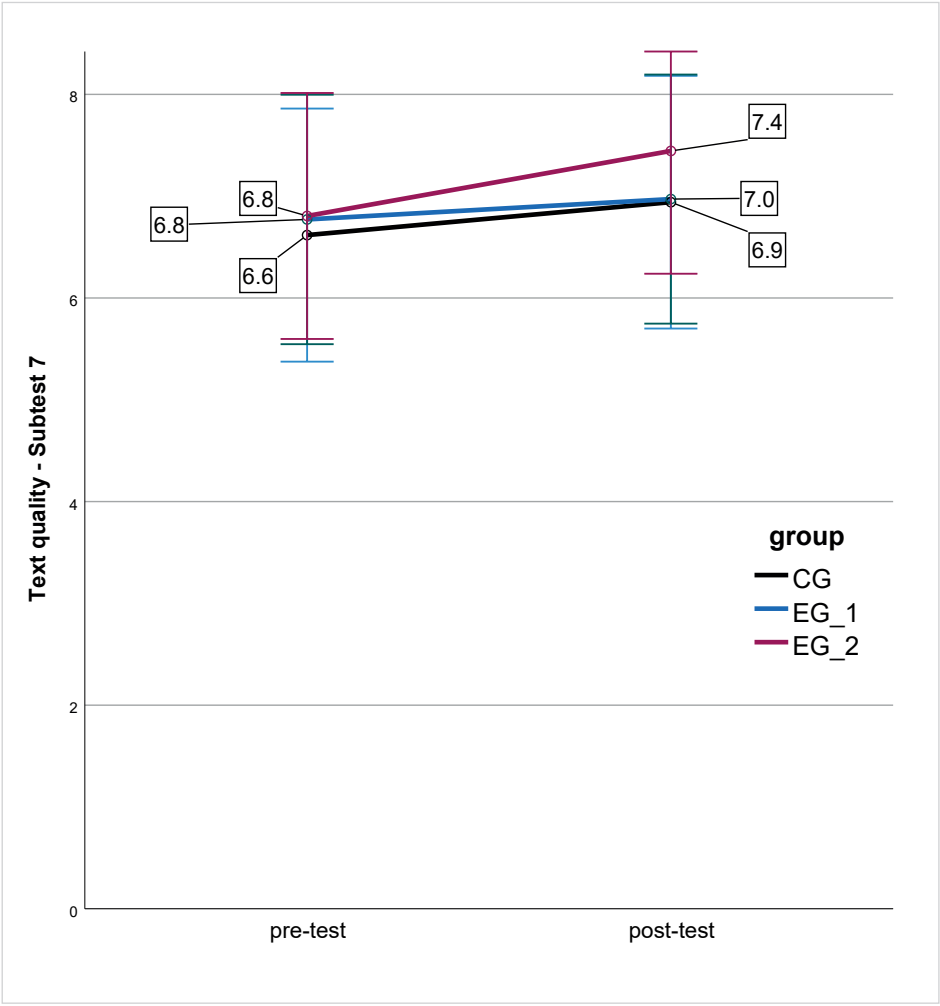


Figure 6.2: Development of text quality over time and group

measures ANOVAs are run. The variables, DLD, AD(H)D, school type, language, gender, severity, and LD+, are separately analysed in the model. As elaborated above, the Bonferroni correction is applied (Bortz & Schuster, 2010) by adjusting the p-level to $p \leq .005$ to avoid α-error accumulation.

The only significant second-order effect (see Appendix Table J1) over time is AD(H)D. When AD(H)D is integrated into the model to detect any second-order effects, this results in an effect of the interaction with the development over time, $p = .014$, and a moderate effect size of $\eta_p^2 = .060$. However, this effect is statistically not significant at the adjusted level of $p < .005$. It is important to emphasise that no significant effect of the treatment condition is found when the interaction of AD(H)

D x group x time is integrated into the model. In other words, the presence of the AD(H)D might impact the development of text quality over time in all groups but cannot be explained by the application of STT or the writing instruction by videos because the development of the control group is alike.

As elaborated above, the output of the statistical analyses allows the research to detect both interactions between factors within the subjects and also effects between subjects. In the case of the present analyses, several characteristics show significant differences, with moderate to strong effect sizes, between participants according to the presence of certain characteristics (see Appendix Table J2) Such effects show differences between subgroups that are not due to the intervention. Therefore, these results do not answer the RQs above but may provide insights into the writing skills of the subgroups that could be of interest for further research. Together with the qualitative results, they might also help to understand the challenges of using such a tool in the classroom.

The analyses show significant differences at the level of $p < .001$ for DLD, the type of school, severity of dyslexia, and language (mono- or multilingualism). The effect sizes for DLD, school type, and severity, $\eta_p^2 > 14$, are strong, while the one for language, above $\eta_p^2 = .06$, is moderate (Cohen, 1988). All these characteristics have no significant interaction with the treatment condition. Therefore, the characteristics of participants do not significantly affect the development of text quality under different treatment conditions.

6.1.2.3 RQs 1 C and 1D – Text length

Text length was discussed, in Chapter 4, Section 4.2.1.2, to be an important predictor for text quality and is therefore examined in connection with text quality. Text length is measured by the total words written (TWW) in the TOWL-4 text writing activity, which lasted for 15 minutes.

All assumptions for the TWW-score are met according to visual inspection and test statistics. A one-way ANOVA is run to check for the comparability of the groups, which shows no significant difference between the groups, with $F\,(2,103) = 1.17$, and $p = .315$. The box plot of TWW shows two outliers, which are excluded from the analysis.

The *M* values of TWW in the TOWL-4 assessment, shown in Table 6.3 increase between pre- and post-test. The development over time and among all groups is significant, with $F\,(2,100) = 9.06$, and $p = .003$, with a strong effect size of $\eta_p^2 = .082$. However, there are no significant group differences and no interaction between the main factors, group and time, found (Table 6.4 and Figure 6.3).

Both hypotheses $H1_{1C}$ *and* $H1_{1D}$ are thus rejected. There is neither a significantly stronger increase from pre- to post-test in text length from students of EG 1 compared to EG 2 and CG due to the STT condition, nor in EG 1 and EG 2 compared to the CG due to the writing instruction.

Table 6.3: Descriptive statistics; M and SD of text length

	Group	n	Mean (M)		Standard deviation (SD)	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
TOWL: TWW	CG	34	92.71	100.76	34.81	30.78
	EG 1	35	84.63	96.66	36.97	36.22
	EG 2	34	94.86	106.91	34.52	41.49

Table 6.4: Main effects and interaction in text length

Total words TOWL	Sum of squares (SS_T)	Mean of squares (M)	F (2, 100)	p	η^2
Main effects within (time)	5968.90	5968.90	9.061	.003**	.082
Interaction (time*group)	181.63	90.81	.138	.871	.003
Main effects between (group)	3715.22	1857.61	.995	.373	.019

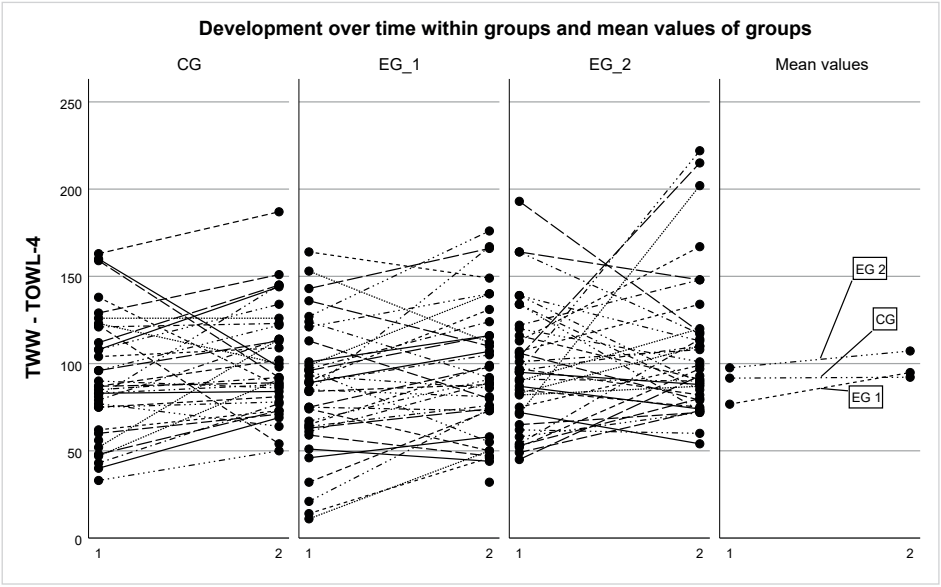


Figure 6.3: Development of text length over time and group, presented by individuals and groups

6.1.2.4 Exploration of effects of participants' characteristics on text length

Some subgroups, according to certain characteristics in participants, show significant differences in their development over time or between their baselines, i. e., output at pre-test. But again, no interaction with the treatment condition is observed (see Appendix Table J3).

The two characteristics, school type and gender, show significant differences between baselines of the subgroups. While school type shows a moderate effect size, the one for gender is strong (see Appendix Table J4). Both do not interact with the

treatment condition. This means that no characteristics of the participants significantly impact the development of text length under different treatment conditions.

6.1.3 Research Question 2 – writing fluency and reading fluency

The results on writing fluency, RQ 2A, and reading fluency, RQ 2B, are presented separately in the following sub-sections.

6.1.3.1 RQ 2A – Writing fluency

Writing fluency is measured by CBM-W, i. e., CWS for accuracy and TWW for the amount of text written.

No outliers are observed. All assumptions for the CWS are met. According to the Shapiro-Wilk test, $p > .05$, normal distribution is given in the three groups. Levene statistics show homogeneity of variance, $p = .491$. The one-way ANOVA, run to compare the outcome at the pre-test, indicates no difference between the three groups, with $F\ (2,101) = .609$, and $p = .546$.

The data on N 104 students are valuable for the analysis. The descriptive statistics presenting the mean (M) and standard deviation (SD) show an increase in the CWS in each group between pre- and post-test (Table 6.5).

Table 6.5: Descriptive statistics; M and SD for writing fluency

	Group	n	Mean (M)		Standard deviation (SD)	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
CWS	CG	34	13.26	20.38	7.45	10.37
	EG 1	34	14.65	16.68	8.79	11.58
	EG 2	36	15.36	20.44	7.88	12.51

Table 6.6 shows the above-described increase of CWS, which is highly significant over time, with $F\ (2,101) = 21.59$, and $p < .001$, and with a strong effect size, $\eta_p^2 = .176$ (Cohen, 1988). However, the treatment condition and time do not interact significantly, with $F\ (2,101) = 2.06$, p .133.

Table 6.6: Main effects and interaction in writing fluency measures

Correct word sequences	Sum of squares (SS_T)	Mean of squares (M)	F (2,101)	p	η^2
Main effect within (time)	1169.17	1169.17	21.586	**<.001*** **	.176
Interaction effect (group*time)	223.13	111.56	2.060	.133	.039
Main effect between (group)	175.64	87.82	.610	.545	.012

The individual lines in Figure 6.4 illustrate a similar picture of an increase with unsystematic development within each group. Both the descriptive data and visual

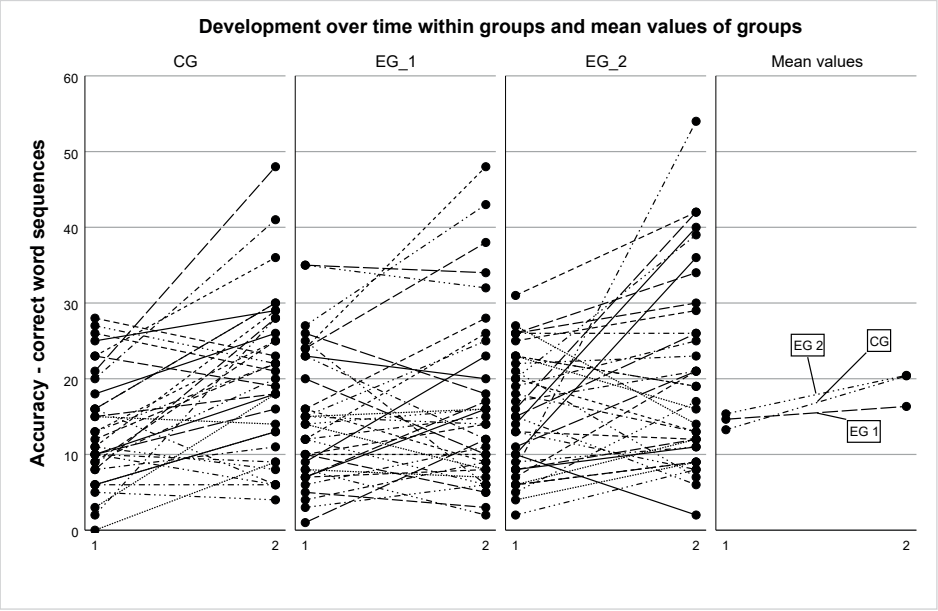


Figure 6.4: Development of writing fluency over time and group, presented by individuals and groups

inspection of the slopes in Figure 6.4 on the right, showing the mean values of each group, might give the impression of a difference between the groups that does not correspond to the output of the repeated measures ANOVA.

To avoid missing an existing group difference due to the non-significant different baseline values, a further case of the GLM is run at this point. A univariate ANOVA is computed to measure differences in the change between pre- and post-test. Therefore, the 'change in CWS' between pre- and post-tests is calculated and used as the output variable. The baseline value of CWS is then added to the UNIANOVA-model as the control variable. The results of the univariate test confirm the results above, with $F\ (2,101) = .420$, and $p = .658$. There is no significant difference between the three groups due to the treatment condition.

$H1_{2A}$ is therefore rejected. There is no significant difference in the development of writing fluency skills between pre- and post-tests in students of EG 1 and in students of EG 2 and CG. The use of the tool thus did not impact the development in a significantly positive or negative manner.

6.1.3.2 Exploration of effects of participants' characteristics on writing fluency

The exploration of participants' characteristics indicate that no characteristics suppress the effect of the treatment condition, since, at the $p \leq .005$ level, there are no significant differences in the development over time within the subgroups and

no interactions between the subgroups and the treatment condition over time (see Appendix Table J5).

By contrast, several subgroups differ significantly in their baseline of writing fluency. DLD, enrolment in specialised school, gender, and severe dyslexia, all show a significantly lower baseline for writing accuracy, with $p \leq .002$. These results show moderate and strong effect sizes. However, there are no interactions of these subgroups over time (see Appendix Table J6). In other words, no characteristics of the participants significantly impact the development under different treatment conditions.

6.1.3.3 RQ 2B – Reading fluency

The data on *N* 107 are analysed for differences between pre- and post-test measures on reading fluency. Reading fluency is measured by the SLS 2–9 (see Chapter 5, Section 5.4.4). Both visual inspection and tests give reason to assume that the assumptions for the score of SLS 2–9 are met. Box plots showed outliers in the CG, but since these are in the range of the two EGs, these are not excluded from the analyses.

The one-way ANOVA to compare the outcome of the reading fluency variable at pre-test time shows no significant difference, with $F(2, 104) = .417$, and $p = .660$.

As shown in Table 6.7, all groups increase in the *M* number of sentences read between pre- and post-test.

Table 6.7: Descriptive statistics; M and SD for reading fluency

	Group	n	Mean (M)		Standard deviation (SD)	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
SLS 2–9; raw value	CG	35	32.11	37.97	10.55	10.40
	EG 1	36	30.03	34.22	11.97	12.39
	EG 2	36	32.00	36.69	9.99	10.92

This increase over all groups is significant, with $F(2, 104) = 119.70$, and $p < .001$, with a very strong effect size of $\eta_p^2 = .530$. The overall increase is visible in the graph on both the individual and the group level. There is no interaction between time and group (see Table 6.8 and Figure 6.5). Therefore, there is no difference between the three treatment conditions.

Table 6.8: Main effects and interactions in reading fluency

Score SLS 2–9	Sum of squares (SS_T)	Mean of squares (M)	F (2,104)	p	η^2
Main effects within (time)	1292.36	1292.36	120.567	**<.001*** **	.537
Interaction (time*group)	25.75	12.88	1.201	.305	.023
Main effects between (group)	331.54	165.77	.706	.496	.013

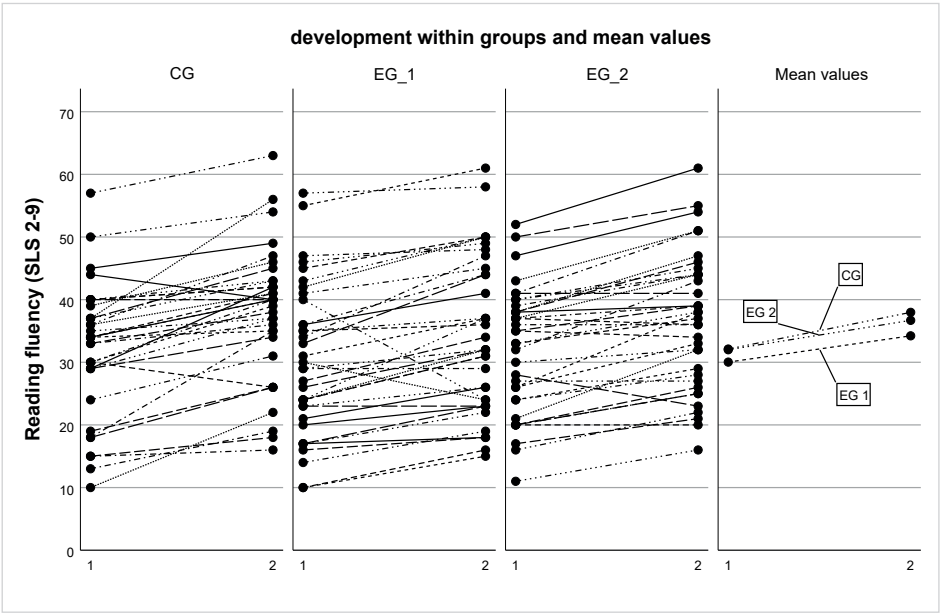


Figure 6.5: Development of reading fluency over time and group, presented by individuals and groups

It was hypothesised that EG 1 improves more strongly in reading skills due to potentially more reading activity at the screen, which was encouraged by the revision with the read aloud function, i. e., TTS, during the dictating procedure. However, $H1_{2B}$ is rejected, because there is no significant difference in the development of reading fluency between pre- and post-test in the comparison of students of EG 1 and of students of EG 2 and CG.

6.1.3.4 Exploration of effects of participants' characteristics on reading fluency

On a subgroup level, the analyses of the two-way repeated measures ANOVAs show violations of the homogeneity of variances regarding the subgroups DLD and severity. The Box test for homogeneity of covariances is not significant, at $p = .236$, but four Levene's tests are significant at the .05 level (p .011 – p .015). As stated in Section 6.1, the small subgroup of students with mild dyslexia, i. e., n 2, is subsequently excluded.

On a subgroup level, the results show one significant difference at $p \leq .005$, namely between the students with severe and moderate dyslexia. At the baseline, the subgroup with severe dyslexia reads significantly fewer sentences, with $M_{Diff} = -13.31$, during the 3-minute task than the group with moderate dyslexia (see Appendix Tables J7 and J8). However, no second-order interaction or subgroup difference proves

significant. Therefore, it is summarized that no characteristics in students affect the treatment condition significantly.

6.1.4 Research Question 3 – dictating fluency

RQ 3A examines differences between the two modalities, dictating and handwriting; and RQ 3B investigates differences in the development of dictating fluency. In contrast to the presentation of the RQs (see Chapter 5, Section 5.2.3) and their discussion (see Chapter 7, Section 7.1.3), the results in this section are presented according to the values analysed, and not regarding the comparison of modalities and development.

These RQs are answered by a slightly different approach concerning the measures, the dataset, and the model calculated. As discussed in Chapter 5, Section 5.4.2, the writing fluency measures used are: 1) the total words produced in five minutes and measured in both modalities (TWW for handwriting, TWD for dictating), which score is used as a productivity score; and 2) the %CWS serves as value for accuracy – since the %CWS score is a relative score or, in other words, a productivity-independent score, the number of words, i. e., the dictating productivity, does not primarily impact this score (Malecki & Jewell, 2003).

To calculate differences between the two modalities, 1) at pre-test and 2) at post-test, *t*-tests with two dependent means are run for both writing fluency and writing productivity. A third, dependent samples *t*-test is run to investigate whether there are differences between these measurements in the dictated texts of the pre-test and those of the post-test. As outlined above, a post hoc power analysis is run to evaluate the power in the sample of EG 1 only. The statistical power for two-tailed testing, with an effect size of d .5 and the α level of .05, is $1 - \beta = 0.81$ for the analysis of 34 scores and $1 - \beta = 0.83$ for 36 scores.

Since *t*-tests are considered to be rather robust with $n > 30$ (Bortz & Schuster, 2010) and because the importance of checking the assumption of normal distribution is questioned (Hemmerich, 2024a), only visual inspection of the box plots for outliers is applied, to be aware of potentially problematic data. The visual inspection of the box plots shows two outliers for both the variables, TWD and %CWS, which are thus excluded from the sample. Because the visual inspection subsequently shows a tolerable normal distribution, *t*-tests are run.

6.1.4.1 RQ 3A1 and 3B1 – Writing productivity – TWW and TWD

The first measure to investigate differences between text production modalities is writing productivity, measured in TWW and TWD in a 5-minute writing sample.

As presented in the two upper rows of Table 6.9, the comparison of the two conditions for the same students shows, at both times, a higher *M* of total words produced in dictated texts than in handwritten texts.

Table 6.9: Descriptive statistics, M and SD for writing productivity in different modalities

Modality	Measurement time	N	M	SD
TWW handwritten	Pre-test	32	32.72	15.66
TWD dictated	Pre-test	32	51.59	25.74
TWW handwritten	Post-test	33	33.55	15.25
TWD dictated	Post-test	33	74.12	50.69
TWD dictated	Pre-test	34	49.97	25.83
TWD dictated	Post-test	34	72.71	50.60

The difference in the score at pre-test is significant; with $M = 18.88$, $SD = 27.74$; t (31) = 3.849 and $p < .001$. This difference is consistent with the post-test score, with $M = 40.58$, $SD = 51.97$, t (32) = 4.485 and $p < .001$. Both differences have $d = .68$ and $d = .78$, as medium, the latter almost strong, effect sizes (Cohen, 1988).

The comparison of the dictated texts of pre- and post-test shows an increase in the number of TWD, see the bottom row of Table 6.9 and Figure 6.6. While the students dictated close to 50 words in five minutes at pre-test, they dictated almost 73 words at post-test. This is a significant increase, with $M = 22.74$, $SD = 48.86$, t (33) = 2.71 and $p = .011$, with a close to moderate effect size of $d = .47$.

The two hypotheses concerning text length, $H1_{3A1}$ *and* $H1_{3B1}$, are therefore accepted. Firstly, the dictated texts by EG 1 are significantly longer than the handwritten texts. Secondly, the dictated texts at post-test are significantly longer than the dictated texts at pre-test.

6.1.4.2 RQ 3A2 and 3B2 – Writing accuracy in dictated texts – %CWS

The descriptive statistics of the tests comparing the %CWS in the handwritten modality with the %CWS in the dictating modality are presented in the two upper rows of Table 6.10. The comparison of the two conditions for the same students shows, at both times, a higher *M*-%CWS in dictated texts than in handwritten texts.

