

Young students' motivation to write

An investigation of K-5 students' own
responses and of the self-reports used to capture
their voices

by

Aline Alves-Wold

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)



Universitetet
i Stavanger

Faculty of Arts and Education

Norwegian Reading Centre
2024

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

©2024 Aline Alves-Wold

ISBN: 978-82-8439-234-9

ISSN: 1890-1387

PhD: Thesis UiS No. 757

Cover photo: Illustration developed by the Norwegian Centre for Reading Education and Research on assignment from the Norwegian Directorate for Education and Training (NDET) during the design of a national screening test for reading difficulties.

Acknowledgements

“Você é capaz!” – “You are capable!” – is the best inheritance my mother could have left to our family. Whenever we doubted ourselves, she would always say these words to us, and now, even though she cannot say them anymore, they live within us. She was not a motivation research expert, but she was the greatest motivator we could ever have asked for. She encouraged us to follow our dreams and gave us the belief that we could always accomplish our goals, whatever they were, because we are *capable*. Thank you, mom, our PhD dream has come true, and I wish you were here to celebrate it with me.

However, I humbly acknowledge that the successful realization of this dream would not have been possible without the support of a great team, which I also would like to thank:

Main supervisor Per Henning Uppstad, thank you for your valuable critique and feedback that have pushed me out of my comfort zone. Facing criticism in academia is indeed hard at times, but without the push, one may risk stagnating, and as remarked by the renowned photographer Gina Milicia, “a comfort zone is a beautiful place, but nothing ever grows there.”

Co-supervisor Erin McTigue, thank you for your utmost support, both with your invaluable research expertise, but also with your heartfelt encouragement. There have been some life-changing crises during this project, but you were always there with wise advice and inspiring words that have helped me keep going. I am forever grateful.

Co-supervisor Bente R. Walgermo, thank you for being the glue in this team, and for your unwavering support. You have been a great partner with whom I could talk about everything from project-related issues to the most trivial matters of everyday life. I am very thankful for having someone with whom I could walk so near in this journey.

Co-author Njål Foldnes, thank you for enriching this project with your mathematical expertise, and for your dedication to this project. It has been an uplifting collaboration.

Oddny J. Solheim and Tracey Hodges, thank you for your thorough assessment of my work during my 50% evaluation, and for your detailed comments that have strengthened this thesis.

To all my amazing friends – near and far – and my fantastic colleagues at the Reading Center and IGIS, thank you for your unfaltering support and encouragement. You have been a treasured source of motivation and strength.

I would also like to thank my family for being my solid bedrock. Your unconditional love and support are the fertile ground that allows my dreams to grow. Thank you for your patience and for holding my hand throughout this journey. I would never have come this far without you.

Finally, to each and every one of you who have played a part in the successful completion of this project, I want to express my sincere gratitude. Your indispensable support will forever hold a special place in my heart. Thank You for all these blessings in my life.

Stavanger, 2023

Summary

The present project contributes knowledge of K-5 students' motivation to write and foregrounds the students' voices by examining students' self-reports, like surveys, interviews, and alternative reports (e.g., drawings and written reflections). In this endeavor, three studies were carried out. First, a systematic literature review was conducted, and 56 empirical studies investigating K-5 classrooms (1996-2020) were synthesized. Given the massive data derived from this review and the different focus of the research questions investigated in each study, the results of the review are divided into two articles. The first focused on factors influencing the students' writing motivation, and to inform the development of valid instruments, the second focused on the self-reports used in the reviewed studies. Thereafter, findings from study 2 revealed a need for investigating potential influences of scale format on young students' responses. To test this hypothesis, a quasi-experimental study was conducted with Norwegian first and third graders, and findings are presented in a third article.

In study 1, nine factors influencing young students' writing motivation were identified, and these are presented mnemonically as the *ABCs of Writing Motivation*, containing the following factors: [A] Appeal, [B] Beliefs, [C] Choice, [D] Difficulty, [E] Environment, [F] Feedback, [G] Goals, [H] Help, and [I] Instructor. Practice examples that support the implementation of these factors are also provided in article 1 in table form, which can be used by educators and researchers as a tool for planning writing lessons or interventions. Moreover, to convey transparently about what these factors entail, detailed examples of students' utterances and responses, and of the contexts in which these factors were identified are also provided.

Study 2 investigated the types of writing tasks addressed in the self-reports used in the included studies, the motivation constructs assessed,

and the emphasis given to the students' voices. Results indicate that (a) the genre *storytelling* was used most often to operationalize writing in the studies, (b) although 32 motivation constructs were identified, *attitude* was the construct most often assessed through surveys and interviews, (c) students' voices were weighted differently across studies, and (d) researchers often lacked sufficient rationale documentation for their choices/design of motivation measures.

Based on the widespread use of pictorial supports (e.g., animals, faces, geometrical figures) used in motivation measures with young students identified in study 2, yet lack of sufficient rationales for this design, study 3 worked to systematically compare pictorial supports. Specifically, two types of pictorial support were compared in a 5-point Likert scale: *faces* vs. *circles*. Quantitative results indicate that young students often skip motivation questions and avoid the extreme ends of the scale more often when face-scales are used.

In conclusion, given the inner nature of motivation, the present project shows that self-reports are a rich source of data for providing an insight into K-5 students' perspectives on their motivations to write. However, to capture the students' voices as accurately as possible, further attention is needed regarding the development of valid instruments for this purpose. First, the methodology used in the studies (i.e., quantitative, qualitative, or mixed methods) seems to be related to differences on the emergence of factors identified in study 1, indicating that open-ended questions may give students more room to express their unique views. Second, findings from study 2 indicate that future research assessing students' writing motivation should (1) include more varied writing tasks in the assessments, which more realistically reflect the myriad of activities, tools, and genres that writing entails, and (2) report clearly whose voices are being heard (e.g., teachers, students, or researchers) and the suitability of this choice in relation to the study's goals, design, and findings. Finally, despite the widespread use of face-scales in measures of motivation, findings from study 3 show that this type of

scale has a biasing effect on students' responses, and that circle-scales seem to be more appropriate for measuring young students' interest and self-beliefs. However, more studies are needed to strengthen this finding and to contribute with further guidelines for the development of valid scales.

Table of Contents

Acknowledgements.....	iv
Summary.....	vi
1 Introduction.....	13
1.1 Aims and research questions.....	15
1.1.1 State-of-the-art and knowledge needs: Research gaps addressed.....	16
1.1.2 Empirical connections between the three studies.....	18
1.1.3 Theoretical grounding of the studies.....	20
1.1.4 Contributions to the field of writing motivation.....	20
1.2 Defining key terms of this thesis' narrative.....	21
1.3 Dissertation structure.....	26
2 Theoretical and Empirical Background.....	27
2.1 Writing.....	28
2.1.1 Main theories for conceptualizing writing.....	30
2.1.2 The scope of writing in early literacy settings.....	33
2.2 Motivation to write.....	34
2.2.1 What is motivation?.....	35
2.2.2 Motivation within the writing domain.....	37
2.2.3 Motivation and engagement.....	42
2.3 Assessing writing motivation through students' self-reports.....	44
2.4 Classroom practices for fostering writing motivation.....	46
2.5 Highlighting theoretical links to the present work.....	49
3 Methods.....	50
3.1 Developmental mixed methods in a multiple study program of inquiry.....	50
3.2 Evidence-based education and systematic reviews: Why am I doing it?.....	53
3.2.1 Evidence-based education.....	54
3.2.2 Systematic reviews.....	60
3.2.3 Summary.....	65
3.3 Studies 1 and 2: Systematic literature review.....	66
3.3.1 Data collection.....	66
3.3.2 Data analysis.....	69
3.4 Study 3: Quasi-experimental study.....	76
3.4.1 Data collection.....	76

3.4.2	Data analysis	79
3.5	Ethical considerations	81
3.6	Study trustworthiness.....	83
4	Results.....	86
4.1	Study 1	86
4.1.1	RQ: What factors emerge from K-5 students' self-reports as influencing their motivation to write?	86
4.2	Study 2.....	90
4.2.1	RQ1: What types of writing tasks are addressed in self-reports measuring students' motivation to write?.....	90
4.2.2	RQ2: What motivation constructs are measured and how are they operationalized?	91
4.2.3	RQ3: What emphasis is given to students' voices in the studies?	93
4.3	Study 3	95
4.3.1	RQ: Do students report, or skip reporting, levels of writing interest and self-beliefs differently depending on the type of pictorial support used in the Likert scales? If so, to what extent are the differences explained by grade, gender, and type of motivational construct measured?	95
5	Discussion	97
5.1	Findings regarding the students' own responses on their motivations to write	97
5.1.1	Students' self-reports are a rich source of data.....	97
5.1.2	The decisive role of the teacher.....	99
5.1.3	Motivational factors are intertwined.....	100
5.1.4	Age-appropriate implementations of factors	101
5.2	Findings regarding student self-reports design	102
5.2.1	The type of writing addressed matters.....	102
5.2.2	The definition and operationalization of motivation constructs matter	103
5.2.3	Whose voices are being heard matters.....	104
5.2.4	The type of pictorial support used matters	105
5.3	Implications for classroom practice	106
5.3.1	Teacher planning.....	106
5.3.2	Tasks	108
5.3.3	Tools	108
6	Philosophical reflections	110