The differences between the scores of handwritten and dictated %CBM at pre-test are significant, with $M = 38.57$, and $SD = 3.45$, and t (31) = 11.18, and $p < .001$. These differences are consistent with the post-test score, with $M = 34$, $SD = 18.23$, and t (32) = 10.71 and $p < .001$. Both differences have $d = 1.98$ and $d = 1.87$, as very strong effect sizes. This means that the students produce a significantly higher %CWS when they dictate texts with STT than when they write by hand.

The comparison of the accuracy in dictated texts from pre- and post-tests, by the productivity-independent variable %CWS, presented in the third row in Table 6.10 and in Figure 6.7, results in no difference, with $M = 3.06$, $SD = 17.17$, and t (33) = 1.04 and $p < .307$. This means that, regardless of the amount of text produced, the texts contain the same %CWS at post-test as they do at pre-test.

However, comparing the accuracy by the productivity-dependent variable CWS, presented in the bottom row of Table 6.10, has $M = 16.12$, $SD = 38.42$, and $t(33) = 2.45$ and $p = .020$. The effect size is $d = .42$, which corresponds to a small effect.

Table 6.10: Descriptive statistics; M and SD for writing accuracy in dictated and handwritten texts

Modality	Measurement time	N	M	SD
%CWS handwritten	Pre-test	32	44.17	19.27
%CWS dictated	Pre-test	32	82.73	14.71
%CWS handwritten	Post-test	33	46.57	19.73
%CWS dictated	Post-test	33	80.56	17.64
%CWS dictated	Pre-test	34	82.72	14.67
%CWS dictated	Post-test	34	79.67	18.14
CWS-dictated	Pre-test	34	41.53	25.27
CWS-dictated	Post-test	34	57.65	38.11

$H1_{3A2}$ hypothesises that the accuracy in dictated text is higher than in texts written by hand. It is thus accepted, because the dictated texts from EG 1 contain significantly more correct spelling than the texts written by hand.

The second hypothesis, $H1_{3B2}$, assuming a statistically significant increase in writing accuracy from pre- to post-test, is rejected. The dictated texts do not contain a higher percentage of %CWS at post-test than at pre-test. However, the %CWS remains consistent despite the increase of TWD.

No hypothesis was formulated concerning CWS, because this measure primarily serves the common understanding and contextualisation of the findings with respect to the current state of research.

To summarise, four points concerning dictated texts are stated: 1) students with dyslexia in Grade 5 produce significantly more text, measured in total words, when they dictate than when they write by hand; 2) after the 18-week intervention period, students dictate significantly more words than they do before the intervention, and the application of STT over time results therefore in longer texts; 3) these students produce a significantly higher %CWS in dictated texts than in handwritten texts, thus the accuracy is better in dictated texts than in handwritten texts; and 4) while the total number of CWS in dictated texts increases due to the increased TWD, the percentage of CWS from TWD remains the same – in other words, the accuracy of dictated texts remains the same, while the length of dictated texts increases.

6.1.4.3 Exploration of effects of participants' characteristics on dictating fluency

It is conceivable that certain characteristics of the students interact with the treatment and result in different effects of the application of the tool. However, no characteristics show significant interactions due to the treatment condition. Additionally,

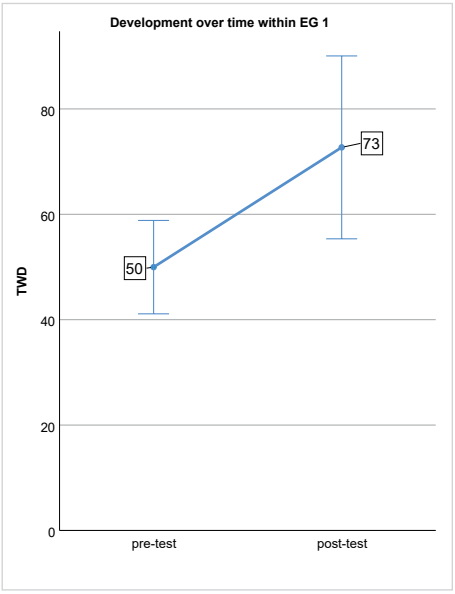


Figure 6.6: Development TWD over time

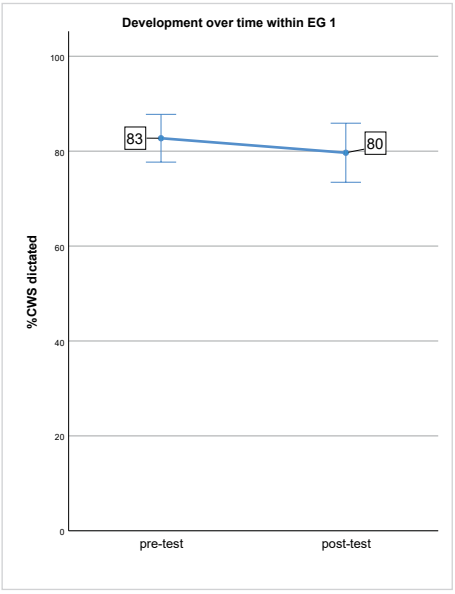


Figure 6.7: Development %CWS over time

problems with normal distribution and homogeneity of variance in the subgroups of the EG 1-sample contradicted further interferential statistical analyses among subgroups of students that dictated their texts. Characteristics explored for the development of TWD and %CWS were the four most mentioned during the interviews, DLD, AD(H)D, school type, severity of dyslexia (moderate and severe).

Table 6.11 presents the development of TWD and %CWS between pre- and post-tests among the subgroups. Arrows and bold font indicate strong increases and reductions of the mean values within subgroups, discussed below.

Table 6.11: Descriptive statistics, M and SD, of TWD and %CWS among subgroups

	subgroups	N	Mean (M)		Standard deviation (SD)	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
TWD	DLD no	20	55.15	**89.75** ↑	25.83	57.93
	DLD yes	13	43.23	48.46	25.76	23.45
TWD	AD(H)D no	22	48.09	66.73	23.87	45.88
	AD(H)D yes	11	55.18	**87.00** ↑	30.70	60.48
TWD	Regular school	21	70.05	89.10	57.47	54.00
	Specialised school	15	46.93	74.67	41.13	84.71
TWD	Severe dyslexia	22	45.41	77.14	23.20	58.88
	Moderate dyslexia	8	50.25	71.38	24.39	27.09
%CWS	DLD no	22	84.38	85.43	15.26	15.66
	DLD yes	13	75.17	**69.61** ↓	19.37	19.15

	subgroups	N	Mean (M)		Standard deviation (SD)	
%CWS	AD(H)D no	24	79.28	**75.90** ↓	18.66	18.47
	AD(H)D yes	11	84.60	**87.51** ↑	13.67	16.53
%CWS	Regular school	21	85.72	85.64	16.13	16.27
	Specialised school	15	72.82	70.26	16.69	17.64
%CWS	Severe dyslexia	22	82.01	77.48	15.42	18.27
	Moderate dyslexia	10	72.53	78.57	18.21	18.09

The mean of the TWD increases in all subgroups form pre-test to post-test. However, the difference between the subgroups of students with and without the diagnosis of DLD appears to be important: while students without DLD increase their TWD by almost 34.6 words dictated in 5 minutes, the students with DLD increase only by 5.3 words. Regarding accuracy, the %CWS in students with DLD decrease from pre- to post-test. These results might indicate that students with DLD benefit less from the of STT-tool.

Somewhat surprising is the increase of the M value according to the subgroups of students with and without AD(H)D. While the students without AD(H)D increase the TWD by 18.6, those with AD(H)D increase by 31.8 dictated words in 5 minutes. Regarding accuracy, those without AD(H)D decrease the percentage of correct word sequences by 3.4, while those with AD(H)D increase by 2.9 %CWS. These findings might indicate that AD(H)D does not affect the benefits of using STT.

6.1.4.4 Exploration of effects of participants' device preference on dictating fluency

A specific focus is directed at 'device preference'. The students' statement about which modality they preferred is used as a grouping variable. The data are analysed to explore whether there is a difference in the dictated texts in relation to this preference. The statistical analyses are underpowered due to the small subsample sizes; for TWD 17 (handwriting preferred), 11 (no preference), and 6 (dictating preferred); for %CWS 15 (handwriting preferred), 13 (no preference), and 6 (dictating preferred).

Table 6.12: Descriptive statistics, M and SD, of TWD and %CWS according to device preferences

Groups according to device preferences		**N**	**Mean (M)**		**Standard deviation (SD)**	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
TWD	Handwriting	17	46.12	63.47	26.13	45.15
	No preference	11	52.91	79.09	28.61	64.53
	Dictating	6	55.50	87.17	21.82	37.56
%CWS	Handwriting	15	89.79	83.25	24.35	18.29ha
	No preference	13	76.14	75.00	13.05	19.44
	Dictating	6	79.29	81.07	12.59	15.60

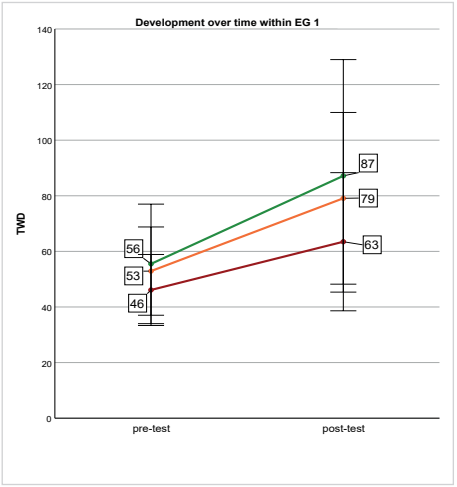


Figure 6.8: TWD in subgroups device preferences

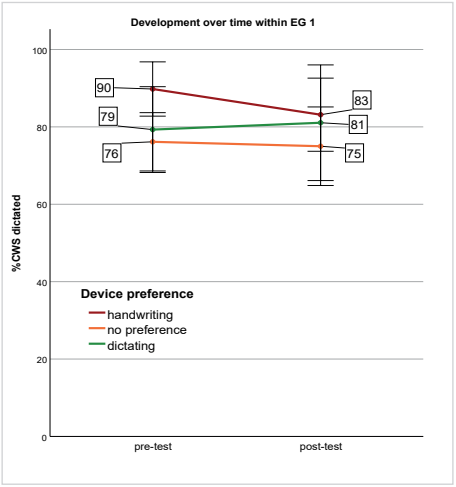


Figure 6.9: %CWS in subgroups device preferences

The results exploration is therefore only illustrated by the descriptive data in Table 6.12 and by graphs in Figures 6.8 and 6.9.

The graph in Figure 6.8 of the development of TWD shows an increase over all groups. On a visual inspection and regarding the scores, the slopes might differ slightly. However, a more important observation might be that the standard deviation of TWD is bigger at post-test than at pre-test.

While the differences between the subgroups in %CWS are small at both measurement times, and the standard deviations overlap strongly, the graphs in Figure 6.9 show slightly different trends. Only the participants that prefer dictating increase in the %CWS. The two other subgroups decrease slightly in the share of %CWS.

6.1.5 Research Question 4 – Writing motivation

The development of writing motivation before and after the intervention is investigated by the translated and adapted survey SWAS by Wright et al. (2019). RQ 4A asks whether the STT intervention affects writing motivation, while RQ 4B asks whether the writing instruction by videos have an influence on writing motivation.

The normal distribution is violated for the EG 2 according to the Shapiro-Wilk test ($p = .028$), and the graphs show two outliers at pre-test and one outlier at post-test. These are thus excluded from the data, which changes the Shapiro-Wilk result to a normal distribution in all three groups $p > .425$ at pre-test and $p > .251$ at post-test. Homogeneity of variances is also given. The three groups do not differ significantly at pre-test measure of overall writing motivation, with $F(2,101) = 1.09$, and $p = .341$.

Writing motivation is measured by a scale with four components, each ranging from 0 to 3 points. Table 6.13 presents the *M* values of the four components added

up, ranging from 0 to 12. Figure 6.10 illustrates the development. The results show that the overall motivation decreases in CG and EG 1 while it increases in EG 2.

Table 6.13: Descriptive statistics; M and SD for overall writing motivation

Writing motivation	Group	N	Mean (M)		Standard deviation (SD)	
			Pre-test (1)	*Post-test (2)*	*Pre-test (1)*	*Post-test (2)*
Sum of means over all components	CG	35	6.60	6.24	1.89	2.04
	EG 1	36	5.97	5.69	2.15	1.97
	EG 2	33	6.44	6.93	1.62	1.73

However, the above presented differences are not significant, as shown in Table 6.14 and illustrated by the overlapping error bars in Figure 6.10. There is no significant effect over time or of the treatment condition, and there is no interaction between the time and the treatment condition.

Table 6.14: Main effects and interactions in overall writing motivation

Writing motivation	Sum of squares (SS_T)	Mean of squares (M)	F (2,101)	P	η^2
Main effects within (time)	.11	.11	.11	.741	.001
Interaction (time*group)	4.87	2.43	2.42	.095	.046
Main effects between (group)	29.68	14.84	2.35	.100	.045

The hypothesis for RQ 4A assumes an increase in writing motivation in EG 1. $H1_{4A}$ is therefore rejected, as no differences between the groups emerge from the analysis. Meanwhile, RQ 4B asked about effects of writing motivation due the writing instructions by videos. $H1_{4B}$ is therefore also rejected, since no differences between the three groups are significant. However, regarding descriptive data and the slopes of the three groups (Figure 6.10), the different trends are obvious: while the CG and the EG 1 show a very similar negative gradient, the EG 2 slope points in a different direction.

6.1.5.1 Exploration of effects of participants' characteristics on overall writing motivation

The same procedure as for RQs 1 and 2 is applied for the exploration of potential effects of participants' characteristics. Repeated measure ANOVAs are run. Again, only significant results $p < .005$ are interpreted and further discussed, to avoid α-error accumulation.

No interactions between participants' characteristics and the treatment condition are found, but two characteristics of students show a significant difference in motivation. Firstly, students with AD(H)D report significantly lower writing motivation scores than those without AD(H)D, with $F\,(2,102) = 9.20$, $p = .003$. The differ-

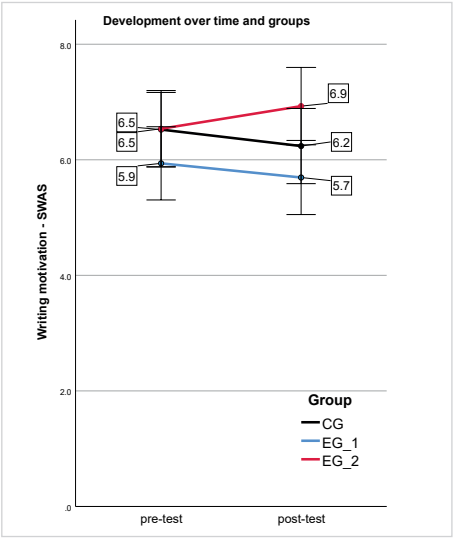


Figure 6.10:
Development of writing motivation over time and group

ence of sum of *M* on the scale and over all groups is *M* -1.22 points. This result has a moderate effect size, $\eta_p^2 = .087$. Secondly, the subgroups show a significant difference according to gender, with $F(2,102) = 12.51$, $p < .001$. Girls report higher writing motivation than boys. The mean difference between the scores is 1.31 points, which has also a moderate effect size, $\eta_p^2 = .112$. No interaction of subgroups and the treatment condition over time is observed. This means that, although students with AD(H)D and students differ according to their gender in their writing motivation, this difference has no effect on the treatment condition.

6.2 Results of qualitative research strand

Challenges of the implementation of STT in the classroom are explored at the end of the intervention study: firstly, by interviews with teachers and specialists of the students who used STT, i. e., EG 1; and, secondly, by three questions to the students of this group. This section presents the results of this qualitative supplemental component. Quantifications of the statements of the respondents provide an overview of frequent and recurring topics. The most prevalent challenges are then summarised and illustrated by examples.

6.2.1 Research Question 5 – Challenges of STT in school

The question about challenges of the implementation of STT in the writing classroom is explored based on eleven interviews with class teachers and specialists. The interview guide, presented in Chapter 5, Section 5.4.5.1, was developed on a theoretical basis and revised according to feedback from professionals. The interview data that had been collected and elaborated on was then analysed, according to Mayring

(2022), by a qualitative content analysis conducted with the software MAXQDA and coded by the code system presented in Chapter 5, Section 5.4.5.2.

The code system comprises six top-level codes, see Table 6.15. The first and the third, and the fourth top-level codes, were originally estimated as the most pertinent to answer RQ 5 about challenges and/or necessary adaptations for the implementation of STT in the classroom. However, the top-level Code 4 provides fewer insights into the question than originally expected, as explained below in Section 6.2.1.3. The top-level Code 2 does not focus on challenges but on benefits of STT, described by the professionals. These coded segments are considered valuable to supplement the above-presented results for RQs 1 to 4. The last two top-level codes, Codes 5 and 6, provide some information about the suitability of the chosen method of instructional videos to implement STT in the classroom. However, this is not the focus of RQ 5 and is therefore not further elaborated on here.

Table 6.15: Top-level code statistics from eleven interviews

Top-level codes		Coded segments	Percentage
1	*Challenges: Reported difficulties of STT in school or SLT/SE*	349	42.77
2	*Benefits: Reported benefits of STT in school or SLT/SE*	104	12.75
3	*Raised or implemented supportive adaptations, factors, resources*	145	17.77
4	*Motives and attitude of teachers/specialists*	113	13.85
5	*Experiences on STT-introduction in SLT/SE*	49	6.00
6	*Experiences on instructions by videos*	56	6.86
	TOTAL	816	100.00

During the qualitative content analysis, the researcher distributed a total of 816 codes to the linguistic material. The resulting top-level code statistics, presented in Table 6.15, reveal which codes are assigned most frequently. The first top-level code (upper grey frame), about challenges of STT in school, shows by far the highest prevalence, with 349 coded segments. This result is not surprising: firstly, because the aim of the interview was to explore challenges; secondly, because the interview guide evoked such; and thirdly, because professionals of the dropouts were interviewed. Dropping out is considered, to some extent, due to important challenges. However, the high number of 349 coded segments, which corresponds to 43 % of the coding, confirms the expected presence of challenges. The third top-level code (lower grey frame), on the other hand, was not initially evident but generated by the inductive coding process. This code summarises 145 segments about supportive factors and adaptations, mentioned as necessary for success or applied during the study. The following sections present a discussion of these two topics, i. e., challenges and adaptations, in detail: firstly, by mapping the code structure including the number of coded segments on lower levels; and secondly, by presenting thematic summaries with examples.

6.2.1.1 Reported challenges of STT for students with dyslexia in school

The code structure of the challenges, illustrated in Figure 6.11, includes two levels of subcodes represented by the circles. The first level subcodes, reading clockwise, are: 1) Adverse factors in the student; 2) Reported problems of the output accuracy of STT; 3) Technical issues with the STT-tool; 4) Difficulty with the dictation process; 5) Adverse factors in the school environment; 6) Negative emotions when dictating; and 7) Adverse factors outside of school.

The most frequent statements by teachers and specialists concern adverse factors in the students. By far, the two most frequently mentioned barriers are 'difficulties in oral language production' and 'behavioural and work process issues', mentioned, respectively, 49 and 48 times. All codes that are coded 20 times or more are presented and discussed in the paragraphs below Figure 6.11.

Limited language skills: Barriers in oral language production appear in nine interviews and the statements concerning multiple language levels and abilities that affect dictating: interviewees name distinct pronunciation, retrieval of vocabulary,

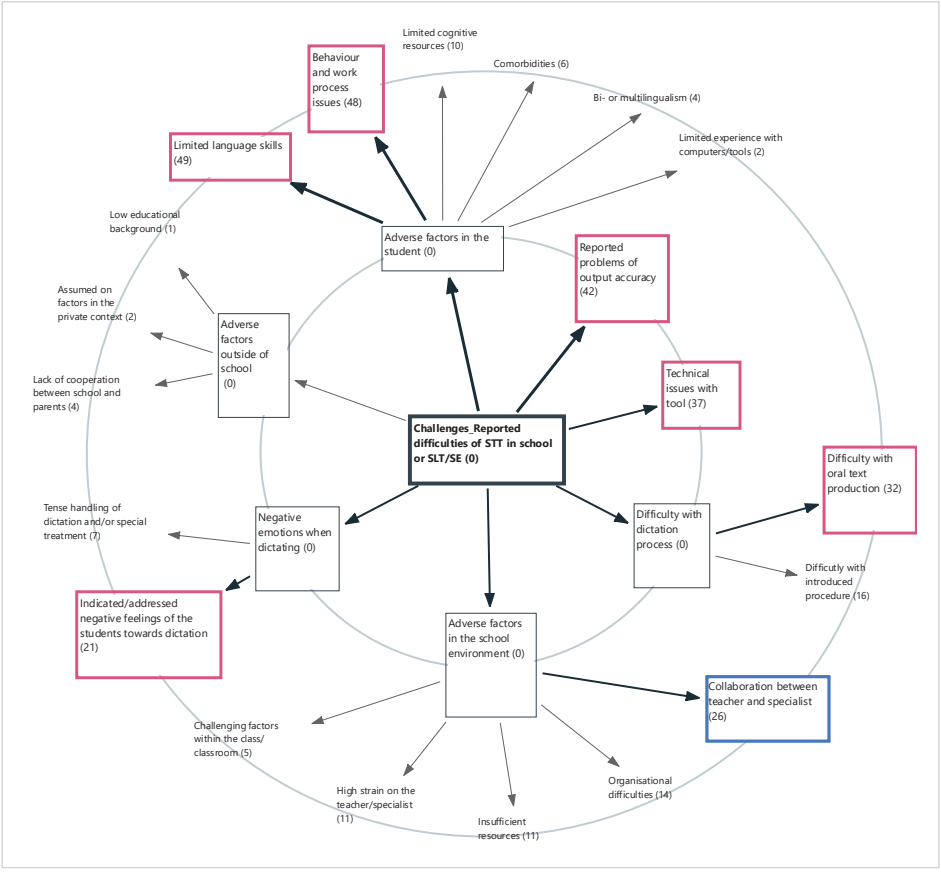


Figure 6.11: Code map of challenges of STT in school, number of codes ordered clockwise according to frequency

sentence structure, influent or nasal speech, prosody, dialectal influences, metalinguistic knowledge, and language planning. The professionals commonly relate these issues to problems in the accuracy of STT; for instance, non-fluent speaking or unclear pronunciation affect word recognition accuracy, while problems in planning and structuring sentences and setting punctuation marks hamper the comprehensibility of the text. One teacher, who accompanied several participants, described different outcomes depending on the students' pronunciation:

> *I have those students in the class who speak relatively clearly, and STT works better with them. And then I have those students for whom the software probably has no chance of really transcribing correctly.* (Interview (Int.) 10, Position (Pos.) 44)

A speech pathologist described the problem of a language disorder at the morpho-syntactic level:

> *She speaks with very peculiar sentence structures. Sentence structure is a big issue for her. And then, of course, Dragon sometimes did something with it* (meaning the spoken input). (Int. 03, Pos. 119)

A special needs teacher expressed the challenge of the difference between informal talking and oral text production, as necessary in the dictating modality:

> *He first had to learn how to produce a complete sentence and not just babble.* (Int. 02, Pos. 109)

Behaviour and work process issues: The coded segments of 'behaviour and work process issues' are present in ten interviews. Similarly to the limited oral language abilities, the segments involve various aspects of behaviour. Aspects captured by this code describe features in students that affect dictating; for example, high awareness of the disorder and subsequent reluctance toward dictating, or low tolerance of frustration. Other aspects capture problems of attention, patience or distractibility that affect a structured way of working. A speech pathologist expressed this as follows:

> *He just wanted to do everything quickly. Yes, I think the problem for him was that he had to say a sentence and to do so, he would have had to think about it carefully and then articulate that sentence. When he had to concentrate on what he wanted to say, he couldn't concentrate on articulating clearly and not speaking too quickly. It's like an adverse combination of everything …* (Int. 09, Pos. 20)

Reported problems of output accuracy: As already mentioned, low output accuracy of STT is, according to the professionals, a major challenge. The code distributed to 42 segments originate from nine interviews. In contrast to the two already discussed codes, the issues mentioned are less diverse. A majority described substantial differences between the students' input and the transcription. While some of these deviations are explained by indistinct pronunciation, other transcriptions are, for the professional, unexpectedly divergent. The professionals expressed surprise because

they either made some experiences themselves which were much more reliable than the students' experiences, or they perceived the students' input as clear and did not expect a so-called hallucination error. One speech pathologist expressed astonishment about the accuracy and quoted a student's anger:

> *The student exclaimed, "Dragon always blocks and doesn't write what I say." Although the student is someone who actually has a very clear pronunciation* (Int. 07, Pos. 25).