6.1	Actuality and potentiality: From potential skills to actual performances	111
6.2	Optimal conditions vary: A thought experiment based on Optimality Theory	113
6.2.1	A brief overview of Optimality Theory	115
6.2.2	Experimenting with motivation factors as OT constraints	117
6.3	The importance of thick descriptions	120
7	Limitations and future directions	123
	References	127
	Appendices	171
	Appendix 1 – Article 1	171
	Appendix 2 – Article 2	226
	Appendix 3 – Article 3	264
	Appendix 4 – Supplementary material: Figure S1	318
	Appendix 5 – Supplementary material: Table S1	320

Table of Figures

Figure 1 – Study focus development from a broader to a narrower perspective as when zooming in with a camera, representing focus levels for articles 1-3.....	16
Figure 2 –Visual representation of the underlying studies of the present thesis.	19
Figure 3 – Illustration of broad landscape of students’ perspectives and limited snapshot of it as captured through self-reports.....	24
Figure 4 – Visual representation of the theoretical anchoring of the three articles that compose this thesis within the writing domain and the research fields of motivation, education, and assessment.	27
Figure 5 – Illustration of a cyclical process between motivation and engagement.	43
Figure 6 – Whitehurst’s model of evidence-based education.	56
Figure 7 – Simplified version of the basic components of a classroom writing community.	73
Figure 8 – Two types of pictorial support: <i>faces</i> and <i>circles</i>	77
Figure 9 – Flipped design for grade 1.....	79

List of Tables

Table 1 – OT tableau structure	116
Table 2 – OT ranking for American English	117
Table 3 – OT ranking for British English	117
Table 4 – Hypothetical OT tableaux structure for optimal writing practices	118
Table 5 – Hypothetical OT ranking for kindergarteners practicing spelling	119
Table 6 – Hypothetical OT ranking for kindergarteners practicing spelling for a spelling contest.....	119

1 Introduction

Literacy skills, including writing and reading, are central in basic education and fundamental for lifelong learning (UNESCO, 2017, p. 9). Writing in particular is considered an essential skill that has been associated with achievement in the educational, social, professional and civic spheres (Graham et al., 2013). However, although reading and writing should “go hand in hand” in early literacy instruction, many researchers have pointed out that writing tends to be deprioritized, not only in the research field, but also in the classroom, especially in lower grades (Barton, 2007; Griffo et al., 2015; Håland et al., 2019; McBride et al., 2017; Read, 2005; Skar et al., 2022).

In part, it can be argued that this negligence has consequences for writing instruction and development, as reports show that despite writing’s importance, many still struggle to develop the necessary skills to attain writing success (Graham, Bollinger, et al., 2012; Troia, 2014; Wyatt-Smith & Jackson, 2016). However, even though we find evidence of an apparent lack of focus on writing in education and research, the complex nature of writing *in itself* is challenging for skill development. That is, becoming a proficient writer requires numerous experiences and takes many years to develop (Bazerman et al., 2017), which can pose challenges not only for the development of writing skills, but also of another indispensable ingredient for writing success, namely *writing motivation*. In fact, motivation is emphasized by UNESCO (2017, p. 7), who claims that the improvement of literacy is a global concern, and progress towards achieving this goal “is only possible if those young people and adults who face literacy challenges are *motivated* to engage in learning.”

Indeed, motivation has lately moved from a peripheral to a central place within educational research and is now seen as a fundamental element for learning (Cook & Artino, 2016). With regards to writing, in

particular, motivation has also recently received increased focus (Camacho et al., 2021; Ekholm et al., 2018; Troia et al., 2012), and practices that may foster writing motivation in educational settings have been proposed (Boscolo & Gelati, 2018; Bruning & Horn, 2000). However, many of the studies investigating writing motivation base their findings on researchers' and/or teachers' observations and may overlook the students' own perspectives regarding their motivations to write.

This lack of focus on students' voices is partially because the use of self-reports, including surveys, questionnaires and interviews, with young children has been largely debated, as some suggest several reliability issues like overestimation of the children's own abilities, and limited capacity and lack of opportunities for comparing themselves with other children and with their own previous performances (Mata, 2011, p. 289). Nevertheless, Sturgess et al. (2002) argue that based on the literature examined, there is compelling evidence supporting the validity, desirability, and usefulness of using self-report measures with young children, and that self-reports have proven to be reliable in various contexts and for various purposes. In addition, various studies have shown that children in primary grades are indeed able to give an indication of their motivations to write (Graham, Berninger, et al., 2012; Mata, 2011), even though some may lack a more robust vocabulary to express themselves (Paquette, 2008).