This statement illustrates that: firstly, the accuracy issues may be – but only to some extent – related to indistinct pronunciation; and secondly, the problematic accuracy may cause negative emotions when dictating, e. g. frustration.

Technical issues: 'Technical issues' appear in ten interviews. The 37 coded segments address problems with the software itself, the compatibility of the software and the device, and the suitability of the software for primary students. A class teacher described this problem and addressed the difficulty of solving technical problems during classroom management:

> *Problems with the device itself or with the software were a recurring issue. I think, we need to work as a team. As a teacher, I would need more support for that*. (Int. 01, Pos. 30)

Dictation process: The 'dictation process' manifests itself as a further challenge for students with dyslexia. While the introduced procedure (thinking of a sentence, dictating the sentence and setting punctuation marks, and revising the sentence) posed a problem for some students (16 coded segments), the challenge to get used to the oral text production is named by 34 coded segments in eight interviews. Although this code is related to the limited oral language abilities of some, it appears also that students without an oral language disorder must learn and get used to the dictating process. A speech pathologist described such a challenge, where both the student and the speech pathologist had to adapt their habits:

> *If I discussed something with the girl and we forgot to pause Dragon, then, of course, Dragon continued transcribing.* (Int. 06, Pos. 89)

Adverse factors in the school environment: In contrast to the factors within the students (red frames in Figure 6.12), the blue frame around 'adverse factors in the school environment' illustrates challenges outside of the students. The problem of limited 'collaboration between teachers and specialists' for the successful implementation of STT in school is addressed in seven interviews (26 segments), but also 'limited resources' (11 segments) and 'organisational difficulties' (14 segments) are mentioned, respectively, in 7 and 6 interviews. On the one hand, teachers named the necessity of support to the students to acquire the skills to use STT. But professionals also stated that teachers need support to allow a smooth transfer of the software use from the individual setting, e. g. SLT, to the classroom. On the other hand, speech patholo-

gists perceived the need for more information from the teacher about the classroom situation, to adapt their support to the students concerned. By contrast, the special needs teacher that supported the students during the writing classes did not perceive a lack of cooperation but described, instead, the constant support provided to the students during the intervention.

Indicated and/or addressed negative feeling towards dictation: The review of the current state of research in Chapter 4, Section 4.2.4 indicates potential negative emotions such as feelings of shame due to perceived stigmatisation. However, the predominant emotion named by the interviewees is frustration and anger when the transcription was not accurate. Only one speech pathologist reported feelings of shame being indicated by the student.

6.2.1.2 Supportive features for success of STT in school

The second top-level code, described in detail and illustrated in Figure 6.12, is 'supportive factors and adaptations' for the successful use of STT in school. The blue frames indicate frequently coded factors and adaptations in the environment. The red frame summarises supportive factors within the student.

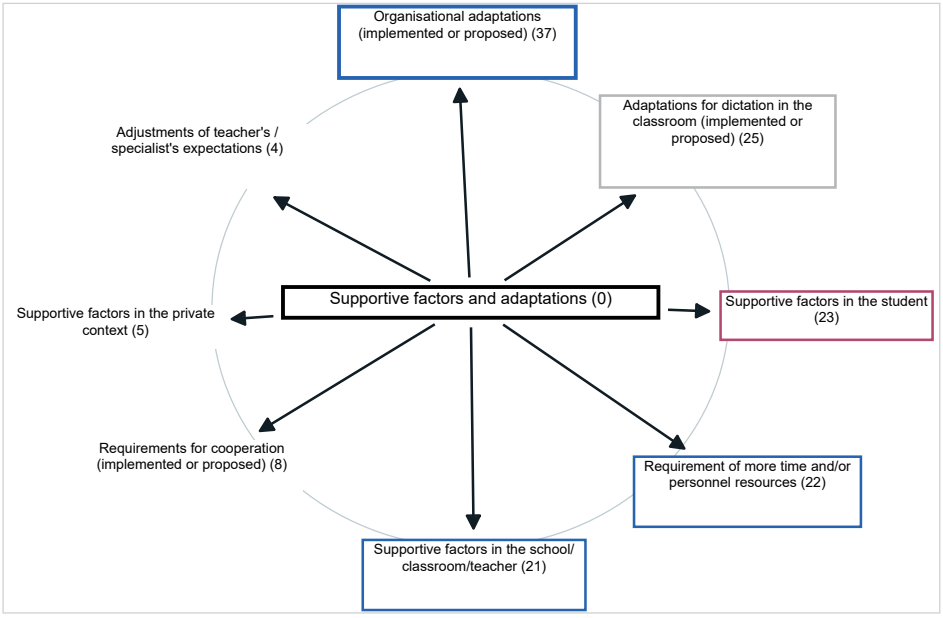


Figure 6.12: Code map of supportive features, number of codes ordered clockwise according to frequency

Organisational adaptations: Necessary adaptations regarding the organisation appeared in nine interviews. Professionals named, by 37 coded segments, what adaptations they apply or consider necessary to facilitate dictation in school. Such adaptations predominantly concern the need of a quiet place for dictating. Due to noise

in the classroom and/or the need of a private place to speak aloud, group rooms or side rooms are easily assigned to the dictating student(s). Sufficient cooperation between teachers and specialists was also mentioned as a condition for successful implementation. As mentioned above, while a lack of effective collaboration impacts the successful use of STT, interviewees observed the benefits of close collaboration.

Adaptations for dictation in the classroom: In contrast to the organisational adaptations, the 'adaptations for dictation in the classroom' code covers measures that directly support the dictation process. Both speech pathologists and special needs teachers reported a close monitoring of the single steps before and during dictation, such as providing reminders for sentence planning or clear pronunciation. In addition, the professionals largely agree on two facts: firstly, that the students need time to acquire the dictating skills; and secondly, that regular training is necessary. Two specialists reported the adaptation of the dictation process by the additional step of oral rehearsal, i. e., the instruction of articulating a sentence before using STT. However, interviewees stated that the support needed varies according to the dictating behaviour and the language skills of the students.

Some professionals spontaneously made trials with the dictating function in MS Word and reported an equal or better accuracy for the students.

Supportive factors in the student: Teachers and specialist named factors on the students' side that facilitate the use of STT in school. Not surprisingly, aspects of oral language skills were mentioned. While low oral language skills represent a challenge, higher oral language skills are estimated to be advantageous. The same applies to work behaviour, motivation, and cognition. For example, calm, concentrated and persistent working behaviour were mentioned as being beneficial. In addition, also mentioned, among supportive factors, were technological affinity and a relaxed approach to tools. One teacher who accompanied three participants described the most successful student as follows:

> *There is a child who is strong in German* (language arts)*, who chatters like a book, who has something to say. She also has imagination; she can make up a story. But she simply can't write it down, she just can't get it on the paper. And STT helped her a lot. It really helped her, she dictated things when she realised, yes, I can do it. The same applies for reading. I can listen to it; I don't have to read it. Because reading is just as difficult as writing.* (Int. 11, Pos. 9)

Requirement of more time and/or personnel resources: The topic of the significant time and support needed appeared in interviews with both teachers and specialists. Professionals from specialised schools stated that this new writing technology requires much more time and should be implemented and applied across grade levels to be successful. In addition, in the single setting, they stated that the question of time matters. However, some statements indicated that dictation may be a means to the end to foster further language skills such as oral language; or in other words, that STT may function, to some extent, as a remedial tool.

Supportive factors in the school/classroom/teacher: Two factors within the school environment were repeatedly named, with one being the classroom atmosphere. The interviewees described their respective schools as being ones that regard integration as an established priority. The second supportive factor is the knowledge of tools and ease with technology use among teachers and specialists.

6.2.1.3 Motives and attitudes of teachers/specialists

As mentioned above, the informative value in the present data of the evoked statements about motives and attitudes among teachers and specialists, i. e., top-level Code 4, is considered limited or questionable, although evoked by the interview guide and although both motives and attitudes are considered essential for the successful implementation of a tool in school. The reason for the questionable top-level code is that the length and the degree of structure of these interviews did not allow the study to delve into the professionals' notions on digitalisation, as discussed in Chapter 3, Section 3.3.3. Furthermore, participation in the study was on a voluntary basis. Thus, the predominantly favourable statements, i. e., 59 of 84 coded segments presented in Table 6.18, may only picture the likelihood of participating in such a study in the first place, rather than providing insights into challenges of STT in the classroom.

The ambivalent statements and attitudes of rejection in the present analysis concern above-mentioned questions and frustration regarding the limited accuracy for the present group of students, and the suitability for students with additional challenges such as limited oral language abilities and/or challenging working behaviour.

Table 6.16: Subcodes of motives and attitudes towards STT among teachers and specialists

Subcodes	Coded segments
Acceptance in general of tools and/or computers in schools	18
Favourable attitude towards STT	41
Ambivalent statements on the use of STT	19
Rejecting attitude/doubts towards STT	6
TOTAL	84

6.2.1.4 Benefits observed and reported by teachers and specialists

So far, the challenges and the requirements illustrate that multiple conditions must be met for the successful implementation of STT in the classroom. However, teachers and specialists observed not only challenges but also benefits. These observations are subjective and might, to some extent, illustrate the positive attitude of the participating professionals towards STT. However, the discussion below provides an overview on the professionals' perspectives on the benefits, to complete the picture of the students' experiences with STT.

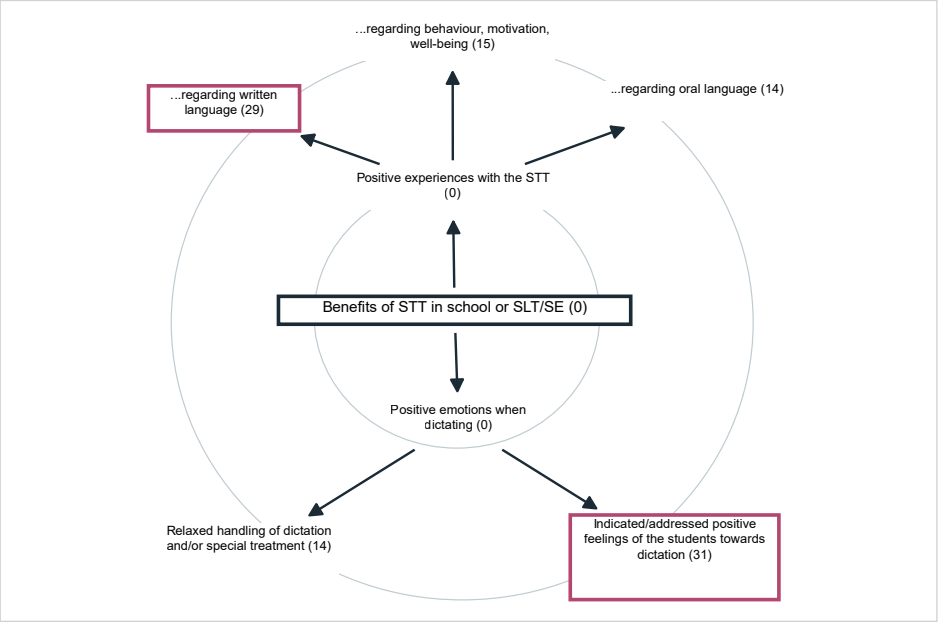


Figure 6.13: Code map of benefits of STT reported by professionals

The code map in Figure 6.13 illustrates the code structure and summarises the benefits. The upper part represents the number of coded segments of subjectively observed benefits for written and oral language and for behaviour, motivation and well-being. The lower part depicts statements by the professionals about students' emotions.

Benefits regarding written language: The professionals described, in the nine interviews, various benefits of dictating for their students, e. g. more text produced and more awareness for written language. Two interviewees stated that refusal to write and writing blockages were no longer an issue.

Positive feelings towards dictation: In nine interviews, segments of positive feelings towards dictation appeared. From the professionals' perspective, the students felt pleased with using the tool, motivated to do so, and enjoyed the writing tasks more than by hand, and some students were described as being proud of using the tool.

6.2.2 Survey of students' device preferences

All 42 EG 1 participants, i. e. including the students who dropped out during the intervention, answered the three additional questions about dictating, presented in Chapter 5, Section 5.4.4.5. Furthermore, 26 students chose to leave a voluntary additional message to the researcher. The following paragraphs show, firstly, the results of the two closed questions. Secondly, the exploration of the students' answers

about their feelings while dictating conclude this results chapter. The feedback on the additional question primarily concerned the instructional videos, which is not the focus here.

6.2.2.1 Students' tool predilection

The students answered the question about their preference of the tool by selecting one of the following answers: I prefer to write texts by hand; I prefer to dictate texts; or I like to dictate and write by hand about equally well. The answers in Table 6.17 show that, at the end of the intervention period, almost half of the group of students with dyslexia preferred writing by hand. The other half of the group had either no preference or preferred dictating.

Table 6.17: Predilection for tool

Statement	number	percentage
I prefer to write texts by hand	20	48 %
I like to dictate and write by hand about equally well.	13	31 %
I prefer to dictate texts	9	21 %

6.2.2.2 Students' own expectations for dictation in the future

The second closed question asked how often the students expected to use STT in the future. The options to select from were, "I think in my future I will dictate often; sometimes; rarely; never".

Table 6.18 indicates that close to half of the group at the end of the study expected to dictate rather frequently, while the other half thought they would dictate rarely or never.

Table 6.18: Students' expectations for dictation in the future

Expectation	n participants	percentage
I will often dictate.	2	5 %
I will dictate sometimes.	18	43 %
I will rarely dictate.	13	31 %
I will never dictate.	9	21 %

6.2.2.3 Students' feelings while dictating

The final compulsory question to the students investigated their feelings during the application of the STT. The students were asked, 'How do you feel when you are dictating?'. Their answers varied in length between one and 81 words. Each answer was coded as an entire segment by one of the following codes in MAXQDA: positive emotions; neutral emotions; ambivalent answer; or negative emotions. The distri-

bution of the answers (see Table 6.21) indicates, with 24 answers coded as positive emotions, that most students felt good or alright when dictating during the intervention study. While half of the students gave only a short answer, some provided more insight such as one student saying:

> *I find it easier for me. I need less time and can write more and then it turns out better.* (Answer 17)

However, almost 20 % did not feel good or comfortable when dictating, and another 20 % had mixed feelings about it. One student expressed his/her own feelings as follows:

> *I don't feel so good. It is difficult. I feel insecure.* (Answer 20)

A student providing a detailed answer about the perceived ambivalence of dictating explained:

> *Sometimes I feel nervous and sometimes normal. When I'm in the classroom and it's really quiet, I get nervous. When I dictate, the other children sometimes look at me. That makes me uncomfortable and then I think I've made a mistake or something.* (Answer 21)

Table 6.19: Subcodes for students' emotions when dictating

Feeling while dictating	n answer	percentage
Positive emotions	24	57 %
Neutral emotions	2	5 %
Ambivalent emotions	8	19 %
Negative emotions	8	19 %

7. Discussion and limitations

Chapter 7 comprises the discussion of each RQ, in Section 7.1. While specific limitations regarding single RQs are discussed within each subsection in Section 7.1, limitations regarding the whole study are presented in Section 7.2.

The impact of writing instruction by videos is controlled for by the additional study group. However, this thesis research focuses on the STT intervention; thus, results of the comparison between the EG 1 and EG 2 to the CG are only discussed when they were meaningful and enlighten the main question, meaning only when the writing instruction by videos are a cause of a study effect.

7.1 Discussion of research results

The results of the study are briefly summarised in the order of the RQs and discussed in the context of theory and the current state of research.

Text production is challenging in general, but due to the symptoms of dyslexia, text production is especially challenging for students with dyslexia, which represents a barrier for their academic and professional career. The need for adapted interventional or compensational measures leads to the question of whether a digital tool could overcome the barrier of dyslexia. Today, numerous digital tools are available to assist the writing process. In Chapter 3, Section 3.2, speech recognition has been argued to be the most consequent manner of circumventing the challenges of lower-order writing processes for students with dyslexia. On the one hand, this thesis study has investigated, by means of a quasi-experimental intervention study with a mixed-methods design, the effectiveness of STT for Grade 5 students with dyslexia regarding writing and writing related skills and writing motivation. On the other hand, the study has aimed to reveal the challenges for the implementation of STT in the classroom.

7.1.1 Research Question 1 – Effect of STT on text quality and text length

RQ 1A asks whether the use of STT by Grade 5 students with dyslexia is an effective method to improve the quality of their texts. The production of meaningful texts was argued, in Chapter 2, to be the main goal of acquiring writing competence. Text quality was assessed in this research using a translated and adapted version of the Test of Written Language – 4th Edition (TOWL-4) (Hammill & Larsen, 2009), before and after the 18-week intervention period. Within this RQ, the length of texts was also investigated, since text length, as part of writing fluency, is an important predictor for text quality (Y.-S. G. Kim et al., 2018; Quinlan, 2004).

The research results show no statistically significant differences in the development of text quality among the three study groups. Therefore, the hypotheses $H1_{1A}$

and $H1_{1B}$ were rejected. The graphic trend lines indicate slight improvement of text quality over time. However, this trend is statistically not significant. Furthermore, the exploration of second-order effects of participants' characteristics on the development of text quality resulted in no interaction with the treatment condition, i. e. the use of STT does not result in a different development of text quality due to the severity of dyslexia, comorbidities, gender, or the type of school in which participants were enrolled.

7.1.1.1 Text quality

Chapter 4, Section 4.2.1 has elaborated the mixed results of studies on STT and students with dyslexia or LD. Due to the mixed results presented in the current state of research, at first sight, the present research's non-significant finding for text quality is thus in line with some former findings but contradicts others. The finding is inconsistent with the increase in holistic text quality reported by MacArthur and Cavalier (2004) and Arcon et al. (2017), but consistent with the non-significant difference in the study by Quinlan (2004). However, on further consideration, the comparison of these findings is not useful to answer RQ 1A, even considering the different groups of participants. The reason for this is that the previous studies compared texts produced in two or three modalities[48], but not the effect of an STT-intervention on text quality, in other words, the transfer of learning due to STT on writing with conventional tools. This applies also to the multiple-baseline single-case study by Almgren Bäck et al. (2024), reporting an increase in variables of text complexity, which are commonly assumed predictors for text quality in primary school students with dyslexia. The only intervention study available investigating, next to TTS, also the effect of STT on conventional techniques is the study by Svensson et al. (2019). However, these authors did not capture changes in writing achievement by objective variables.

Thus, the non-significant research results presented in this thesis on the development of text quality can only be contextualised to a limited extent with respect to the state of research. This complicates the interpretability of these results. From the theoretical perspective of CLT and motivational theories, as presented in Chapter 3, the present findings are unexpected. It was argued (e. g. by Quinlan, 2004) that STT is a potential assistance to the lower-order writing processes due to the reduction of the cognitive load. The freed working memory resources could then potentially be used for higher-order writing processes and, thus, have a positive impact on text quality (Kellogg et al., 2013). However, in both Chapters 2 and 3, numerous challenges for dictation were identified. On the one hand, challenges may be present in the students with dyslexia due to comorbidities such as DLD (Cabbage et al., 2018)

48 Handwriting, dictation to a human scribe, and dictation with STT, in MacArthur and Cavalier (2004) and Arcon et al. (2017); handwriting and dictation with STT in Quinlan (2004).

or ADHD (Schuchardt et al., 2015); but, on the other hand, challenges may also be present in the environment, e. g. factors in the teacher or in the classroom, that impact the implementation of a tool in the classroom (Kerres, 2018; Petko, 2012), or in the technology itself which may, for example, be limited in the accurate recognition of children's voices (Feng et al., 2024) or in German (Silber-Varod et al., 2021). From this perspective, the findings presented in Chapter 6, Section 6.1.2 are less surprising. Regarding these two perspectives, different explanations for the rejection of the formulated hypotheses on text quality appear conceivable.

The first conceivable explanation is that the assumption about the reduction of cognitive load and, therefore, the provision of assistance for lower-order writing processes may not fit for students with dyslexia in Grade 5. Olive's (2014) parallel and cascading model illustrates the complexity of the coordination of conceptualisation, formulation and execution of writing processes. Since both the size of bursts and the pace are different in dictation than in handwriting, this could involve new and/or partly unknown challenges which increase the cognitive load and, subsequently, annihilate the assistance of the tool. Unlike a proficient writer, who is fluent in word and vocabulary use and has working knowledge of syntax (McCollum et al., 2014), a Grade 5 student with dyslexia might not be sufficiently fluent in oral language production to switch easily from handwritten to oral text production.

This explanation is supported by the qualitative research results for RQ 5 (see Chapter 6, Section 6.2.1.1). These illustrate the high demand on oral language skills, addressed in numerous statements by teachers and specialists of the participants, in the interview study in this thesis. In addition, writing with a new tool, such as typing on a keyboard, requires training (A. Sturm, 2017). The same applies to the use of STT (Bouck et al., 2011; Liu et al., 2019). As a result, the cognitive load in dictating might not be decreased for Grade 5 students with dyslexia but be different due to other processes that demand working memory resources. As elaborated in the qualitative results, further factors such as working behaviour may also challenge the use of STT.

The second explanation is that the hypothesis $H1_{1A}$ may fit but the study procedures, as well as the tool itself, may have limited the manifestation of the expected effects. The qualitative research results express clearly that the implementation of the tool for Grade 5 students with dyslexia in the classroom is challenging for several reasons, as discussed in Chapter 6, Sections 6.2.1.1 and 6.2.1.2. Three non-exhaustive reasons are presented here, as follows: the high need of training; the lack of accuracy; and the lack of individual adaptation of the AT:

- Previous research results and the findings for RQ 5 emphasise the great need for training and time for acquisition of the dictating procedure. In the present study, time limitations occurred at different stages. The limited time for the introduction of the software in the single setting, during the first school term, was possibly too short for the students. The same is possibly true for the intervention period. Galuschka et al. (2014) state that interventions for students with RD

lasting for more than 12 weeks have higher mean effect sizes than for shorter interventions – although this difference does not reach statistical significance. Referring to this, the present 18-week intervention is above the indicated 12 weeks; but the question is whether the Galuschka et al. statement applies to the acquisition of the use of STT. Regarding the treatment intensity, higher amounts of treatment, according to Galuschka et al. (2014), appear more effective. Therefore, the intended one-hour per week of writing instruction with the use of STT for EG 1 might not have been sufficient for participants to learn to use STT, especially because the students were simultaneously confronted with the content of writing instruction and the challenges in using STT.

- A further, repeatedly identified issue in the qualitative results in Chapter 6, and in the literature (e. g. Almgren Bäck et al., 2024), is lack of accuracy which might have strongly limited the effectiveness of the results. Mossige, Arendal, et al. (2023) state that, until the STT[49] improves, challenges in relying on STT remain, because of possible failed recognition of spoken words.
- The study procedure is considered to have a possible impact on the manifestation of the treatment effect on text quality and text length; among other challenges, the necessity of the individual adaptation of an AT, addressed in Chapter 4, Section 4.1, must be named. The present study might not have paid adequate attention to this necessity because it followed, as much as possible, the recommendations for an experimental study, for example by selecting a homogeneous group in terms of age, as recommended for intervention studies by Galuschka et al. (2014), or by providing a maximally comparable intervention for all participants. However, this contradicts the nature of an AT which should be individually selected, adapted and instructed (Bausch et al., 2015).

A third explanation for the rejection of the hypothesis $H1_{1A}$ might be that the hypothesis may only partially fit, meaning that STT could be an effective tool for some students but not for all students. The frequently addressed challenge of limited oral language skills, in the qualitative results, and the unsystematic picture of the individual developments in the graph of Figure 6.1, point in the direction of differences in the effectiveness of the tool depending on these skills. However, several subgroup analyses were run and did not result in statistically significant results, see Chapter 6, Section 6.1.2. This indicates that comorbidity-based subgroups on their own cannot explain the likelihood of success of applying STT. The question of whether STT is more likely to be effective for students with an exclusive diagnosis of dyslexia thus cannot be answered by the present study because the number of participants without comorbidity ($n = 23$ over all three groups) is too low to run statistical analyses with sufficient power.

49 Mossige, Arendal, et al. (2023) write about Nordic STT.

In addition, considering only comorbidities would ignore the fact that inter-individual differences among students with dyslexia – having no additional diagnosis – can be important (Almgren Bäck et al., 2024). Further features in students with dyslexia could be decisive for the successful use of STT. For example, Galuschka et al. (2014) mention that severity of dyslexia could impact the effectiveness of an intervention. Meanwhile, Cislaru et al. (2024) consider the connection between oral and written language production from a different perspective. These authors address the issue of a lack of knowledge regarding the underlying writing processes in students with dyslexia. Their investigation on written language processing, firstly, challenges the view that sentences are the unit of language processing. Secondly, these authors observe differences between children with and without dyslexia in terms of the handling of infra-sentential sequences. Thus, it is considered that a writing intervention building on oral language production should possibly focus more on underlying language processing abilities than apparent language skills or diagnosis. This could be decisive, on the one hand, to take informed decisions on whether STT is an adapted tool for the student with dyslexia, and on the other hand, to support the necessary functions to improve the oral language input and, subsequently, the STT output.

For both these second and third explanations, the recommendation mentioned in Chapter 4 for further research on effects on different subgroups applies to some extent. However, at this point, it is emphasised that the question of subgroups goes beyond subgroups according to comorbidities or severity but should also consider underlying language processing. While this thesis research design, if undertaken with a larger sample, might reveal answers to some of these open questions, a different study design could also possibly answer questions not explorable by the present design.

The question of whether its non-significant results might be a result of non-transferable, higher-order writing skills from the dictating modality to the handwriting modality remains open from this thesis research.

In summary, the rejection of the text quality hypothesis is, to some extent, unexpected on a theoretical basis. However, on the basis of the current state of research and the present qualitative results, this rejection is less surprising.

7.1.1.2 Text length

With the same rationale as for the thesis findings discussed above, the present study also investigated text length as a predictor of text quality. In contrast to text quality, the development over the intervention period results in a statistically significant increase. However, this increase does not statistically differ significantly between the three study groups. Therefore, $H1_{IC}$ and $H1_{ID}$ were rejected. Since the CG increases in a similar way, this statistically significant increase is interpreted as maturation and/or the treatment effect of the individual support that the participants of all three

groups received during the intervention period. Regarding the current state of research on text length, the same issue for interpretation applies as for text quality because the previous studies have compared text length under different conditions of text production, which in the present study is done only for RQ 3 on EG 1, that is, for the students who used STT (see Section 7.1.3). In contrast to the directed hypothesis regarding text quality, the hypothesis on text length was formulated as non-directional in this thesis, due to arguments for both an increase and a decrease in text length. Thus, the non-significant result is less surprising.

A first explanation for this finding is that the non-significant difference in text length could confirm that text length, or in other words, productivity – as one part of writing fluency – is mode-specific. Therefore, even in the case of EG 1-students increasing their dictating productivity due to the STT condition (see RQ 3), observed by Schüler (2021), Quinlan (2004), and Almgren Bäck et al. (2024), this would not increase the productivity in the handwriting modality. However, the non-significant difference between the three groups – with a significant main effect over time – indicates that the students in EG 1 improved in the same way as the CG and EG 2, despite less time on handwriting tasks. Assuming that the productivity aspect of writing fluency is mode-specific, a decrease in the handwriting productivity in EG 1 would have been expected, which was not, however, the case in the findings of the present study.

A second possible explanation is that the intensity of the intervention was too low to decrease handwriting productivity in EG 1.

A third potential explanation is that the STT condition balanced a potential decrease in handwriting productivity, resulting in a non-significant result. However. the available study data in this thesis research does not allow it to answer this question.

In sum, RQ 1A and RQ 1C on the effectiveness of STT for text quality and text length cannot, therefore, be answered exhaustively with the present results. Nevertheless, the results for these RQs are relevant. To date, most studies have compared text production by means of different modalities, such as dictating and handwriting. In contrast to those findings, the present results inform, to some extent, about the impact of STT on text quality in handwritten texts. While this thesis cannot show that Grade 5 students with dyslexia make greater progress in text quality and text length due to the STT intervention, its findings do indicate that this specific group of students does not stagnate in their development of text quality and text length due to the STT condition. As the quantitative results for RQ 1 indicate, and argued in Section 7.1.1.1, the time and intensity of the intervention might be a factor affecting the manifestation of an effect. The same applies to text length. It thus cannot be excluded, at this point, that a longer and/or more intense setting would not have resulted in a different finding.

Meanwhile, the qualitative results about benefits of STT, and about the students' answers to the question of how they felt about using STT, indicate that STT appears

suitable for some students, not only for text production but also for oral language skills or for overcoming writing blockages. Both components together, imply, firstly, that Grade 5 students with dyslexia continue to develop their skills in text production, and secondly, that testing STT for students in the individual and/or classroom setting might reveal benefits to students and professionals.

At this stage, the contextualisation regarding the instructional videos and their effects on higher-order writing skills, i. e. RQ 1B, seems appropriate. The aim of this instruction was, firstly, to make the writing instruction for the participants in EG 1 and EG 2 comparable. Secondly, the videos allow to ensure evidence-based instruction, necessary for the effective use of STT according to MacArthur and Cavalier (2004). The manual chosen as the basis for the videos was considered fitting this requirement because of its effectiveness for higher-order writing skills (see Section 5.5.3). From this perspective, a greater improvement in text writing over time in the EG 1 and EG 2 would have been expected. However, as elaborated in Section 5.5.3, the revised version of the instructional videos did not contain the explicit instruction of self-reflexion. This could be one reason, why no significant positive development in writing skills was observed in groups EG 1 and EG 2 over the intervention period.

7.1.2 Research Question 2 – Effect of STT on writing and reading fluency

STT might affect both writing and reading fluency (see Chapter 5, Section 5.2). Writing fluency was assessed by the number of correct word sequences (CWS) in a five-minute writing sample (Chapter 5, Section 5.4.2). Reading fluency was assessed by the SLS 2–9 (Chapter 5, Section 5.4.3).

7.1.2.1 Writing fluency

The comparison of the writing fluency measure at pre- and post-test shows a statistically significant main effect over time in all three groups, meaning that all groups increased their writing fluency in the handwritten modality during the intervention period. However, there is no statistically significant interaction between the three groups over time, i. e. no difference between the three groups. The non-directional hypothesis, $H1_{2A}$, assuming a difference in the development of writing fluency, was therefore rejected, since the STT condition had no significant effect on the development of writing fluency. In other words, the use of STT, compared to writing by hand, is not a statistically more or less effective method to improve writing fluency.

As for text quality and text length, the same problem of comparing the present results with the current state of research applies to writing fluency. The only prior study investigating the effect of the application of STT and TTS on reading and writing, in the conventional modalities, among students with dyslexia is Svensson et al.'s (2019), which did not capture objective measures for writing. Regarding writing

fluency, this subsection discusses the combined variable of writing fluency, comparing both accuracy and productivity.

The non-directional hypothesis for writing fluency indicates that, on the basis of theory and the current state of research, both a positive and a negative effect of STT on writing fluency are conceivable. Due to this open-ended hypothesis, its rejection in this thesis research is not surprising. However, conceivable explanations for the present study findings are discussed here, in order to deduce implications for future research and practice.

Theoretically, it was argued and hypothesised in Chapter 5, Section 5.2.2 that the exposure to correct spelling on the screen might impact writing fluency in a positive way. At least for the handwriting modality. However, this hypothesis does not fit the analysis of data in this thesis.

A first potential explanation before this finding is that the already-mentioned mode specificity, meaning that the students only improved in the trained modality, i. e. handwriting for EG 2 and CG and dictating for EG 1. However, to determine this, evidence about the dictation mode is necessary, which is aimed at by RQ 3 (see Section 7.1.3). In the sense of mode specificity, the STT condition could affect writing fluency negatively; but the present results, with no significant difference, do not speak clearly in favour of a negative impact of dictating. Nevertheless, trendlines in the graphs presented in Chapter 6, Section 6.1.3 can be interpreted differently, and the question arises of whether handwriting fluency stagnates when the time spent writing by hand decreases, for example due to dictating. A difference in the slopes of the *M* values of three groups in Figure 6.4 can be perceived, with a flatter slope for EG 1 than for the two other groups. Thus, EG 1 improves less in writing fluency than EG 2 and CG. Although not significant in this study setting, a different possible result with, for example, a more intense and longer study setting cannot be excluded. Thus, further research should pay attention to writing fluency and its development over longer and/or more intense interventions, to allow more informed decisions for practitioners and people with dyslexia (see Chapter 8).

A second explanation, this time supported by the qualitative analysis presented in this thesis revealing many challenges such as accuracy during the intervention, could be that EG 1 was, to some extent, distracted by the acquisition of STT. This distraction could possibly have hampered the development of writing fluency. In addition, writing instruction by video might have impacted the instructional focus on higher-order processes. This explanation is supported by the slightly different slopes between EG 2 and CG, with CG improving the strongest in writing fluency. However, the present data analysis was not able to answer these questions, which indicates the need for further research.

In sum, students that use STT might not increase their writing development in the handwriting condition in precisely the same way as the students writing by hand. It is considered important to monitor this in future research. At the same time, the difference in writing fluency development is not significant. On the one hand, it

is thus concluded that the use of STT does not harm students using STT, which is relevant knowledge for practitioners. On the other hand, only considering writing fluency in the handwriting modality does not, in this case, provide a complete picture, which is why RQ 3 must be further considered in relation to RQ 2A.

7.1.2.2 Reading fluency

The results for reading fluency show a similar pattern as those for writing fluency. In all three groups, reading fluency measures show a statistically significant increase between pre- and post-test, but no difference between the three study groups. Regarding reading fluency, the trendlines point in a similar direction. Therefore, $H1_{2B}$ was also rejected. This result is unexpected, since, based on theory, a directional hypothesis was formulated, arguing that the time spent reading is increased due to: 1) the revision with TTS, assuming more revision on the screen than on paper; and 2) potentially, simultaneous reading during dictation.

Regarding the current state of research, the rejection of the $H1_{2B}$ is difficult to contextualise, since no studies are found for an STT-intervention on reading fluency. Although the use of the TTS function was evoked by the introduced revision procedure, the intervention did not focus on TTS. Therefore, studies such as the meta-analysis by Wood et al. (2018) on TTS and RD are not meaningful for discussing the present RQ 2B. Nevertheless, the research literature and qualitative results do indicate a possible explanation for the unexpected result, as follows.

Firstly, the assumption of a stronger increase in reading fluency in the EG 1 is based on the possibly supporting revision process in writing and increased time spent reading (Taylor et al., 1990). However, revision is emphasised by Hayes and Olinghouse (2015) as very challenging for developing writers and barely applied by beginning writers. Almgren Bäck et al. (2024) state, for example, that students' accuracy at a word level is, in fact, decreased by revision. The same finding is reported by Quinlan (2004) who investigated text revision and observe that text quality decreases when students apply more revision. Quinlan (2004) suggests that revision is a specific challenge in using STT. Therefore, this challenging process might not free resources but, rather, increase extraneous load or distract attention.

Secondly, the accuracy issues emphasised in the qualitative data analysis (see Chapter 6, Section 6.2) and the current state of research (see Chapter 4, Section 4.2.5) could explain, to some extent, why $H1_{2B}$ does not fit the present sample. If the students were occupied and distracted by accuracy issues, this could have impacted the theoretically conceivable positive impact of STT on reading fluency.

7.1.3 Research Question 3 – Handwriting vs. dictating

Several authors investigating STT in the educational setting have found differences between modalities (e. g. MacArthur & Cavalier, 2004; Quinlan, 2004; Schüler,

2021). RQ 3 investigates such differences between the modalities of handwriting and dictating, i. e., productivity and accuracy differences, and regarding the development of productivity and accuracy between pre- and post-test, as investigated before by Almgren Bäck et al. (2024). The findings are discussed in the following subsection regarding the differences between modalities and in Section 7.1.3.2 regarding the development of text production in the dictation modality. Productivity was measured by the number of words written by hand (TWW) or dictated (TWD). The accuracy was measured by %CWS and CWS.

7.1.3.1 Comparison of modalities

The comparison of the fluency of text production shows, firstly, that the dictated texts outperform handwritten texts in terms of productivity, i. e. the EG 1 students ($n = 36$) dictated statistically significantly longer texts than they wrote by hand. This difference in favour of dictated texts has, at pre-test, a moderate effect, and at post-test a moderate but almost strong effect. Secondly, the dictated texts are more accurate, with statistical significance, than the handwritten texts. This difference in fact shows a very strong effect. Therefore, both hypotheses, $H1_{3A1}$ and $H1_{3B1}$, are accepted. Unlike the previous results, these results can be contextualised more clearly in relation to the current state of research as they also compare two modalities.

The acceptance of these hypotheses is in line with the results from the one available German study, by Schüler (2021), which finds longer and more correct texts with the assistance of STT. However, the participants in the Schüler (2021) study were typically developing Grade 8 students, which limits the comparison with Grade 5 students with dyslexia to some extent. The present research findings are consistent also with Quinlan (2004), showing that the less fluent middle school participants dictated both longer texts and more accurate texts than they wrote by hand. Meanwhile, findings by MacArthur and Cavalier (2004) are in line with those in this thesis regarding accuracy. Those authors found students with LD committing significantly fewer errors in the STT-condition, but in contrast, both MacArthur and Cavalier (2004) and Kraft et al. (2019) found no increase in text length. These inconsistent findings are especially surprising, as the sample of the study by Kraft et al. (2019) comes closest to the sample of the present study. The latter authors investigated texts dictated by 10–12-year-old students with spelling difficulties.

In summary, the present findings are in line with those in previous studies regarding accuracy and, thus, confirm that students dictate more correct texts with STT than they write by hand. This is relevant, since text length is important for text quality (Y.-S. G. Kim et al., 2018; Troia, 2006). Regarding text length, the present study is in line with some but not all prior studies. However, the present research findings are relevant since more accurate texts are easier to understand, which therefore supports the communicative aspect of writing. Especially for students with dys-

lexia, this aspect can be an important advantage in terms of participation and social inclusion.

7.1.3.2 Development of dictating fluency

The results for RQ 3B are mixed. Texts dictated by EG 1 at post-test are significantly longer than texts dictated at pre-test, showing a small but almost moderate effect size. Thus, this first alternative hypothesis, $H1_{3B1}$, is accepted. By contrast, the alternative hypothesis, $H1_{3B2}$, is rejected, because the %CWS in dictated texts at post-test remain stable at approximately 80 % of correct %CWS. However, it is important to interpret this latter with caution (Winkes & Schaller, 2022) because, as elaborated in Chapter 5, Section 5.4.2, the measure of %CWS is not an absolute value but a value relative to the amount of text. It is therefore summarised that, over time, accuracy develops along with the increase of text length. This is illustrated by an example: a participant produces at pre-test 100 words containing 80 CWS; at post-test, the same participant produces a text with 150 words containing 120 CWS; thus, while the %CWS has remained stable, the absolute number has increased at the same rate as the TWD.

The multiple-baseline single-case study by Almgren Bäck et al. (2024) allows this thesis to contextualise the results in relation to the development within the research literature – despite different study designs – because Almgren Bäck et al. investigated dictating productivity over time in students with dyslexia. The results in this thesis are consistent with Almgren Bäck et al.'s (2024) results on dictating productivity. Regarding the accuracy in dictated texts, Almgren Bäck et al. (2024) found a tendency of an overall increase in accuracy on a productivity dependent level, as measured by CWS. The findings in this thesis confirm this increase in the productivity-dependent score of CWS but also add the result of the maintenance of the accuracy in a productivity-independent measure, as %CWS. These additional results about the stability of %CWS indicate that the students affected by dyslexia might benefit from further support to develop higher dictating accuracy, e. g. from an instruction in the effective use of a spell checker.

Descriptive results presented in Chapter 6, Section 6.1.3, illustrate differences in the development of dictating in connection with participants' characteristics. While AD(H)D appears, against expectations, to not hamper the development of dictating fluency, DLD appears more critical to the success of STT as an assistance for these students.

In sum, the present research findings are considered relevant for further research and practice. As elaborated in RQs 1 and 2, EG 1 develops in a comparable way to the CG and EG 2. Although text quality in dictated texts was not investigated in the present study, these thesis results show that the students with dyslexia make significant progress in the dictating modality. While the STT-intervention was interpreted as equal to conventional writing in terms of the development of written

language skills in the handwriting condition, the results indicate that students with dyslexia can progress in their dictating fluency with the use of STT. While these results cannot confirm that STT is a suitably adapted tool for learning for students with dyslexia, they do show that STT does assist students in their text production. Therefore, the available data inform the discussion on compensation for disadvantages, as outlined in Section 2.2.4.4, since STT enables students to produce longer and more accurate texts, and the data on text length indicate continuous progress rather than stagnation.

7.1.4 Research Question 4 – Writing motivation

Theory and research literature largely agree on writing motivation affecting and predicting writing performance (see Chapter 3, Section 3.3.2). This appears especially important for students with disabilities (S. Graham et al., 2018). RQ 4 asks whether the STT intervention has an impact on writing motivation according to the writing motivation model by Wright et al. (2019). Therefore, the translated and adapted 'Self-Beliefs, Writing-Beliefs, and Attitude Survey' (SWAS) was applied.

None of the findings for the two RQs about writing motivation, i. e., 4A and 4B, fit the hypotheses, since there are no interactions over time; and between one of the two interventions, i. e. use of STT or writing instruction, neither shows a statistically significant result. In other words, neither the treatment condition nor the writing instruction by videos have an effect on writing motivation. The hypotheses, $H1_{4A}$ to $H1_{4B}$, are thus rejected. However, the descriptive data and graphs show different trendlines for EG 2 than for EG 1 and CG. This difference in the trend, which is considered to be related to the writing instruction, is discussed after the question on the effect of the STT condition at the end of this subsection.

These results are unexpected, regarding the theory and state of research, because personal empowerment affects motivation (Bandura, 1997) and a positive emotional environment enhances writing motivation. However, the interview data and the questions to the students may provide an alternative explanation for these unexpected results. These qualitative data indicate that the STT intervention did not, or doesn't yet, provide sufficient support to the students with dyslexia to reduce the cognitive burdens placed on them, as expected by J. M. Sturm and Rankin-Erickson (2002) for an intervention with a digital technology. Therefore, students most likely felt not empowered by the tool, and thus, motivational components such as self-efficacy did not increase. The reason for this lack of positive feelings or experiences might be explained by the qualitative results, presented in Chapter 6, Section 6.2, which emphasise various user experiments and numerous challenges experienced by the students in dictating. For example, the dictating procedure and the output accuracy of the software most likely caused frustration related to STT. At this point, both the lack of empowerment, and negative user experiences due to challenges related to the tool, are considered decisive for writing motivation, which may explain

why no effect of the intervention on writing motivation is manifest. To summarise, the formulated hypothesis, $H1_{4A}$, might, on the one hand, theoretically fit RQ 4 once the STT-intervention is less affected by numerous challenges.

However, on the other hand, the formulated hypothesis simply might not fit the RQ 4 because the STT intervention generally might have no impact on writing motivation. This would contradict the findings of increased writing motivation and general school motivation by Svensson et al. (2019), but confirm, to some extent, the unidirectional impact of writing motivation described by S. Graham et al. (2007) and Yeung (2022). It can be concluded, therefore, that writing motivation, i. e., writing attitude, affects writing achievement but writing achievement does not affect writing motivation. Furthermore, the limitations of the translated and adapted SWAS to assess writing motivation, discussed in Chapter 5, Section 5.4.4.4, is recalled. This scale, with a low consistency of self-efficacy and a purposefully limited task specificity, possibly does not fully account for effects on writing motivation due to the intervention in this thesis study.

In sum, the present results on STT and writing motivation raise new questions and are considered as not satisfactorily answering RQ 4. More research is thus considered necessary, on the one hand, on writing motivation and, on the other hand, on STT and writing motivation with a less challenging context, due to low accuracy.

Regarding the RQ 4B, the $H1_{4B}$ is rejected, as stated above. However, the results in Chapter 6, Section 6.1.5 show that EG 2 differs in the direction of the slopes (see Figure 6.10). As the interpretation of trends harbours the risk of overestimating the effects (Schröder et al., 2022), the trends are only pointed out here but without a detailed elaboration of an interpretation: In contrast to the other two groups, motivation increases slightly in EG 2. Therefore, the question is whether this different trend is related to the writing instruction by video. It is conceivable that this writing instruction type only had an effect in EG 2 over the 18-week intervention because EG 1 was too occupied by the STT intervention. The decrease in writing motivation in the CG, visible in the graph, points in the direction of decreasing writing motivation, as described in the literature Chapter 3, Section 3.3.2 and by Pajares (2003) regarding self-efficacy and by Ekholm et al. (2018) regarding writing attitudes. However, as stated in Chapter 5, Section 5.4.4, single components are not considered in this thesis research. More research on both the overall writing motivation and on single components of writing motivation is needed to provide a starting point on writing motivation of students with dyslexia.

7.1.5 Research Question 5 – Challenges of implementation

This thesis research emphasises that the intention of investigating the effectiveness of STT for students with dyslexia in the classroom setting harbours challenges because the success of the implementation of a tool in the classroom, especially an audible tool such as STT, depends on numerous factors. RQ 5, therefore, addresses

the challenges accompanying the implementation of STT in the classroom via an exploratory approach of semi-structured interviews with professionals from EG 1 and questions to participants in EG 1. The data of this qualitative supplemental component reveal challenges and benefits, and confirm some of the expected issues, when students with dyslexia are using STT in the classroom. The following discussion contextualises these results in relation to earlier research findings and the results of RQs 1 to 4.

The qualitative analysis of the interviews emphasises three major topics to answer RQ 5. The first and the third top-level codes reveal, firstly, difficulties met with STT in school or SLT/SE, and secondly, adaptations, factors and resources needed or applied to successfully implement STT in the classroom. The third topic results of this analysis are reported benefits of STT by professionals. The answers from participants in EG 1, firstly, confirm, the difficulties described by professionals such as low output accuracy, and secondly, illustrate the mixed user experiences in statements about the students' emotions.

These topics and reported mixed experiences are, to some extent, in line with the findings by Almgren Bäck et al. (2023). Their five-year, follow-up interviews with students with dyslexia also emphasise mixed user experiences, and describe enabling and hindering factors, many of which are the same as revealed in the present thesis findings. Although the interviewees and the focuses of the interviews in these studies are not identical, both reveal the importance of support. However, in contrast to the interviews in the present thesis with the professionals, the interviews with the adolescents by Almgren Bäck et al. reveal more issues interpreted as perceived stigmatisation, summarised by the authors as emotional responses to the learning environment, such as non-acceptance of the disorder or a lack of acceptance of the assistive tool employed. The challenges revealed by professionals in this thesis research focus more strongly on students' characteristics that influence the success of STT.

In general, the topics revealed by the present research interviews are consistent with possible challenges that were, to a certain extent, expected based on the earlier theoretical discussion and literature review in Chapters 2 to 4. Both working behaviour and oral language abilities were highlighted, as well as numerous issues with the accuracy of the STT output. On the one hand, the oral language abilities of the students with dyslexia and accuracy issues of STT might be related. On the other hand, alternative explanations for low accuracy, such as being due the speaking style (Feng et al., 2024; Silber-Varod et al., 2021) or higher pitch of children (Higgins & Raskind, 1995), lower-performance due to the language, i. e. German, or due to the ASR system (Feng et al., 2024; Silber-Varod et al., 2021) used in the present study, are conceivable, but these can neither be confirmed nor rejected by the present qualitative results.

In sum, the topics revealed and highlighted in this thesis are considered relevant for three reasons. Firstly, they inform on implications for the successful implementa-

tion of STT in the educational environment (see Chapter 8). Secondly, they support understanding of the above-presented research findings, as discussed in Sections 7.1.1 to 7.1.4. Thirdly, the findings of the qualitative and the quantitative approaches, together, indicate limitations of this thesis study (see Section 7.2). At the same time, these findings may also point in the direction of further research. For example, the descriptions of benefits in this thesis might be of interest for future research to investigate the remedial effect of STT on oral language production.

7.2 Limitations

Limitations of the assessment tools and the study materials in this study were discussed in the respective sections in Chapter 5, Sections 5.4 and 5.5 The discussion in Section 7.1 also mentions some limitations in specific contexts. However, there are also limitations that affect the study as a whole, which are addressed as follows.

It was argued in both the Chapters 2 and 3, that the production of texts is the goal of writing. Text production, i.e, text quality, was assessed in this thesis study by narrative texts written by hand. However, this research was unable to investigate the impact of STT on text quality in dictated texts. This is considered a major limitation in trying to answer the question of whether STT is a suitable tool to compensate for barriers of dyslexia. The findings show that the produced texts are longer and more accurate than texts written by hand. However, these findings give no indication of whether the dictated texts outperform handwritten texts in terms of text quality, such as in number of ideas or structure. This study investigated the application of a tool in the educational field with a limited number of students with dyslexia in each class. Therefore, numerous teachers and specialists were involved, which increases the possible amount of variation, for example, in instruction and amount of support time. The study could not control for certain factors, which might limit the reliability of the study. Not documented, and thus not controlled for, were the writing lessons in the classes, i. e., the content and amount of writing instruction in the CG and the exact execution of writing instruction based on the videos in the experimental groups. Regarding the individual setting, for example, the therapy time was recorded, but the exact content and focus of the TAU were not included in the analysis.

The qualitative results illustrate the complexity of investigating the effects of a tool which is burdened with challenges. In addition, the present sample of students with dyslexia, by definition, entails difficulties that can have a negative impact on speech recognition. The observations of these challenges in the educational setting are, on the one hand, an important contribution of the study. On the other hand, these challenges, first and foremost the accuracy of the STT output, have an influence on the effects, and are thus difficult to interpret. This difficulty in interpretation is considered as lying in lacking a distinction between cause and effect. In other words, the accuracy issues may, to some extent, be the reason why no effects were detected and why some students struggled with the dictating procedure. Therefore,

the validity of the study could be decreased because these nil results might reflect, to some extent, accuracy issues rather than the effect of dictating for the present sample. However, the recognition difficulties may also be the result of the oral language difficulties of this specific group.

Therefore, further research is needed to investigate the potential of STT for students with dyslexia, ideally with a tool that bears fewer challenges on the technological side. Regarding the accuracy issues, retrospectively, an accuracy measurement at pre-test is considered appropriate. Firstly, participants with too low accuracy could have been excluded from the sample. Secondly, the pre-test measure could possibly have informed about a minimum accuracy level necessary for the successful application of STT. Arcon et al. (2017) did assess in their study the accuracy of speech recognition for each participant and subsequently excluded those with low accuracy, defined as between 50 to 59 %. However, the literature provides no recommended cut-off value.

As stated in Chapter 5, Section 5.4, no norm values were available for the variables investigated. The only standardised test used, i. e. the SLS 2–9, did not allow this thesis research to compare the results to a norm sample, due to a considerable part of the outcome values being below the norm values. Therefore, this study relied only on non-standardised assessment tools, which could, in the case of significant results, lead to an overestimation of the effect sizes.

There may be further limitations due to the study setting. The effects of the STT intervention in the classroom, and for the specific group of students with dyslexia, imposed some requirements on the study setting. Some of the necessary adaptations are considered as limiting the effectiveness of the study, others as enhancing it; namely, recruiting participants meeting the inclusion criteria and with professionals committed to participating in the study – teachers of EG 1 and EG 2 had to participate with their class – required a setting that did not exceed the workload of teachers and the frame of the curriculum. Therefore, the intensity of the intervention was set to one hour per week for 18 weeks, and instructional videos were developed to instruct the students requiring a minimum preparation time for professionals. While the instruction of the students by videos was considered suitable, the intensity of the intervention might have been too low to be effective. Furthermore, the question is raised whether the professionals were sufficiently knowledgeable to instruct and support the students adequately with the tool.

8. Conclusion and implications for future research and practice

The following sections provide, firstly, a general conclusion of the findings with a critical reflection on the study, focusing in particular on the relevance and generalisability of the findings. Section 8.2.2 outlines implications for future research; followed by implications for the educational and therapeutic practice in Section 8.2.3.

8.1 General conclusion

This thesis research has aimed at identifying the effectiveness of the use of STT for Grade 5 students with dyslexia and has explored challenges in the implementation of STT as an AT in the educational environment. The summarised answer to these two overall research questions, based on quantitative and qualitative analyses and in line with the definition of AT by D. Edyburn (2015), (see Chapter 3, Section 3.2.3), is as follows:

> STT can be applied by students with dyslexia as an assistance in the digital text production to produce longer and more accurate texts. As a learning tool to acquire writing skills and increase writing motivation, however, STT is, to date, not a more effective approach for students with dyslexia than conventional methods.

Furthermore, the results indicate that STT can be an effective tool for the development of oral language and literacy skills for some students with dyslexia, but they also reveal numerous challenges to reach this stage. Certain challenges can be overcome with the necessary knowledge and resources. However, other challenges may indicate that STT in the individual setting of a student with dyslexia is not the right support, or not at specific stages, for example in earlier school grades.

Before discussing the impact of the findings, however, it is important to emphasise that, for three reasons, these results can only be generalised to a limited extent. Firstly, due to further expected technological advances of the ASR systems, this study gives only a picture on the current state of their development. Secondly, the results apply to the complex classroom setting of this study, as was described earlier in this thesis. An individual study setting is considered less challenging than the classroom setting and might, thus, generate different and potentially better results. Thirdly, in terms of age, the present, intervention study was conducted with a homogeneous group of students. The age group was primarily chosen for practical reasons, relating to the feasibility of the recruitment of the study sample, and with sufficient statistical power. The present findings provide, therefore, information about the effect on this group. Factors such as the amount of text that must be produced in higher classes or the generally increased use of digital devices from secondary school onwards

(Schröder et al., 2022) might impact the perceived need for an assistive tool or the ease to use it.

However, the strength of the quantitative findings of this study is considered in presenting a current state of the impact of STT for the investigated group of students up to the present time. A further strength of these thesis findings is the provision of a considerable amount of information about the challenges in implementing STT in the classroom, due to the complex study setting. Knowledge about these challenges can, on the one hand, be applied for future guidelines and more efficient support in the field. On the other hand, this knowledge can also be used for improving future research settings in this field. The following discussion reflects on the extent to which the present results extend, confirm or question the current state of research.

Two relevant differences of the present study, compared to previous studies, lead to an extension of previous knowledge in the field. The first difference to most cited studies concerns the aim of the study and its outcome. To date, studies on STT mostly compared the effect of the tool in different modalities, i. e. handwriting, typing, and dictating. Although the present study also investigated, in RQ 3, the modality-specific effect of STT, the main study focus has lain on the modality-independent impact on writing performance and writing motivation. This modality-independent impact on writing is interpreted as the impact of STT as a learning tool on conventional writing skills. The second relevant difference concerns the setting and the focus on writing. While Scandinavian studies are, to some extent, comparable because they also investigated the specific group of students with dyslexia and also less resourced languages than English, i. e. Swedish and Norwegian, these studies either did not investigate STT in the classroom, i. e. Almgren Bäck et al. (2024), or did not capture objective writing skills, i. e. Svensson et al. (2019).

The present findings can extend previous research by the finding that STT, at its present stage and applied in writing classes, is not, or not yet, a tool to generally improve writing development with conventional tools because statistically, the method applied in the setting of this study is currently neither better nor worse than conventional support. Therefore, at present, STT cannot remove the barrier of dyslexia for learning to write. However, regarding the sample of students with dyslexia, this thesis research shows that, despite comorbidities, these students develop similarly to the others. Therefore, the STT condition does not negatively affect the development of writing due to comorbidities.

A further extension of the current state of research is that Grade 5 students with dyslexia can use STT in a mid-resourced language such as German and, despite a diglossic environment, without stagnating in their development of literacy skills, since statistically no differences are found compared to the development of students' literacy skills in conventional writing modalities such as handwriting and typing.

So far, these conclusions to the thesis findings are presented from a perspective that is generally orientated towards the values, processes and products of a 'book culture', as Steinhoff (2023) calls this perception of writing. The setting of the present

research, with the question of improving conventional writing skills through dictation with STT, can however, from this perspective, be fundamentally questioned because this reinforces the view that the computer or STT is used as a learning medium in which the hardware or software only play a secondary role. Steinhoff (2023) questions the persistence of this view of writing because it limits the acknowledgement of the computer and the environment as artefacts that are part of the writing process, which he calls 'co-active'. Considering that the present study's RQs 1, 2 and 4 investigate STT as a learning medium, RQ 3 thus has aimed at investigating digital text production, as possibly more appropriate for the digital writing environment. However, while the 'co-active' approach between the human and the computer (Steinhoff, 2023) is considered of great interest for research and practice, the importance of understanding a digital writing tool with a broad perspective appears important as a starting point and in order to be able to take informed decisions. Therefore, after these conclusions on the impact of STT on writing in the 'book culture' perspective, the following discussion addresses writing with the focus on digital text production.

Regarding the finding of significantly longer and more accurate texts in the dictating modality, this confirms previous findings, for example by MacArthur and Cavalier (2004) in English and with students with LD. Thus, STT can assist students with dyslexia in German in their digital text production. This confirmation is relevant for two reasons. Firstly, since writing fluency is important for text quality (Y.-S. G. Kim et al., 2018), STT can be a support for students with dyslexia. In addition, this can be important for the development of writing because longer texts allow teachers to provide more feedback on the writing product (S. Graham & Harris, 2019b). Secondly, more accurate text is easier to understand for the reader. This improves the communicative aim of writing and might, in the end, have an impact on participation (Shanahan, 2019).

The present findings, investigated through a larger sample, also confirm those in the multiple baseline single-case study by Almgren Bäck et al. (2024) suggesting that students improve text length over time when they dictate. Although this effect is not transferable to other conditions, this finding suggests that students who train on dictating improve in their text production in terms of text length. However, as discussed in the limitations (see Chapter 7, Section 7.2), this thesis cannot contribute to the state of research on text quality in dictated texts.

The present intervention study did not focus on writing fluency or on the question of transferability of writing fluency. However, these findings do not challenge the statement by Hayes (2012) that writing fluency is mode-specific.

8.2 Implications for future research

The already named lack of information about the development of the text quality in the dictating condition in this thesis research is considered one of the main thesis implications for future research. However, numerous further implications result

from the present research analyses of both the qualitative and the quantitative data. This subsection outlines some of these.

The qualitative results have emphasised characteristics in the oral language of the students as one major challenge in the application of STT. Although the present study controlled for several factors, i. e. comorbidities, severity of dyslexia, bi- or multilingualism, gender, and LD, the findings do not allow a distinct determination of a cause by one of these factors. Nevertheless, in the sample of EG 1 to answer RQ 3, compared to students without DLD, students with DLD showed a weaker increase in the development of dictating, as to what could be assumed as a risk factor. However, factors influencing the success of dictation need further consideration. The call by Schüler (2020) and Quinlan (2004) for more research regarding effects on different subgroups of students is supported by these thesis results. However, perhaps subgroups by diagnostic criteria are less meaningful than differences in language processing. For example, the coordination of written language processes as described by Olive (2014), put in relation with oral language processing, might be a starting point for future research on the application of STT by students.

Working behaviour is also frequently identified in this thesis as a critical point, not only in the dictation process but also in the oral language planning and the revision process. Further knowledge about essentials of dictating behaviour might therefore be useful in decisions taken on for whom STT is promising and to inform practitioners to support students in their dictating behaviour.

As mentioned above, this thesis didn't sufficiently capture teachers' notions about digitalisation or their motives and attitudes toward digitalisation or put these in relation to the effect of the tool. This could be an interesting point to investigate, being at the frontier of media education and research in special needs education on the effects of an AT as STT.

Questions also remain unanswered regarding writing motivation. As discussed above, the present findings contribute less than expected in relation to the current state of research and key issues in this field. However, writing motivation as an important predictor for writing achievement is considered to be under-researched for the group of students of dyslexia but also in relation to STT.

The quantitative findings suggest that STT is no more effective than conventional methods in the acquisition of written language. However, on the one hand, frequent accuracy issues in the present study might have impacted the results. On the other hand, further technological advancements in terms of accuracy for various user groups can be expected. Therefore, a later study replication, but also the investigation of a different user group in terms of age, might show different results. In the context of technological advances, the question arises as to how AI-based writing tools employing LLMs, such as ChatGPT, affect the text production of students with dyslexia. The potential benefits and challenges of these tools for written language acquisition and for supporting various writing processes also require comprehensive research to fully understand their impact and to guide their effective implementa-

tion. The quantitative findings in this thesis also raised the question of the specificity of writing fluency. Future research in the field of STT might inform about handwriting, and vice versa.

The evaluation of minimum accuracy, discussed in Chapter 7, Section 7.2 as a limitation of this study, is considered as an implication for future research to enhance practice. This thesis research did not set an initial minimum accuracy level. However, regarding the question of evaluating for whom STT is promising, the accuracy rate might be an important variable to investigate in terms of research and practice. The provision of indications such as a necessary accuracy rate to start working with STT, or the accuracy level adapted for implementing the tool in school, would be considered helpful for practitioners.

8.3 Implications for practice

The present research findings are mixed in terms of effectiveness of STT as a learning tool or as an assistive tool. In addition, the user experiences are mixed. There are reasons in favour, and against, the application of STT for students with dyslexia, and an individual evaluation of this question is needed. The findings presented in this thesis can generally justify the application of STT as an assisting tool to produce more text and more accurate text. In addition, STT can be the method of predilection for certain students in individual cases, for example motivated by a specific career aspiration or the desire for more independence, although STT does not outperform conventional writing techniques in terms of writing acquisition. As presented in Chapter 6, Section 6.2, different types of challenges must be considered. Some challenges are within the students, especially the working behaviour and the oral language skills. Regarding challenges of the school setting, the necessary resources and willingness to support a student in the application of the tool are necessary. According to Almgren Bäck et al. (2023), acceptance of a tool, and of the difficulty due to dyslexia, must exist for both the individual and the school, to successfully use a tool in the school setting.

If a decision is taken to use, or assess the use, of STT, specialists and teachers need to know the challenges involved and to adapt to those. A speech pathologist, for example, might introduce the software as adapted to the student's oral language skills in order to simultaneously foster language awareness and accurate text production by dictation. Meanwhile, a teacher might have to plan their lessons in a way that allows a dictating student to receive the content instruction and to dictate in a separate space with a sufficient amount of time.

However, to date, no guidelines such as those described by Bausch et al. (2015) in the USA are known for school teams in Switzerland, to carefully evaluate and implement STT in school and to define responsibilities. The development of such guidelines might be a necessary step at a stakeholder level to improve the use of individually suited writing tools for all students, and in preparing students with dys-

lexia for their future academic, professional and everyday life, where writing plays an increasingly important role.

9. Summary

Writing is a key competence for academic and professional career as well as for everyday life. However, writing competence is challenging to acquire. Based on three models of writing, the complexity of writing processes, their coordination, and the numerous factors within the writer and in the writing context that play a role, have been illustrated. However, the acquisition of writing is challenging, especially for people affected by dyslexia (Schulte-Körne & Galuschka, 2019). Therefore, writing represents a barrier for those affected. Dyslexia manifests as primary, and often with additional secondary, symptoms. These symptoms can be intensified or persistent due to comorbidities which are frequently present in people with dyslexia. Therefore, solutions to support writing and learning to write, or to compensate for the inadequate competence, are much needed. Meanwhile, digital writing tools are increasingly applied for writing (Schröder et al., 2022). This thesis research thus argues that speech-to-text technology (STT), as one such digital tool, is a potential tool to provide support and/or assistance for students with dyslexia in their acquisition of writing skills and in their writing.

STT is based on automatic speech recognition technology which has, recently, been greatly improved. The application of STT that assists lower-order writing processes has been argued in this thesis, on the theoretical basis of Cognitive Load Theory (Sweller, 1994), to potentially free up cognitive resources for higher-order writing processes, and on the basis of motivational theories to increase writing motivation, for example due to empowerment (Bandura, 1997). However, the implementation of a tool such as STT in the educational setting is challenging (Kerres, 2018), although digital tools have become, in the digital age, part of the so-called 'New Literacies' (Leu et al., 2013).

This thesis study has aimed at filling existing research gaps about the effectiveness of STT for students with dyslexia and about challenges in its implementation. Some, mostly English studies exist on the use of STT for defined groups such as students with learning disability or foreign language learners; and a few Scandinavian studies have examined STT specifically for students with dyslexia. However, since many studies investigate STT in the dictating modality, the present study has attempted to investigate the effect of STT on text production both by hand and by dictation. While one German study has investigated STT for secondary school students, there is no evidence in German about its use for primary school students with dyslexia.

An intervention study, with a mixed-methods design, was thus considered appropriate for this thesis research to investigate whether (STT) can be beneficially applied to reduce barriers for students affected by dyslexia. To meet the conditions for the successful use of STT as emphasised in the research literature, the present study provided specific and evidence-based writing instruction by means of video to one experimental group (EG 1) which used STT. To control for both the use of STT

and the provided writing instruction, two groups were installed to control for these effects, namely a control group (CG) and a second experimental group (EG 2). *N* 107 Grade 5 students with dyslexia, equally distributed into the three study groups, completed the 18-week intervention study, conducted in the writing classroom.

However, the application of STT in school is considered challenging in the literature. Therefore, the intervention study, representing the quantitative core component of this thesis research, was supplemented by a qualitative component. To investigate the effect of STT, the variables, i. e. text quality, text length, writing fluency, reading fluency and writing motivation, were selected. To assess these, assessment tools were either selected or translated and adapted. To explore challenges in the implementation, 11 interviews with an interview guide were conducted with professionals of the EG 1, and the students of EG 1 answered three questions about their perceptions on the use of STT in school.

The main models, run in SPSS, were repeated measure ANOVAs, calculating main effects and their interactions. In addition, several characteristics of students such as comorbidities or severity were analysed by second-order interaction, to evaluate whether such characteristics reinforced or suppressed an effect. Interview data and answers from the students were analysed in MAXQDA by qualitative content analysis based on Mayring (2022).

All results analysing the effect of STT on conventional writing and writing-related skills show no statistically significant interaction between the treatment condition and over time. This means that, to date, STT is not a more effective method to acquire writing skills than conventional methods. However, students using STT do not stagnate in their development either. The same applies to writing motivation, which results in no significant interaction. Regarding the dictation modality, only analysed in the subsample $n = 36$ of EG 1, the students produced significantly longer texts and more accurate text in the dictating modality than by handwriting. In addition, these students dictated significantly longer texts at post-test than they did at pre-test. Therefore, there is a significant effect, close to moderate, of STT in the mode-specific dictation task, in contrast to the non-significant results for writing fluency in the handwriting task. To summarise, it can be stated that STT is, to date, not effective as a learning tool; however, STT does assist Grade 5 students in writing by allowing the production of longer and more accurate texts. Therefore, it does show eligibility as an assistive tool for these students.

The qualitative data support understanding of the challenges of the implementation but may also cast light, to some extent, on the quantitative results. The most frequent statements by teachers concerned oral language skills among the students and their working behaviour and working processes. Also mentioned by all interviewees was the low accuracy of the STT output, which was emphasised as a major problem in terms of the successful use of the tool. The responses from the students show a similar pattern: although many students reported a somewhat positive or a positive attitude towards the tool, frustration due to low accuracy was one reported issue.

The triangulation of the quantitative and qualitative results within the discussion has explained, to some extent, the non-significant results from the group comparison, and reveal limitations of the study. The reported issues are assumed to be partly in the students with oral language problems and partly in the accuracy of the software. The technology is, however, expected to further improve in its performance. Therefore, STT still has the potential to be an effective learning medium for writing in addition to already being an effective tool for text production in the dictation modality.

References

Abbott, C. (2007). *E-inclusion: Learning Difficulties and Digital Technologies*. Report 15 (LR15). King's College.

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, *102*(2), 281–298. https://doi.org/10.1037/a0019318

Adlof, S. M., Baron, L. S., Bell, B. A., & Scoggins, J. (2021). Spoken Word Learning in Children With Developmental Language Disorder or Dyslexia. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*(7), 2734–2749. https://doi.org/10.1044/2021_JSLHR-20-00217

Afonso, O., Connelly, V., & Barnett, A. L. (2019). Struggling with Writing: An Examination of Writing Difficulties in Specific Language Impairment, Developmental Dyslexia and Developmental Coordination Disorder. In C. Perret, & T. Olive (Eds.), *Studies in writing: Volume 39. Spelling and writing words: Theoretical and methodological advances* (pp. 112–127). Brill. https://doi.org/10.1163/9789004394988_007

Aldosari, B., Babsai, R., Alanazi, A., Aldosari, H., & Alanazi, A. (2023). The Progress of Speech Recognition in Health Care: Surgery as an Example. In J. Mantas, P. Gallos, E. Zoulias, A. Hasman, M. S. Househ, M. Charalampidou, & A. Magdalinou (Eds.), *Healthcare Transformation with Informatics and Artificial Intelligence* (pp. 414–418). https://doi.org/10.3233/SHTI230519

Almgren Bäck, G., Lindeblad, E., Elmqvist, C., & Svensson, I. (2024). Dyslexic students' experiences in using assistive technology to support written language skills: A five-year follow-up. *Disability and Rehabilitation. Assistive Technology*, *19*(4), 1217–1227 https://doi.org/10.1080/17483107.2022.2161647

Almgren Bäck, G., Mossige, M., Bundgaard Svendsen, H., Rønneberg, V., Selenius, H., Berg Gøttsche, N., Dolmer, G., Fälth, L., Nilsson, S., & Svensson, I. (2024). Speech-to-text intervention to support text production among students with writing difficulties: A single-case study in Nordic countries. *Disability and Rehabilitation. Assistive Technology*, *19*(8), 3110–3129. https://doi.org/10.1080/17483107.2024.2351488

Alves, R. A [Rui Alexandre], Castro, S. L., & Olive, T. (2008). Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill. *International Journal of Psychology: Journal International De Psychologie*, *43*(6), 969–979. https://doi.org/10.1080/00207590701398951

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, *40*(1), 57–87. https://doi.org/10.1111/1469-7610.00424

Applebee, A. N., & Langer, J. A. (2009). EJ Extra: What Is Happening in the Teaching of Writing? *The English Journal*, *98*(5), 18–28. https://www.jstor.org/stable/40503291

Arcon, N., Klein, P. D., & Dombroski, J. D. (2017). Effects of Dictation, Speech to Text, and Handwriting on the Written Composition of Elementary School English Language Learners. *Reading & Writing Quarterly*, *33*(6), 533–548. https://doi.org/10.1080/10573569.2016.1253513

Ayer, G., & Joss Almassri, C. (April 2023). *Lese- und/oder Rechtschreibstörung (Dyslexie) in der Regelschule: Information zuhanden der Lehrpersonen zu Dyslexie, Massnahmen der Differenzierung im Unterricht und zum Nachteilsausgleich.* SZH. https://edudoc.ch/record/230201?v=%5B%22%5B%27pdf%27%5D%22%5D

Bachmann, T., & Becker-Mrotzek, M. (2017a). Schreibaufgaben situieren und profilieren. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 201–220). Waxmann.

Bachmann, T., & Becker-Mrotzek, M. (2017b). Schreibkompetenz und Textproduktion modellieren. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 25–54). Waxmann.

Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition*, *11*(2), 185–199. https://doi.org/10.1016/0010-0277(82)90025-7

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 91–106. https://doi.org/10.1016/j.cortex.2018.05.015 (Special issue: Review).

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. W. H. Freeman/Times Books/Henry Holt & Co.

Bausch, M. E., Jones Ault, M., & Hasselbring, T. S. (2015). Assistive Technology in Schools: Lessons Learned from the National Assistive Technology Research Institute. In D. L. Edyburn (Ed.), *Advances in Special Education Technology. Efficacy of Assistive Technology Interventions* (pp. 13–50). Emerald Publishing. https://doi.org/10.1108/S2056-769320150000001002

Bazerman, C., Graham, S., Applebee, A. N., Matsuda, P. K., Berninger, V. W., Murphy, S., Brandt, D., Wells Rowe, D., & Schleppegrell, M. (2017). Taking the Long View on Writing Development. *Research in the Teaching of English*, *51*(3), 351–360. https://www.jstor.org/stable/44821267

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Becker-Mrotzek, M. (2014). Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur. In J. Grabowski (Ed.), *Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur* (pp. 51–71). Verlag Barbara Budrich. https://doi.org/10.2307/j.ctvddzg18.5

Beer, J. de, Engels, J., Heerkens, Y., & van der Klink, J. (2014). Factors influencing work participation of adults with developmental dyslexia: A systematic review. *BioMed Central*, *77*(14), 1–22. http://www.biomedcentral.com/1471-2458/14/77

Beers, S. F., Mickail, T., Abbott, R., & Berninger, V. (2017). Effects of transcription ability and transcription mode on translation: Evidence from written compositions, language bursts and pauses when students in grades 4 to 9, with and without persisting dyslexia or dysgraphia, compose by pen or by keyboard. *Journal of Writing Research*, *9*(1), 1–25. jowr.org/index.php/jowr/article/view/632/569

Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition. The psychology of education and instruction*. L. Erlbaum Associates; Routledge. https://doi.org/10.4324/9780203812310

Berner, K., & Alves, A. N. (2021). A scoping review of literature using speech recognition technologies by individuals with disabilities in multiple contexts. *Disability and Reha-*

bilitation: Assistive Technology, *18*(7), 1139–1145. https://doi.org/10.1080/17483107.2021.1986583

Berner, K., & Alves, A. N. (2023). A scoping review of literature using speech recognition technologies by individuals with disabilities in multiple contexts. *Disability and Rehabilitation. Assistive Technology*, *18*(7), 1139–1145. https://doi.org/10.1080/17483107.2021.1986583

Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 345–363). Guilford Press.

Berninger, V. W., Winn, W. D., Stock, P., Abbott, R. D., Eschen, K., Lin, S.-J., Garcia, N., Anderson-Youngstrom, M., Murphy, H., Lovitt, D., Trivedi, P., Jones, J., Amtmann, D., & Nagy, W. (2008). Tier 3 specialized writing instruction for students with dyslexia. *Reading and Writing*, *21*(1–2), 95–129. https://doi.org/10.1007/s11145-007-9066-x

Berthele, R. (2019). Alemannisch und der Deutschunterricht. *Linguistik Online*, *98*(5), 387–409. https://doi.org/10.13092/lo.98.5945

Bhide, A., Shah, P. S., & Acharya, G. (2018). A simplified guide to randomized controlled trials. *Acta Obstetricia Et Gynecologica Scandinavica*, *97*(4), 380–387. https://doi.org/10.1111/aogs.13309

Blackhurst, E. A. (2005). Perspectives on Applications of Technology in the Field of Learning Disabilities. *Learning Disability Quarterly*, *28*(2), 175–178. https://doi.org/10.2307/1593622

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th edition). Springer. https://doi.org/10.1007/978-3-642-12770-0

Boscolo, P., & Gelati, C. (2013). Best Practices in Promoting Motivation for Writing. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (2nd edition, pp. 284–308). The Guilford Press.

Bosse, I., Schluchter, J.-R., & Zorn, I. (2018). Einleitung: Ziel des Handbuchs. In I. Bosse, J.-R. Schluchter, & I. Zorn (Eds.), *Handbuch Inklusion und Medienbildung* (pp. 9–13). Beltz.

Bosse, M.-L., Chaves, N., & Valdois, S. (2014). Lexical orthography acquisition: Is handwriting better than spelling aloud? *Frontiers in Psychology*, *5*, 1–9. https://doi.org/10.3389/fpsyg.2014.00056

Bouck, E. C., Flanagan, S., Joshi, G. S., Sheikh, W., & Schleppenbach, D. (2011). Speaking Math – A Voice Input, Speech Output Calculator for Students with Visual Impairments. *Journal of Special Education Technology*, *26*(4), 1–14. https://doi.org/10.1177/016264341102600401

Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100. https://doi.org/10.1177/0265532214542994

Bouwer, R., Lesterhuis, M., Smeth, F. de, van Keer, H., & Maeyer, S. (2024). Comparative approaches to the assessment of writing: Reliability and validity of benchmark rating and comparative judgement. *Journal of Writing Research*, *15*(3), 498–518. https://doi.org/10.17239/jowr-2024.15.03.03

Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE Life Sciences Education*, *15*(4), 1–6. https://doi.org/10.1187/cbe.16-03-0125

Brouwer, K. L. (2012). Writing motivation of students with language impairments. *Child Language Teaching and Therapy*, *28*(2), 189–210. https://doi.org/10.1177/0265659012436850

Brüggemann, M. (2018). Berufsfeld Grundschule. In I. Bosse, J.-R. Schluchter, & I. Zorn (Eds.), *Handbuch Inklusion und Medienbildung* (pp. 111–118). Beltz.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S [Sharon] (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology*, *105*(1), 25–38. https://doi.org/10.1037/a0029692

Bruning, R., & Horn, C. (2000). Developing Motivation to Write. *Educational Psychologist*, *35*(1), 25–37. https://doi.org/10.1207/S15326985EP3501_4

Buchner, A., Erdfelder, E., Faul, F., & Lang, A.-G. (2014). *G*Power 3.1.9.2* (Version 3.1) [Computer software].

Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4th edition.). *Psychologie*. Pearson.

Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen 1 (2002 & rev. Stand am 1. Juli 2020).

Cabbage, K. L., Farquharson, K., Iuzzini-Seigel, J., Zuk, J., & Hogan, T. P. (2018). Exploring the Overlap Between Dyslexia and Speech Sound Production Deficits. *Language, Speech, and Hearing Services in Schools*, *49*(4), 774–786. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0008

Camacho, A., Alves, R. A [Rui A.], & Boscolo, P. (2021). Writing Motivation in School: A Systematic Review of Empirical Research in the Early Twenty-First Century. *Educational Psychology Review*, *33*(1), 213–247. https://doi.org/10.1007/s10648-020-09530-4

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/s0010-0277(03)00164-1

Catts, H. W., Patton Terry, N., Lonigan, C. J., Compton, D. L., Wagner, R. K., Steacy, L. M., Farquharson, K., & Petscher, Y. (2024). Revisiting the definition of dyslexia. *Annals of Dyslexia*, *74*, 282–302. https://doi.org/10.1007/s11881-023-00295-3

Centre de compétences Fritic. (2024, June 15). https://www.fritic.ch/de#gsc.tab=0

Chen, F., & Jokinen, K. (2010). Preface. In F. Chen, & K. Jokinen (Eds.), *Speech Technology: Theory and applications* (pp. v–ix). Springer. https://doi.org/10.1007/978-0-387-73819-2

Chusseau, V. (2018). *Enfants et adolescents à haut potentiel intellectuel et troubles spécifiques du langage écrit : Proposition d'une démarche pour le diagnostic orthophonique de dyslexie/dysorthographie* [Master thesis]. Université de Nantes, Nantes.

Cislaru, G., Feltgen, Q., Khoury, E., Delorme, R., & Bucci, M. P. (2024). Language Processing Units Are Not Equivalent to Sentences: Evidence from Writing Tasks in Typical and Dyslexic Children. *Languages*, *9*(5), 1–28. https://doi.org/10.3390/languages9050155

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 95–121). McGraw-Hill.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition). Erlbaum Associates.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159. https://psycnet.apa.org/doi/10.1037/0033-2909.112.1.155

Conradi, K., Jang, B. G., & McKenna, M. C. (2014). Motivation Terminology in Reading Research: A Conceptual Review. *Educational Psychology Review*, *26*(1), 127–164. https://doi.org/10.1007/s10648-013-9245-z

Creagh, S., Thompson, G., Mockler, N., Stacey, M., & Hogan, A. (2023). Workload, work intensification and time poverty for teachers and school leaders: A systematic research synthesis. *Educational Review*, *77*(2), 661–680. https://doi.org/10.1080/00131911.2023.2196607

DATAtab Team. (2024). *Zweifaktorielle Varianzanalyse mit Messwiederholung*. https://datatab.de

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Recognition of Spoken Digits. *The Journal of the Acoustical Society of America*, *24*(6), 637–642. https://doi.org/10.1121/1.1906946

Dawson, K., Antonenko, P., Lane, H., & Zhu, J. (2019). Assistive Technologies to Support Students with Dyslexia. *TEACHING Exceptional Children*, *51*(3), 226–239. https://doi.org/10.1177/0040059918794027

D-EDK. (2024, July 2). *Lehrplan 21*. https://www.lehrplan21.ch/

Deutsche Gesellschaft für Kinder und Jugendpsychiatrie, Psychosomatik und Psychotherapie e. V. (DGKJP). (2015, April 23). *Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese- und/oder Rechtschreibstörung* (Evidenz- und konsensbasierte Leitlinie (S3) AWMF-Registernummer 028–044). https://register.awmf.org/assets/guidelines/028-044l_S3_Lese-Rechtschreibst%C3%B6rungen_Kinder_Jugendliche_2015-06-abgelaufen.pdf

Dhanjal, A. S., & Singh, W. (2024). A comprehensive survey on automatic speech recognition using neural networks. *Multimedia Tools and Applications*, *83*(8), 23367–23412. https://doi.org/10.1007/s11042-023-16438-y

Di Betta, A. M., & Romani, C. (2006). Lexical learning and dysgraphia in a group of adults with developmental dyslexia. *Cognitive Neuropsychology*, *23*(3), 376–400. https://doi.org/10.1080/02643290442000545

Di Folco, C., Guez, A., Peyre, H., & Ramus, F. (2022). Epidemiology of reading disability: A comparison of DSM-5 and ICD-11 criteria. *Scientific Studies of Reading*, *26*(4), 337–355. https://doi.org/10.1080/10888438.2021.1998067

Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: the role of curriculum based measures. *British Educational Research Journal*, *41*(4), 575–595. https://doi.org/10.1002/berj.3162

Dockrell, J. E., Marshall, C. R., & Wyse, D. (2016). Teachers' reported practices for teaching writing in England. *Reading and Writing*, *29*, 409–434. https://doi.org/10.1007/s11145-015-9605-9

Duff, D. M., Hendricks, A. E., Fitton, L., & Adlof, S. M. (2023). Reading and Math Achievement in Children With Dyslexia, Developmental Language Disorder, or Typical Development: Achievement Gaps Persist From Second Through Fourth Grades. *Journal of Learning Disabilities*, *56*(5), 371–391. https://doi.org/10.1177/00222194221105515

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

EDA. (2024, April 24). *About Switzerland: Political system*. https://www.eda.admin.ch/aboutswitzerland/en/home/politik-geschichte/politisches-system/foederalismus.html

Edyburn, D. L. (Ed.). (2015). *Efficacy of Assistive Technology Interventions*. Emerald Publishing. https://doi.org/10.1108/S2056-769320150000001011

Edyburn, D. L. (2010). Would You Recognize Universal Design for Learning if You Saw it? Ten Propositions for New Directions for the Second Decade of UDL. *Learning Disability Quarterly*, *33*(1),33-41. https://doi.org/10.1177/073194871003300103

Edyburn, D. L. (2015). Expanding the Use of Assistive Technology while Mindful of the Need to Understand Efficacy. In D. L. Edyburn (Ed.), *Advances in Special Education Technology. Efficacy of Assistive Technology Interventions* (pp. 9–20). Emerald Publishing. https://doi.org/10.1108/S2056-769320150000001001

Edyburn, D. L. (2021). Universal Usability and Universal Design for Learning. *Intervention in School and Clinic*, *56*(5), 310–315. https://doi.org/10.1177/1053451220963082

Ekholm, E., Zumbrunn, S [S.], & DeBusk-Lane, M. (2018). Clarifying an Elusive Construct: A Systematic Review of Writing Attitudes. *Educational Psychology Review*, *30*(3), 827–856. https://doi.org/10.1007/s10648-017-9423-5

Farquharson, K., Hogan, T. P., & Fox, A. B. (2021). Factors that influence non-word repetition performance in children with and without persistent speech sound disorders. *International Journal of Language & Communication Disorders*, *56*(6), 1218–1234. https://doi.org/10.1111/1460-6984.12663

Fayol, M. (1999). From on-line management problems to strategies in written production. In M. Torrance, & G. Jeffery (Eds.), *Cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 13–24). Amsterdam University Press.

Feenstra, H. (2014). *Assessing writing ability in primary education: On the evaluation of text quality and text complexity* [Doctoral thesis, University of Groningen, Groningen]. CrossRef.

Feilke, H. (2017). Schreibdidaktische Konzepte. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 153–171). Waxmann.

Feng, S., Halpern, B. M., Kudina, O., & Scharenborg, O. (2024). Towards inclusive automatic speech recognition. *Computer Speech & Language*, *84*, 1–15. https://doi.org/10.1016/j.csl.2023.101567

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th edition). Sage.

Filippidou, F., & Moussiades, L. (2020). A Benchmarking of IBM, Google and Wit Automatic Speech Recognition Systems. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *IFIP Advances in Information and Communication Technology. Artificial Intelligence Applications and Innovations* (Vol. 583, pp. 73–82). Springer International Publishing. https://doi.org/10.1007/978-3-030-49161-1_7

Fink, A. (2013). *How to Conduct Surveys: A Step-by-Step Guide* (3rd edition). Sage.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: an introduction to theory and research*. Reading, Mass: Addison Wesley Publication Co.

Flütsch Keravec, S. (forthcoming). Motivation to write and dyslexia: Differences and similarities between groups and subgroups of students with and without dyslexia.

Frith, U. (1985). Beneath the Surface of Developmental Dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Psychology Library Editions: Vol. 8. Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading* (pp. 301–330). Taylor & Francis Group. https://doi.org/10.4324/9781315108346-18

Furui, S. (2010). History and Development of Speech Recognition. In F. Chen, & K. Jokinen (Eds.), *Speech Technology: Theory and applications* (pp. 1–18). Springer. https://doi.org/10.1007/978-0-387-73819-2_1

Gabrial, B. (2008). History of Writing Technologies. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 27–39). L. Erlbaum Associates.

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PloS One*, *9*(2), 1–12. https://doi.org/10.1371/journal.pone.0089900

Galuschka, K., & Schulte-Körne, G. (2016). The Diagnosis and Treatment of Reading and/or Spelling Disorders in Children and Adolescents. *Deutsches Ärzteblatt International*, *113*(16), 279–286. https://doi.org/10.3238/arztebl.2016.0279

Gebhardt, M. (2023). *Pädagogische Diagnostik: Leistung, Kompetenz und Entwicklung messen, bewerten und für individuelle Förderung interpretieren*. Version 0.1. https://doi.org/10.5283/EPUB.54073

Ghayas, A., Ahmad Lawaye, A., Ahmad Mir, T., & Rana, P. (2024). Toward Developing Attention-Based End-To-End Automatic Speech Recognition. In A. E. Hassanien, O. Castillo, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2023, Volume 2* (pp. 147–161). Springer.

Giannouli, V., & Banou, M. (2020). The intelligibility and comprehension of synthetic versus natural speech in dyslexic students. *Disability and Rehabilitation. Assistive Technology*, *15*(8), 898–907. https://doi.org/10.1080/17483107.2019.1629111

Gillespie, A., & Graham, S. (2014). A Meta-Analysis of Writing Interventions for Students With Learning Disabilities. *Exceptional Children*, *80*(4), 454–473. https://doi.org/10.1177/0014402914527238

Girden, E. R. (1992). *ANOVA: Repeated measures. Quantitative applications in the social sciences*. Sage. https://doi.org/10.4135/9781412983419

Girgensohn, K., & Sennewald, N. (2012). *Schreiben lehren, Schreiben lernen: Eine Einführung*. WGB (Wissenschaftliche Buchgesellschaft).

Glaser, C., & Palm, D. (2014). *Aufsatztraining für 4. bis 6. Klassen: Ein Lehrermanual mit Unterrichtsmaterialien zur Förderung von Schreibkompetenz und Arbeitsverhalten*. Hogrefe.

Glogowska, M. (2011). Paradigms, pragmatism and possibilities: Mixed-methods research in speech and language therapy. *International Journal of Language & Communication Disorders*, *46*(3), 251–260. https://doi.org/10.3109/13682822.2010.507614

Grabowski, J. (2014). Kompetenz: ein bildungswissenschaftlicher Begriff. In J. Grabowski (Ed.), *Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur* (pp. 9–28). Verlag Barbara Budrich. https://doi.org/10.2307/j.ctvddzg18.3

Graham, J. M. (2008). The General Linear Model as Structural Equation Modeling. *Journal of Educational and Behavioral Statistics*, *33*(4), 485–506. https://www.jstor.org/stable/20172130

Graham, S. (2018). A Revised Writer(s)-Within-Community Model of Writing. *Educational Psychologist*, *53*(4), 258–279. https://doi.org/10.1080/00461520.2018.1481406

Graham, S. (2019). Changing How Writing Is Taught. In American Educational Research Association (Ed.), *Review of Research in Education. Changing Teaching Practice in P-20 Educational Settings* (Vol. 43, pp. 277–303). Sage Journals.

Graham, S., Ahumada, S., Bañales, G., Puente, A., Guajardo, M., & Muñoz, I. (2022). Intermediate-Grade Teachers' Adaptations for Weaker Writers: A National Survey in Urban Schools in Chile. *Journal of Learning Disabilities*, *55*(2), 87–98. https://doi.org/10.1177/0022219420941795

Graham, S., Berninger, V., & Abbott, R. (2012). Are Attitudes Toward Writing and Reading Separable Constructs? A Study With Primary Grade Children. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *28*(1), 51–69. https://doi.org/10.1080/10573569.2012.632732

Graham, S., Berninger, V., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology*, *32*(3), 516–536. https://doi.org/10.1016/j.cedpsych.2007.01.002

Graham, S., Capizzi, A., Harris, K. R., Hebert, M., & Morphy, P. (2014). Teaching writing to middle school students: a national survey. *Reading and Writing*, *27*(6), 1015–1042. https://doi.org/10.1007/s11145-013-9495-7

Graham, S., Daley, S. G., Aitken, A. A [Angelique A.], Harris, K. R., & Robinson, K. H. (2018). Do writing motivational beliefs predict middle school students' writing performance? *Journal of Research in Reading*, *41*(4), 642–656. https://doi.org/10.1111/1467-9817.12245

Graham, S., Harbaugh-Schattenkirk, A. G., Aitken, A. A [A. Angelique], Harris, K. R., Ng, C., Wilson, J. M., & Wdowin, J. (2023). Writing Motivation Questionnaire: Factorial and Construct Validity with Middle School Students. *Educational Psychology Review*, *35*(5), 1–28. https://doi.org/10.1007/s10648-023-09742-4

Graham, S., & Harris, K. R. (1989). Improving Learning Disabled Students' Skills at Composing Essays: Self-Instructional Strategy Training. *Exceptional Children*, *56*(3), 201–214. https://doi.org/10.1177/001440298905600305

Graham, S., & Harris, K. R. (2019a). Designing an Effective Writing Program. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (3rd edition, pp. 3–25). The Guilford Press.

Graham, S., & Harris, K. R. (2019b). Evidence-Based Practices in Writing: Chapter 1. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (3rd edition, pp. 3–25). The Guilford Press.

Graham, S., & Harris, K. R. (2020). Writing and Students with Learning Disabilities. In M. J. Andrew, R. A. Sperling, & K. J. Newton (Eds.), *Handbook of Educational Psychology and Students with Special Needs* (pp. 487–509). Routledge. https://doi.org/10.4324/9781315100654

Graham, S., Harris, K. R., Bartlett, B. J., Popadopoulou, E., & Santoro, J. (2016). Acceptability of Adaptations for Struggling Writers. *Learning Disability Quarterly*, *39*(1), 5–16. https://doi.org/10.1177/0731948714554038

Graham, S., Harris, K. R., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment*. A Carnegie Corporation Time to Act report.

Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, *30*(2), 207–241. https://doi.org/10.1016/j.cedpsych.2004.08.001

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-Based Writing Practices and the Common Core: Meta-analysis and Meta-synthesis. *The Elementary School Journal*, *115*(4), 498–522. https://doi.org/10.1086/681964

Graham, S., Kiuhara, S. A., Harris, K. R., & and Fishman, E. J. (2017). The Relationship among Strategic Writing Behavior, Writing Motivation, and Writing Performance with Young, Developing Writers. *The Elementary School Journal*, *118*(1), 82–104. http://www.journals.uchicago.edu/t-and-c

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, *104*(4), 879–896. https://doi.org/10.1037/a0029185

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, *27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.2307/1163620

Hammill, D. D., & Larsen, S. C. (2009). *TOWL 4: Test of Written Language Fourth Edition*. Examiner's Manual (4th edition). Pro-Ed.

Harris, K. R., Graham, S., & Mason, L. H. (2003). Self-Regulated Strategy Development in the Classroom: Part of a Balanced Approach to Writing Instruction for Students With Disabilities. *Focus on Exceptional Children*, *35*(7), 1–16. https://core.ac.uk/download/pdf/235895905.pdf

Harsch, C., Neumann, A., Lehmann, R., & Schröder, K. (2007). Schreibfähigkeit. In B. Beck, & E. Klieme (Eds.), *Sprachliche Kompetenzen: Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (pp. 38–58). Beltz

Haug, K. N., & Klein, P. D. (2018). The Effect of Speech-to-Text Technology on Learning a Writing Strategy, *34*(1), 47–62. https://doi.org/10.1080/10573569.2017.1326014

Hayes, J. R. (2009). From Idea to Text. In R. Beard (Ed.), *The SAGE handbook of writing development* (pp. 65–79). Sage. https://doi.org/10.4135/9780857021069.n5

Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hayes, J. R., & Ann Chenoweth, N. (2007). Working Memory in an Editing Task. *Written Communication*, *24*(4), 283–294. https://doi.org/10.1177/0741088307304826

Hayes, J. R., & Flower, L. S. (1980). Identifying the Organization of Writing Processes. In L. W. Gregg, & R. Steinberg (Eds.), *Cognitive process in writing: An Interdisciplinary Approach* (pp. 3–33). Lawrence Erlbaum.

Hayes, J. R., & Olinghouse, N. G. (2015). Can Cognitive Writing Models Inform the Design of the Common Core State Standards? *The Elementary School Journal*, *115*(4), 480–497. https://doi.org/10.1086/681909

Hebert, M., Kearns, D. M., Hayes, J. B., Bazis, P., & Cooper, S. (2018). Why Children With Dyslexia Struggle With Writing and How to Help Them. *Language, Speech, and Hearing Services in Schools*, *49*(4), 843–863. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0024

Helfferich, C. (2011). *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews* (4th edition). VS Verlag für Sozialwissenschaften/Springer Fachmedien. https://doi.org/10.1007/978-3-531-92076-4

Hemmerich, W. A. (2024a, August 29). *Ungepaarter t-Test: Normalverteilung überprüfen. StatistikGuru.de*. https://statistikguru.de/spss/ungepaarter-t-test/normalverteilung-ueberpruefen-5.html

Hemmerich, W. A. (2024b, August 30). *ANOVA mit Messwiederholung: Sphärizität bestimmen. StatistikGuru.de*. https://statistikguru.de/spss/rm-anova/sphaerizitaet-bestimmen.html

Hennes, A.-K., Michels, V., Becker-Mrotzek, M., Koch, A. M., Möller, J., Woerfel, T., & Fleckenstein, J. (2023). Der SMisK – ein Fragebogen zur Erfassung der Schreibmotivation im schulischen Kontext. *Didaktik Deutsch*, *28*(55), 59–84. https://doi.org/10.21248/dideu.682

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Higgins, E. L., & Raskind, M. H. (1995). Compensatory Effectiveness of Speech Recognition on the Written Composition Performance of Postsecondary Students with Learning Disabilities. *Learning Disability Quarterly*, *18*(2), 159–174. https://www.jstor.orgn/stable/1511202

Hollier, R. (Ed.) (1983). *Applications and potential of speech recognition*. IFS Publications LTD.

Hopfenbeck, T. N. (2019). Writing assessment, comparative judgement and students' evaluative expertise. *Assessment in Education: Principles, Policy & Practice*, *26*(1), 1–5. https://doi.org/10.1080/0969594X.2019.1577959

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A Practical Guide to Curriculum-Based Measurement* (2nd edition). Guilford Publications.

Hurschler Lichtsteiner, S. (2020). Differenzierende Beurteilung der Handschrift – ein Bestandteil der Schreibförderung. *Leseforum.Ch*(3), 1–26. https://www.leseforum.ch/sysModules/obxLeseforum/Artikel/711/2020_3_de_hurschler.pdf

IBM. *SPSS Software | IBM* (Version 28) [Computer software]. https://www.ibm.com/spss

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464. https://doi.org/10.30935/cedtech/13605

Institut St. Joseph du Guintzet. (2021, July 27). *Sprachheilschule*. https://www.guintzet.ch/de/sprachheilschule/

Interkantonale Lehrmittelzentrale. (2024, July 2). *Falesia Lehrmittelspektrum*. https://ilz.ch/lehrmittel/lehrmitteleinsatz

Ise, E., & Schulte-Körne, G. (2010). Spelling deficits in dyslexia: Evaluation of an orthographic spelling training. *Annals of Dyslexia*, *60*(1), 18–39. https://doi.org/10.1007/s11881-010-0035-8

Jaccard, J., & Turrisi, R. (2003). *Interaction Effects in Multiple Regression* (2nd edition). Sage.

Joyner, R. E., & Wagner, R. K. (2020). Co-occurrence of Reading Disabilities and Math Disabilities: A Meta-Analysis. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *24*(1), 14–22. https://doi.org/10.1080/10888438.2019.1593420

Kanton Zürich, Bildungsdirektion. (2022, February 10). *Lese-Rechtschreib-Störung LRS* (Empfehlung Umgang LRS).

Karpagavalli, S., Deepika, R., Kokila, P., & Usha Rani, K. (2011). Automatic Speech Recognition: Architecture, Methodologies and Challenges – A Review. *International Journal of Advanced Research in Computer Science*, *2*(6), 326–331.

Keelor, J. L., Creaghead, N. A., Silbert, N. H., Breit, A. D., & Horowitz-Kraus, T. (2023). Impact of text-to-speech features on the reading comprehension of children with reading and language difficulties. *Annals of Dyslexia*, *73*(3), 469–486. https://doi.org/10.1007/s11881-023-00281-9

Kellogg, R. T. (2001). Competition for Working Memory among Writing Processes. *The American Journal of Psychology*, *114*(2), 175–191. https://www.jstor.org/stable/1423513

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working Memory in Written Composition: An Evaluation of the 1996 Model. *Journal of Writing Research*, *5*(2), 159–190. https://doi.org/10.17239/jowr-2013.05.02.1

Kerres, M. (2018). *Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote* (5th edition). *De gruyter studium*. de Gruyter. https://doi.org/10.1515/9783110456837

Kerres, M., Buntins, K., Buchner, J., Drachsler, H., & Zawacki-Richter, O. (in press). Interaktive, adaptive und künstlich-intelligente Lernprogramme: Eine kritische Analyse aus Sicht der Mediendidaktik. In C. de Witt, C. Gloerfeld, & S. E. Wrede (Eds.), *Künstliche Intelligenz in der Bildung* (pp. 1–19). Springer VS.

Kim, S.-E., Chernyak, B. R., Seleznova, O., Keshet, J., Goldrick, M., & Bradlow, A. R. (2024). Automatic recognition of second language speech-in-noise. *JASA Express Letters*, *4*(2), 1–7. https://doi.org/10.1121/10.0024877

Kim, Y.-S. G., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2018). Theorization and an Empirical Investigation of the Component-Based and Developmental Text Writing Fluency Construct. *Journal of Learning Disabilities*, *51*(4), 320–335. https://doi.org/10.1177/0022219417712016

Kim, Y.-S. G., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, *109*(1), 35–50. https://doi.org/10.1037/edu0000129

Kirby, P., & Snowling, M. J. (2022). *Dyslexia: A history*. McGill-Queen's University Press. https://doi.org/10.1515/9780228015390

Kirk, C., & Gillon, G. T. (2009). Integrated morphological awareness intervention as a tool for improving literacy. *Language, Speech, and Hearing Services in Schools*, *40*(3), 341–351. https://doi.org/10.1044/0161-1461(2008/08-0009)

Kirschner, P. A., & van Merriënboer, J. J. (2013). Do Learners Really Know Best? Urban Legends in Education. *Educational Psychologist*, *48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Klassen, R. (2002). Writing in Early Adolescence: A Review of the Role of Self-Efficacy Beliefs. *Educational Psychology Review*, *14*(2), 173–203. https://doi.org/10.1023/A:1014626805572

Klicpera, C., Schabmann, A., Gasteiger-Klicpera, B., & Schmidt, B. M. (2020). *Legasthenie – LRS* (6th edition). Ernst Reinhardt. https://doi.org/10.36198/9783838554822

Knapp, T. R. (1978). Canonical Correlation Analysis: A General Parametric Significance-Testing System. *Psychological Bulletin*, *85*(2), 410–416. https://doi.org/10.1037/0033-2909.85.2.410

Knezek, G., & Christensen, R. (2016). Extending the will, skill, tool model of technology integration: adding pedagogy as a new model construct. *Journal of Computing in Higher Education*, *28*(3), 307–325. https://doi.org/10.1007/s12528-016-9120-2

Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In O. Deutschmann, H. Flasche, B. König, M. Kruse, W. Pabst, & W.-D. Stempel (Eds.), *Romanistisches Jahrbuch* (Vol. 36, pp. 15–43). Walter de Gruyter.

Koning, B. B. de, Tabbers, H. K., Rikers, R. M., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior*, *27*(1), 41–45. https://doi.org/10.1016/j.chb.2010.05.010

Kraft, S., Thurfjell, F., Rack, J., & Wengelin, Å. (2019). Lexikala analyser av muntlig, tangentbordsskriven och dikterad text producerad av barn med stavningssvårigheter. *Nordic Journal of Literacy Research (NJLR)*, *5*(3), 102–122. https://doi.org/10.23865/njlr.v5.1511

Kuhn, K., Kersken, V., Reuter, B., Egger, N., & Zimmermann, G. (2023). Measuring the Accuracy of Automatic Speech Recognition Solutions. *ACM Transactions on Accessible Computing*, *16*(4), 1–23. https://doi.org/10.1145/3636513

Lai, E. R., Wolfe, E. W., & Vickers, D. (2015). Differentiation of Illusory and True Halo in Writing Scores. *Educational and Psychological Measurement*, *75*(1), 102–125. https://doi.org/10.1177/00113164414530990

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, *100*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data, *33*(1), 159–174. https://www.jstor.org/stable/2529310

Lee, X. C. (2011). *The Application of Speech Recognition Technology for Remediating the Writing Difficulties of Students with Learning Disabilities* [Dissertation]. University of Washington, Washington.

Lenhard, A., & Lenhard, W. (2017). Diagnoseverfahren zur Erfassung schriftsprachlicher Leistungen. In M. Philipp (Ed.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (pp. 174–198).

Leu, D. J. (1997). Exploring Literacy on the Internet: Caity's Question: Literacy as Deixis on the Internet. *The Reading Teacher*, *51*(1), 62–67. https://www.jstor.org/stable/20201860

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New Literacies: A Dual-Level Theory of the Changing Nature of Literacy, Instruction, and Assessment. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical Models and Processes of Reading* (6th edition, vol. 197, pp. 1150–1181). https://doi.org/10.1598/0710.42

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. *ACL-MIT Press series in natural-language processing*. MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001?locatt=mode:legacy

Linnemann, M., Stephany, S., Lemke, V., Bulut, N., Haider, H., Roth, H.-J., & Becher-Mrotzek, M. (2022). The Dimensionality of Writing and Reading Fluency and its Impact on Comprehension and Composition. *Journal of Writing Research*, *14*(2), 185–227. https://doi.org/10.17239/jowr-2022.14.02.02

Liu, K. K., Thurlow, M. L., Press, A., & and Dose, M. J. (2019). *A Review of the Literature on Computerized Speech-to-Text Accommodations: NCEO Report 414.* University of Minnesota.

Locke, R., Scallan, S., Mann, R., & Alexander, G. (2015). Clinicians with dyslexia: A systematic review of effects and strategies. *The Clinical Teacher*, *12*(6), 394–398. https://doi.org/10.1111/tct.12331

Mabuan, R. A. (2024). ChatGPT and ELT: Exploring Teachers' Voices. *International Journal of Technology in Education*, *7*(1), 128–153. https://doi.org/10.46328/ijte.523

MacArthur, C. A. (2009). Reflections on Research on Writing and Technology for Struggling Writers. *Learning Disabilities Research & Practice, 24*(2), 93–103. https://doi.org/10.1111/j.1540-5826.2009.00283.x

MacArthur, C. A. (2019). Evaluation and Revision. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (3rd edition, pp. 278–297). The Guilford Press.

MacArthur, C. A., & Cavalier, A. R. (2004). Dictation and Speech Recognition Technology as Test Accommodations. *Exceptional Children*, *71*(1), 43–58. https://doi.org/10.1177/001440290407100103

Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools*, *40*(4), 379–390. https://doi.org/10.1002/pits.10096

Malik, M., Malik, M. K., Mehmood, K., & Makhdoom, I. (2021). Automatic speech recognition: a survey. *Multimedia Tools and Applications*, *80*(6), 9411–9457. https://doi.org/10.1007/s11042-020-10073-7

Matheretter. (2024, June 21). *Zufallszahlengenerator für beliebige Zahlen – Matheretter*. https://www.matheretter.de/rechner/zufallszahlengenerator

Matre, M. E. (2022). Speech-to-Text Technology as an Inclusive Approach: Lower Secondary Teachers' Experiences. *Nordisk Tidsskrift for Pedagogikk Og Kritikk (NTPK)*, *8*, 233–247. https://doi.org/10.23865/ntpk.v8.3436

Matre, M. E., & Cameron, D. L. (2024). A scoping review on the use of speech-to-text technology for adolescents with learning difficulties in secondary education. *Disability and Rehabilitation. Assistive Technology*, *19*(3), 1103–1116. https://doi.org/10.1080/17483107.2022.2149865

MAXQDA. *MAXQDA | Qualitative und Mixed Methods-Forschung* (Version 24.2.0) [Computer software]. VERBI Software. https://www.maxqda.com/de/produkte/maxqda

Mayer, A., & Lindberg, S. (2016). *Lese-Rechtschreibstörungen (LRS)*. Ernst Reinhardt/utb.

Mayer, R. E., & Fiorella, L. (2014). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 279–315). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015

Mayring, P. (2022). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (13th edition). Beltz.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "Specifics" of Specific Reading Disability and Specific Language Impairment. *Journal of Child Psychology and Psychiatry*, *41*(7), 869–874. https://doi.org/10.1111/1469-7610.00674

McArthur, G. M., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H.-C., Kohnen, S., Banales, E., Anandakumar, T., Marinus, E., & Castles, A. (2018). Phonics training for English-speaking poor readers. *The Cochrane Database of Systematic Reviews*, *Issue 11*, CD009115. https://doi.org/10.1002/14651858.CD009115.pub3

McCollum, D., Nation, S., & Gunn, S. (2014). Effects of a Speech-to-Text Software Application on Written Expression for Students with Various Disabilities. *National Forum of Special Education Journal*, *25*(1), 1–13.

McCrimmon, A. W., & Climie, E. A. (2011). Test Review: Test of Written Language – Fourth Edition. *Journal of Psychoeducational Assessment*, *29*(6), 592–596. https://doi.org/10.1177/0734282911406646

McMaster, K. L., & Campbell, H. (2008). New and Existing Curriculum-Based Writing Measures: Technical Features Within and Across Grades. *School Psychology Review*, *37*(4), 550–566. https://doi.org/10.1080/02796015.2008.12087867

Methodenberatung, U. Z. (2024, August 30). *Einfaktorielle Varianzanalyse (mit Messwiederholung)*. https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/evarianzmessw.html#3._Die_einfaktorielle_Varianzanalyse_mit_Messwiederholung_mit_SPSS

Mey, G., & Mruck, K. (2011). Grounded-Theory-Methodologie: Entwicklung, Stand, Perspektiven. In G. Mey, & K. Mruck (Eds.), *Grounded Theory Reader* (2nd edition, pp. 11–50). VS Verlag für Sozialwissenschaften/Springer Fachmedien. https://doi.org/10.1007/978-3-531-93318-4_1

Microsoft. (2024, July 2). *PowerPoint*. https://www.microsoft.com/en-au/microsoft-365/powerpoint

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PloS One*, *9*(7), e103537. https://doi.org/10.1371/journal.pone.0103537

Moll, K., & Landerl, K. (2009). Double Dissociation Between Reading and Spelling Deficits. *Scientific Studies of Reading*, *13*(5), 359–382. https://doi.org/10.1080/10888430903162878

Moll, K., Landerl, K., Snowling, M. J., & Schulte-Körne, G. (2019). Understanding comorbidity of learning disorders: Task-dependent estimates of prevalence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *60*(3), 286–294. https://doi.org/10.1111/jcpp.12965

Moran-Ellis, J., Alexander, V. D., Cronin, A., Dickinson, M., Fielding, J., Sleney, J., & Thomas, H. (2006). Triangulation and integration: processes, claims and implications. *Qualitative Research*, *6*(1), 45–59. https://doi.org/10.1177/1468794106058870

Moreno, R., & Park, B. (2010). Cognitive Load Theory: Historical Development and Relation to Other Theories. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 9–28). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.003

Morgan, D. L. (2014). *Integrating Qualitative and Quantitative Methods: A Pragmatic Approach*. Sage. https://doi.org/10.4135/9781544304533

Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: a meta-analysis of research findings. *Reading and Writing*, *25*(3), 641–678. https://doi.org/10.1007/s11145-010-9292-5

Morse, J. M., & Niehaus, L. (2009). *Mixed Method Design: Principles and Procedures*. LEF COAST PRESS.

Mossige, M., Almgren Bäck, G., Svensson, I., Berg Gøttsche, N., Rønneberg, V., Dolmer, G., Selenius, H., Fälth, L., & Bundgaard Svendsen, H. (2023). Study Protocol: Text Performers – Using Speech-to-Text Technology to Support Students With Dyslexia During Text Production. *Nordic Journal of Literacy Research*, *9*(2), 99–123. https://doi.org/10.23865/njlr.v9.3768

Mossige, M., Arendal, E., Kongskov, L., & Svendsen, H. B. (2023). How do technologies meet the needs of the writer with dyslexia? An examination of functions scaffolding the transcription and proofreading in text production aimed towards researchers and practitioners in education. *Dyslexia (Chichester, England)*, *29*(4), 408–425. https://doi.org/10.1002/dys.1752

Myhill, D., & Jones, S. (2009). How Talk Becomes Text: Investigating the Concept of Oral Rehearsal in Early Years' Classrooms. *British Journal of Educational Studies*, *57*(3), 265–284. https://doi.org/10.1111/j.1467-8527.2009.00438.x

Neumann, A. (2017). Zugänge zur Bestimmung von Textqualität. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 203–219). Waxmann.

Noakes, M. A., Schmitt, A. J., McCallum, E., & Schutte, K. (2019). Speech-to-text assistive technology for the written expression of students with traumatic brain injuries: A single case experimental study. *School Psychology (Washington, D. C.)*, *34*(6), 656–664. https://doi.org/10.1037/spq0000316

Nordström, T., Nilsson, S., Gustafson, S., & Svensson, I. (2019). Assistive technology applications for students with reading difficulties: Special education teachers' experiences and perceptions. *Disability and Rehabilitation. Assistive Technology*, *14*(8), 798–808. https://doi.org/10.1080/17483107.2018.1499142

Nuance Communications. (2024, June 15). *All-New Dragon Home Speech Recognition Version 15. Nuance*. https://www.nuance.com/dragon/dragon-for-pc/home-edition.html

Oesterreicher, W. (2008). Revisited: die 'zerdehnte Sprechsituation'. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, *130*(1), 1–21. https://doi.org/10.1515/bgsl.2008.002

Ok, M. W., Rao, K., Pennington, J., & Ulloa, P. R. (2022). Speech Recognition Technology for Writing: Usage Patterns and Perceptions of Students with High Incidence Disabilities. *Journal of Special Education Technology*, *37*(2), 191–202. https://doi.org/10.1177/0162643420979929

Olive, T. (2014). Toward a Parallel and Cascading Model of the Writing System: A Review of Research on Writing Processes Coordination. *Journal of Writing Research*, *6*(2), 173–194. https://doi.org/10.17239/jowr-2014.06.02.4

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition*, *30*(4), 594–600. https://doi.org/10.3758/BF03194960

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, *19*(2), 139–158. https://doi.org/10.1080/10573560308222

Pajares, F., & Valiante, G. (1997). Influence of Self-Efficacy on Elementary Students' Writing. *The Journal of Educational Research*, *90*(6), 353–360. https://doi.org/10.1080/00220671.1997.10544593

Pajares, F., & Valiante, G. (2001). Gender Differences in Writing Motivation and Achievement of Middle School Students: A Function of Gender Orientation? *Contemporary Educational Psychology*, *26*(3), 366–381. https://doi.org/10.1006/ceps.2000.1069

Palmiter Bajorek, J. (2019). *Voice Recognition Still Has Significant Race and Gender Biases: Technology And Analytics*. https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases

Parker, R. I., & Vannest, K. (2009). An Improved Effect Size for Single-Case Research: Nonoverlap of All Pairs. *Behavior Therapy*, *40*, 357–367. https://doi.org/10.1016/j.beth.2008.10.006

Pennington, J., Wook Ok, M., & Rao, K. (2018). Beyond the Keyboard: A Review of Speech Recognition Technology for Supporting Writing in Schools. *International Journal for Educational Media and Technology*, *12*(2), 47–55. https://ijemt.org/index.php/journal/article/view/199

Perelmutter, B., McGregor, K. K., & Gordon, K. R. (2017). Assistive Technology Interventions for Adolescents and Adults with Learning Disabilities: An Evidence-Based Systematic Review and Meta-Analysis. *Computers & Education*, *114*, 139–163. https://doi.org/10.1016/j.compedu.2017.06.005

Peterson-Karlan, G. R. (2011). Technology to Support Writing by Students with Learning and Academic Disabilities: Recent Research Trends and Findings. *Assistive Technology Outcomes and Benefits*, *7*(1), 39–62.

Petko, D. (2012). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the 'will, skill, tool' model and integrating teachers' constructivist orientations. *Computers & Education*, *58*(4), 1351–1359. https://doi.org/10.1016/j.compedu.2011.12.013

Philipp, M. (2015). *Schreibkompetenz: Komponenten, Sozialisation und Förderung*. A. Francke Verlag/utb. https://doi.org/10.36198/9783838544571

Philipp, M. (2017). Wirksame Schreibförderung – metaanalytische Befunde. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 187–202). Waxmann.

Philipp, M. (2019). *Grundlagen der effektiven Schreibdidaktik und der systematischen schulischen Schreibförderung* (7th edition). Schneider Verlag Hohengehren

Pildal, J., Hróbjartsson, A., Jørgensen, K. J., Hilden, J., Altman, D. G., & Gøtzsche, P. C. (2007). Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *International Journal of Epidemiology*, *36*(4), 847–857. https://doi.org/10.1093/ije/dym087

Piolat, A., Olive, T., Roussey, J.-Y., Thunin, O., & Ziegler, J. C. (1999). SCRIPTKELL: A tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 113–121. https://doi.org/10.3758/bf03207701

Price, K. M., Wigg, K. G., Misener, V. L., Clarke, A., Yeung, N., Blokland, K., Wilkinson, M., Kerr, E. N., Guger, S. L., Lovett, M. W., & Barr, C. L. (2022). Language Difficulties in School-Age Children With Developmental Dyslexia. *Journal of Learning Disabilities*, *55*(3), 200–212. https://doi.org/10.1177/00222194211006207

Pruden, M., Kerkhoff, S. N., Spires, H. A., & Lester, J. (2017). Enhancing Writing Achievement Through a Digital Learning Environment: Case Studies of Three Struggling Adolescent Male Writers. *Reading & Writing Quarterly*, *33*(1), 1–19. https://doi.org/10.1080/10573569.2015.1059780

Quinlan, T. (2004). Speech Recognition Technology and Students With Writing Difficulties: Improving Fluency. *Journal of Educational Psychology*, *96*(2), 337–346. https://doi.org/10.1037/0022-0663.96.2.337

Rädiker, S., & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA: Text, Audio und Video*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-22095-2

Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2014). *Quantitative Methoden 2: Einführung in die Statistik für Psychologen und Sozialwissenschaftler* (4th edition). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43548-9

Regan, K., Evmenova, A. S., Hughes, M. D., Rybicki-Newman, M. P., Gafurov, B., & Mastropieri, M. A. (2021). Technology-mediated writing: It's not how much, but the thought that counts. *E-Learning and Digital Media*, *18*(5), 480–495. https://doi.org/10.1177/2042753021996387

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *The American Psychologist*, *74*(4), 445–458. https://doi.org/10.1037/amp0000356

Richards, L., & Morse, J. M. (2013). *Readme first for a user's guide to qualitative methods* (3rd edition). Sage. https://doi.org/10.4135/9781071909898

Rønneberg, V., Torrance, M., Uppstad, P. H., & Johansson, C. (2022). The process-disruption hypothesis: How spelling and typing skill affects written composition process and product. *Psychological Research*, *86*(7), 2239–2255. https://doi.org/10.1007/s00426-021-01625-z

Rose, J. (2009). *Identifying and Teaching Children and Young People with Dyslexia and Literacy Difficulties: An independent report form Sir Jim Rose to the Secretary of State for Children, Schools and Families*. https://www.thedyslexia-spldtrust.org.uk/media/downloads/inline/the-rose-report.1294933674.pdf

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, *55*(1), 68–78. 10.1037110003-066X.55.1.68

Saddler, B., & Asaro-Saddler, K. (2013). Response to Intervention in Writing: A Suggested Framework for Screening, Intervention, and Progress Monitoring. *Reading & Writing Quarterly*, *29*(1), 20–43. https://doi.org/10.1080/10573569.2013.741945

Saddler, B., & Graham, S. (2007). The Relationship between Writing Knowledge and Writing Performance among More and Less Skilled Writers. *Reading & Writing Quarterly*, *23*(3), 231–247. https://doi.org/10.1080/10573560701277575

Sahli Lozano, C., Greber, L., & Steiner, F. (2016). Der Nachteilsausgleich an Berner Primarschulen: Ergebnisse einer Befragung der Schulleiterinnen und Schulleiter. *Schweizerische Zeitschrift f*ür *Heilpädagogik*, *22*(11), 28–35. https://phrepo.phbern.ch/id/eprint/813

Santangelo, T., & Graham, S. (2016). A Comprehensive Meta-analysis of Handwriting Instruction. *Educational Psychology Review*, *28*(2), 225–265. https://doi.org/10.1007/s10648-015-9335-1

Schäfer, K.-H. (2001). Medienpädagogik als Teildisziplin der Allgemeinen Erziehungswissenschaft. In S. Aufenanger (Ed.), *Jahrbuch Medienpädagogik* (Vol. 1, pp. 17–45). https://doi.org/10.1007/978-3-322-97494-5_2

Scheerer-Neumann, G. (2018). *Lese-Rechtschreib-Schwäche und Legasthenie: Grundlagen, Diagnostik und Förderung* (2nd edition). *Lehren und Lernen*. W. Kohlhammer. https://doi.org/10.17433/978-3-17-034159-3

Scheerer-Neumann, G., Schnitzler, C. D., Hübner, A., Mühlbauaer, E., & Ritter, C. (2010). *ILeA Individuelle Lernstandsanalysen: Lehrerheft 4*. Deutsch (2nd edition).

Schellenberg, C., Hofmann, C., & Georgy-Tscherry, P. (2017). Gerechtere Bildungschancen dank Nachteilsausgleich? Ergebnisse eines Forschungsprojektes zur Umsetzung des Nachteilsausgleichs auf Laufbahnen auf Sekundarstufe II. *Schweizerische Zeitschrift Für Heilpädagogik*, *23*(10), 18–25. https://edudoc.ch/record/226245?ln=de

Schneider, H., & Anskeit, N. (2017). Einsatz digitaler Schreibwerkzeuge. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 283–298). Waxmann.

Schneider, P., & Niederberger, M. (2020). Mixed-Methods-Studien in der Gesundheitsförderung. Ergebnisse eines systematischen Reviews deutschsprachiger Publikationen. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, *153–154*, 10–22. https://doi.org/10.1016/j.zefq.2020.05.002

Schröder, L., Urton, K., Vierbuchen, M.-C., Hertel, S., Knaak, T., Nobel, K., Barwasser, A., Grünke, M., Schulden, M., & Hillenbrand, C. (2022). Digitale Förderung der narrativen Schreibkompetenz bei Schülerinnen und Schülern mit diagnostiziertem Unterstützungsbedarf: Ein systematisches Literatur-Review. *Empirische Sonderpädagogik,* (1), 47–78. https://doi.org/10.25656/01:25530

Schuchardt, K., Fischbach, A., Balke-Melcher, C., & Mähler, C. (2015). Die Komorbidität von Lernschwierigkeiten mit ADHS-Symptomen im Grundschulalter [The comorbidity of learning difficulties and ADHD symptoms in primary-school-age children]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *43*(3), 185–193. https://doi.org/10.1024/1422-4917/a000352

Schüler, L. (2020). Diktieren mit Spracherkennung als Form der medienunterstützten Textproduktion – ein Forschungsbericht. *Didaktik Deutsch*, *48*(2), 71–85. https://doi.org/10.25656/01:21679

Schüler, L. (2021). Schreibflüssigkeit im Medienvergleich: Handschrift – Tastaturschreiben – Diktieren mit Spracherkennung. *Zeitschrift f*ür Angewandte Linguistik, (75), 330–363. https://doi.org/10.1515/zfal-2021-2077

Schulte-Körne, G. (2021). Verpasste Chancen: Die neuen diagnostischen Leitlinien zur Lese-, Rechtschreib- und Rechenstörung der ICD-11 [Missed opportunities: The new ICD-11 diagnostic guidelines for reading, spelling and calculation disorders]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *49*(6), 463–467. https://doi.org/10.1024/1422-4917/a000791

Schulte-Körne, G., & Galuschka, K. (2015). *Diagnostik und Behandlung von Kindern und Jugendlichen mit Lese- und/oder Rechtschreibstörung: Evidenz- und konsensbasierte Leitlinie (S3_abgelaufen* (028–044). Deutsche Gesellschaft für Kinder und Jugendpsychiatrie, Psychosomatik und Psychotherapie e. V. (DGKJP).

Schulte-Körne, G., & Galuschka, K. (2019). *Lese-/Rechtschreibstörung (LRS). Leitfaden Kinder- und Jugendpsychotherapie: Band 26*. Hogrefe.

Sensor Review (Ed.). (1981). *Britain Scores in Easier Man-Machine Communication* (Emerald Backfiles No. 4).

Shanahan, T. (2019). Reading-Writing Connections. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (3rd edition, pp. 298–321). The Guilford Press.

Silber-Varod, V., Siegert, I., Jokisch, O., Sinha, Y., & Geri, N. (2021). A cross-language study of speech recognition systems for English, German, and Hebrew. *Online Journal of Applied Knowledge Management*, *9*(1), 1–15. https://doi.org/10.36965/OJAKM.2021.9(1)1-15

Snowling, M. J. (2019). *Dyslexia//Dyslexia: A Very Short Introduction.* Oxford University Press. https://doi.org/10.1093/actrade/9780198818304.001.0001

Snowling, M. J., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: Past, present and future. *Oxford Review of Education*, *46*(4), 501–513. https://doi.org/10.1080/03054985.2020.1765756

Söderström, S. (2016). Socio-material practices in classrooms that lead to the social participation or social isolation of disabled pupils. *Scandinavian Journal of Disability Research*, *18*(2), 95–105. https://doi.org/10.1080/15017419.2014.972449

Staat Freiburg. (2023). *Nachteilsausgleich in der obligatorischen Schule*. https://www.fr.ch/de/bildung-und-schulen/obligatorische-schule/nachteilsausgleich-in-der-obligatorischen-schule

Steiner, E., & Benesch, M. (2021). *Der Fragebogen: Von der Forschungsidee zur SPSS-Auswertung* (6th edition). facultas/utb. https://doi.org/10.36198/9783838587882

Steinhoff, T. (2023). Der Computer schreibt (mit): Digitales Schreiben mit Word, Whatsapp, ChatGPT & Co. als Koaktivität von Mensch und Maschine. *Medien im Deutschunterricht*, *5*(1). https://doi.org/10.18716/OJS/MIDU/2023.1.4

Sturm, A. (2017). Strategiefokussierte Intervention. In M. Becker-Mrotzek, J. Grabowski, & T. Steinhoff (Eds.), *Forschungshandbuch empirische Schreibdidaktik* (pp. 267–281). Waxmann.

Sturm, A., Nänny, R., & Wyss, S. (2017). Entwicklung hierarchieniedriger Schreibprozesse. In M. Philipp (Ed.), *Handbuch: Schriftspracherwerb und weiterführendes Lesen und Schreiben* (pp. 84–104). Beltz Juventa.

Sturm, A., & Schneider, H. (2019). Rechtschreibung als Teil des Verschriftens und Vertextens: Rechtschreibleistungen im Vergleich von Klasse 5 bis Berufsschule. *Leseforum.Ch*(3), 1–24. https://doi.org/10.26041/fhnw-2029

Sturm, A., & Weder, M. (2016). *Schreibkompetenz, Schreibmotivation, Schreibförderung: Grundlagen und Modelle zum Schreiben als soziale Praxis*. *Lehren lernen*. Kallmeyer in Verbindung mit Klett.

Sturm, J. M., & Rankin-Erickson, J. L. (2002). Effects of Hand-Drawn and Computer-Generated Concept Mapping on the Expository Writing of Middle School Students with Learning Disabilities. *Learning Disabilities Research and Practice*, *17*(2), 124–139. https://doi.org/10.1111/1540-5826.00039

Sumner, E., Connelly, V., & Barnett, A. L. (2014). The influence of spelling ability on handwriting production: Children with and without dyslexia. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(5), 1441–1447. https://doi.org/10.1037/a0035785

Svensson, I., & Lindeblad, E. (2019). 20 års svensk forskning i assisterande teknik – vad har vi lärt oss? *Viden Om Literacy* (26), 24–30. https://www.videnomlaesning.dk/media/2946/26_idor-svensson_emma-lindeblad-1.pdf

Svensson, I., Nordström, T., Lindeblad, E., Gustafson, S., Björn, M., Sand, C., Almgren/Bäck, G., & Nilsson, S. (2021). Effects of assistive technology for students with reading and writing disabilities. *Disability and Rehabilitation. Assistive Technology*, *16*(2), 196–208. https://doi.org/10.1080/17483107.2019.1646821

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive Load Theory, Learning Difficulty, and Instructional Design. *Learning and Instruction*, *4*, 295–312. PII: 0959-4752(94)90003-5

Sweller, J. (2023). The Development of Cognitive Load Theory: Replication Crises and Incorporation of Other Theories Can Lead to Theory Expansion. *Educational Psychology Review*, *35*(4), 1–20. https://doi.org/10.1007/s10648-023-09817-2

Sweller, J., & Chandler, P. (1994). Why Some Material Is Difficult to Learn. *Cognition and Instruction*, *12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Tam, C., Reid, D., Naumann, S., & O' Keefe, B. (2002). Perceived benefits of word prediction intervention on written productivity in children with spina bifida and hydrocephalus. *Occupational Therapy International*, *9*(3), 237–255. https://doi.org/10.1002/oti.167

Tate, T. P., Warschauer, M., & Abedi, J. (2016). The effects of prior computer use on computer-based writing: The 2011 NAEP writing assessment. *Computers & Education*, *101*, 115–131. https://doi.org/10.1016/j.compedu.2016.06.001

Taylor, B. M., Frye, B. J., & Maruyama, G. M. (1990). Time Spent Reading and Reading Growth. *American Educational Research Association*, *27*(2), 351–365. https://www.jstor.org/stable/1163013

TechSmith. (2020). *Camtasia 2024 – The Best Video Editing Software Just Got Better!* www.techsmith.com

Totter, A., Keller-Lee, M., & Rast, C. (2024). Digitale Mediennutzung von Schülerinnen und Schülern im Fremdsprachenunterricht der Primarschule. *Zeitschrift Für Grundschulforschung*, *17*(1), 21–38. https://doi.org/10.1007/s42278-023-00189-8

Transkripto.de. (2024, April 22). *Transkription Service: Wir tippen Ihre Interviews. Transkripto.de*. https://www.transkripto.de/transkription-service

Troia, G. A. (2006). Writing Instruction for Students with Learning Disabilities. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 324–336). The Guilford Press.

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: effects of grade, sex, and ability. *Reading and Writing*, *26*(1), 17–44. https://doi.org/10.1007/s11145-012-9379-2

Troia, G. A., Shankland, R. K., & Wolbers, K. A. (2012). Motivation Research in Writing: Theoretical and Empirical Considerations. *Reading & Writing Quarterly*, *28*(1), 5–28. https://doi.org/10.1080/10573569.2012.632729

Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional Levels of Language Writing Measures in Grades Four to Six. *Written Communication*, *36*(2), 231–266. https://doi.org/10.1177/0741088318819473

United Nations. (2024, May 22). *Convention on the Rights of Persons with Disabilities: Articles, Division for Inclusive Social Development (DISD)*. United Nations. https://social.desa.un.org/issues/disability/crpd/convention-on-the-rights-of-persons-with-disabilities-articles

Usman, S., Arafah, B., Marhum, M., Munir, S., Budi, & Tadeko, N. (2024). The Effect of Technological Pedagogical Content Knowledge (TPACK) on Rural Area Students' English Writing. *Journal of Language Teaching and Research*, *15*(3), 884–892. https://doi.org/10.17507/jltr.1503.21

Variantengrammatik. (2020, January 27). *Variantengrammatik des Standarddeutschen: Theoretische Grundlagen*. http://mediawiki.ids-mannheim.de/VarGra/index.php/Theoretische_Grundlagen

Wallner-Paschon, C., & Haider, G. (Eds.). (2002). *PISA Plus 2000. Thematische Analysen nationaler Projekte*. Studien Verlag.

Weder, M. (2021). Keystroke-Logging und Stimulated-Recall in der Orthographie-Forschung. *Bulletin VALS-ASLA (Association Suisse De Linguistique Appliquée)*(91), 85–104. http://edoc.unibas.ch/49838/

West, J. A. (2019). Using new literacies theory as a lens for analyzing technology-mediated literacy classrooms. *E-Learning and Digital Media*, *16*(2), 151–173. https://doi.org/10.1177/2042753019828355

White, D. H., & Robertson, L. (2015). Implementing assistive technologies: A study on co-learning in the Canadian elementary school context. *Computers in Human Behavior*, *51*, 1268–1275. https://doi.org/10.1016/j.chb.2014.12.003

WHO. (2024, April 30). *ICD-11: for Mortality and Morbidity Statistics*. https://icd.who.int/browse/2024-01/mms/en#1516623224

Wilson, J. (2019). Assessing Writing. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (3rd edition, pp. 321–344). The Guilford Press.

Wimmer, H., & Mayringer, H. (2014). *SLS 2–9: Salzburger Lese-Screening für die Schulstufen 2–9*. Manual. Hogrefe Verlag.

Winkes, J. (2014). *Isolierte Rechtschreibung: Eigenständiges Störungsbild oder leichte Form der Lese-Rechtschreibstörung? Eine Untersuchung der kognitiv-linguistischen Informationsver-*

arbeitungskompetenzen von Kindern mit Schriftspracherwerbsstörungen [Dissertation]. Universität Freiburg, Freiburg.

Winkes, J., & Schaller, P. (2022). Lernverlaufsdiagnostik Schreiben (LVD – Schreiben): Reliabilität, Validität und Sensitivität für mittelfristige Lernfortschritte im deutschsprachigen Raum. *VHNplus (Vierteljahresschrift f*ür Heilpädagogik *und ihre Nachbargebiete)*, *91*(0), 1–26. https://doi.org/10.2378/vhn2022.art22d

Wolcott, W., & Legg, S. M. (1998). *An Overview of Writing Assessment: Theory, Research, and Practice*. National Council of Teachers of English.

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis. *Journal of Learning Disabilities*, *51*(1), 73–84. https://doi.org/10.1177/0022219416688170

Wright, K. L., Hodges, T. S., Dismuke, S., & Boedeker, P. (2020). Writing Motivation and Middle School: An Examination of Changes in Students' Motivation for Writing. *Literacy Research and Instruction*, *59*(2), 148–168. https://doi.org/10.1080/19388071.2020.1720048

Wright, K. L., Hodges, T. S., Enright, E., & Abbott, J. (2021). The relationship between middle and high school students' motivation to write, value of writing, writer self-beliefs, and writing outcomes. *Journal of Writing Research*, *12*(3), 601–623. https://doi.org/10.17239/jowr-2021.12.03.03

Wright, K. L., Hodges, T. S., & McTigue, E. M. (2019). A validation program for the Self-Beliefs, Writing-Beliefs, and Attitude Survey: A measure of adolescents' motivation toward writing. *Assessing Writing*, *39*, 64–78. https://doi.org/10.1016/j.asw.2018.12.004

Yeung, P. (2022). Writing motivation in Chinese children with developmental dyslexia. *Journal of Research in Reading*, *45*(2), 223–252. https://doi.org/10.1111/1467-9817.12403

Abbreviations

AD(H)D	Attention Deficit (Hyperactivity) Disorder
ASR	Automated Speech Recognition
AT	Assistive technology
CBM	Curriculum based measurement
CG	Control group
CIWS	Correct minus incorrect word sequences
CLT	Cognitive Load Theory
CWS	Correct word sequences
%CWS	Percentage of correct word sequences from TWW
DLD	Developmental Language Disorder
EFA	Exploratory factor analysis
EG	Experimental group
GLM	General Linear Model
LD	Learning disabilities
MD	Mathematical disorder
N/n	Number
RD	Reading disorder
RQ	Research Question
SLS 2–9	Salzburger Lesescreening, Grades 2–9
SLT	Speech and Language Therapy
STT	Speech-to-text technology
SWAS	Self-Beliefs, Writing-Beliefs, Attitude Survey
TAU	Treatment as usual
TD	Typically developing peers
TOWL-4	Test of Written Language-4th Edition
TTS	Text-to-speech technology
TWD	Total words dictated
TWW	Total words written
UDL	Universal Design for Learning
WER	Word error rate
WSC	Word spelled correctly

List of Figures

Figure 2.1: Basic components of a writing community (s. Graham, 2018, p. 264) 23
Figure 2.2: Current version of the writing model (Hayes, 2012, p. 371) 26
Figure 2.3: Current version of multicomponent model of working memory (Baddeley, 2012, p. 23) 28
Figure 2.4: Schematic representation of writing processes coordination (Olive, 2014, p. 178) 30
Figure 3.1: Generation of (digital) text from spoken language via microphone and speech recognition technology (own representation) 57
Figure 3.2: Conventional speech recognition architecture based on Filippidou & Moussiades (2020, p. 75) and Karpagavalli et al. (2011, p. 327) 60
Figure 3.3: End-to-end ASR Architecture (Ghayas et al., 2024, p. 150) (simplified representation) 60
Figure 3.4: Visual representation of the assumptions underlying the CLT development (Moreno & Park, 2010, p. 18) 69
Figure 3.5: Proposed model of writing motivation according to Wright et al. (2019, p. 74) 76
Figure 3.6: Technological Pedagogical Content Knowledge according to Mishra and Koehler (2006, p. 1025) 88
Figure 5.1: Illustration of design according to Morse and Niehaus (2009, p.118) 120
Figure 5.2: Timeline of quasi-experimental study with mixed methods design 123
Figure 5.3: Timeline and position of the qual. supplemental component within the study 143
Figure 5.4: Development and application of selected and/or adapted study materials 146
Figure 5.5: Self-efficacy 162
Figure 5.6: Beliefs about writing 162
Figure 5.7: Evaluation model of the interview analysis on STT in the classroom (own representation) 166
Figure 5.8: Development and procurement of study materials 168
Figure 6.1: Development of text quality over time and group, presented by individuals and groups 178
Figure 6.2: Development of text quality over time and group 179
Figure 6.3: Development of text length over time and group, presented by individuals and groups 181
Figure 6.4: Development of writing fluency over time and group, presented by individuals and groups 183

Figure 6.5: Development of reading fluency over time and group, presented by individuals and groups 185
Figure 6.6: Development TWD over time 189
Figure 6.7: Development %CWS over time 189
Figure 6.8: TWD in subgroups device preferences 191
Figure 6.9: %CWS in subgroups device preferences 191
Figure 6.10: Development of writing motivation over time and group 193
Figure 6.11: Code map of challenges of STT in school, number of codes ordered clockwise according to frequency 195
Figure 6.12: Code map of supportive features, number of codes ordered clockwise according to frequency 198
Figure 6.13: Code map of benefits of STT reported by professionals 201

List of Tables

Table 5.1: Sample and students' characteristics 135
Table 5.2: Spearman's rank correlation coefficients between features and characteristics of participants 139
Table 5.3: Comparison of groups according to participant characteristics 140
Table 5.4: Demographics of the interviewees 145
Table 5.5: Interrater reliability of translated TOWL-4 version 150
Table 5.6: Intrarater reliability of translated TOWL-4 version 150
Table 5.7: Correlation matrix (Person's r), spelling and writing fluency measures in handwritten texts 153
Table 5.8: Correlation matrix (Person's r), writing fluency measures in dictated texts 154
Table 5.9: Demographics of questionnaire pilot sample 160
Table 5.10: Values of variance and reliability of factor analyses 161
Table 5.11: Questions asked to EG 1-students 163
Table 5.12: Overview of assessment tools, analysed scores, and settings 167
Table 6.1: Descriptive statistics; M and SD for text quality measures 177
Table 6.2: Main effects and interactions text quality 178
Table 6.3: Descriptive statistics; M and SD of text length 181
Table 6.4: Main effects and interaction in text length 181
Table 6.5: Descriptive statistics; M and SD for writing fluency 182
Table 6.6: Main effects and interaction in writing fluency measures 182
Table 6.7: Descriptive statistics; M and SD for reading fluency 184
Table 6.8: Main effects and interactions in reading fluency 184
Table 6.9: Descriptive statistics, M and SD for writing productivity in different modalities 187
Table 6.10: Descriptive statistics; M and SD for writing accuracy in dictated and handwritten texts 188
Table 6.11: Descriptive statistics, M and SD, of TWD and %CWS among subgroups 189
Table 6.12: Descriptive statistics, M and SD, of TWD and %CWS according to device preferences 190
Table 6.13: Descriptive statistics; M and SD for overall writing motivation 192
Table 6.14: Main effects and interactions in overall writing motivation 192
Table 6.15: Top-level code statistics from eleven interviews 194
Table 6.16: Subcodes of motives and attitudes towards STT among teachers and specialists 200
Table 6.17: Predilection for tool 202
Table 6.18: Students' expectations for dictation in the future 202
Table 6.19: Subcodes for students' emotions when dictating 203