Thus, to get a more comprehensive picture of writing motivation in educational settings, explorations of the students' own views regarding their motivations to write are also necessary. In fact, according to Hall and Axelrod (2014), if we are to succeed with motivating students to engage with writing activities and become better writers, we need to gain insights into *elementary* school children's motivation to write, and in this quest, it is crucial to include the students' voices in writing research focusing on the affective domain. In addition, given the importance of early years for the development of foundational writing skills and motivation, investigating writing motivation in elementary settings is

primordial. However, in this exploration, valid instruments for capturing the students' voices are necessary.

Turning to the present work, I thus investigate K-5 students' writing motivation by focusing on students' self-reports. In this examination, in addition to investigate what students report regarding their motivations to write, what these self-reports are measuring and how is also explored.

1.1 Aims and research questions

The present thesis *aims to contribute knowledge about what K-5 students indicate regarding their motivations to write* through self-reports, including surveys, questionnaires, and interviews. In addition, in order to generate knowledge regarding the development of valid instruments, the present study also *examines the types of self-reports used for capturing the students' voices, and the potential influence of self-report format on students' responses*.

In this investigation, I move from a broader perspective that explores factors influencing students' writing motivation, as derived from their self-reports (Study 1), towards narrower viewpoints concerning characteristics of self-report design (Study 2), and of Likert-scale design in particular (Study 3), as illustrated in Figure 1. Common for all three studies is the investigation of data through students' self-reports on their motivations to write.

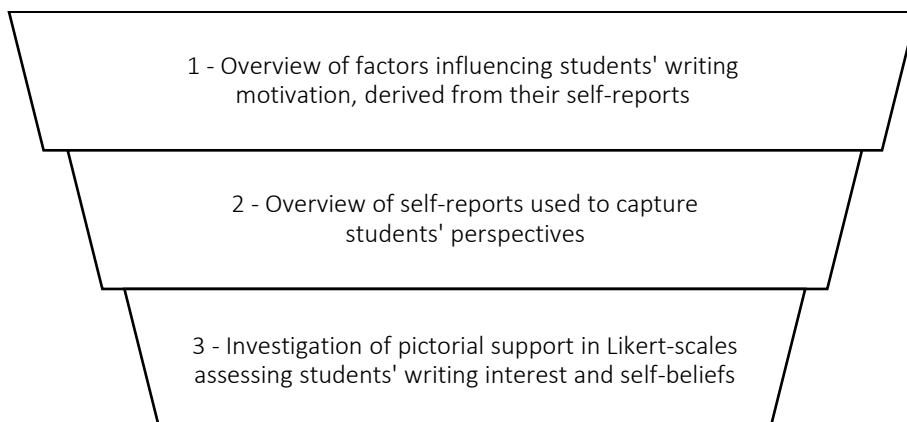


Figure 1 – Study focus development from a broader to a narrower perspective as when zooming in with a camera, representing focus levels for articles 1-3.

1.1.1 State-of-the-art and knowledge needs: Research gaps addressed

Although an impressive body of research on motivation had been promulgated during the 1980s, Hidi and Boscolo (2008, p. 144) argue that motivation research only impacted writing research much later, when writing researchers demonstrated that writing is a complex task that requires not only the coordination and development of *cognitive* skills, but also of *affective* components. The revised framework on writing proposed by Hayes (1996) reflects this new understanding of writing, in which affective components, like goals and beliefs, are given a more predominant role. In fact, according to Alexander and Fox (2004, p. 50), the period from 1996 and onwards represents the “Era of Engaged Learning,” as it marks a change in the way the literacy community views learners and highlights the role of motivation for literacy achievement. Consequently, following this shift, there has been a significant increase in studies focusing on motivation for reading and writing in instructional settings (Boscolo & Gelati, 2018; Graham, Daley, et al., 2018; Guthrie et al., 2004, 2012; Guthrie & Cox, 2001; Hidi & Boscolo, 2008).

In the specific domain of writing, significant advancements in motivation research have been made. For instance, major components of writing motivation have been identified (Troia et al., 2012), the relationship between writing motivation and achievement has been explored (Akyol & Aktaş, 2018; Graham et al., 2007, 2017), clusters of conditions for developing writing motivation in instructional settings have been proposed (Bruning & Horn, 2000), and best practices have been highlighted (Boscolo & Gelati, 2018; Gerde et al., 2012).

Despite these notable advancements in writing motivation research in educational settings, the present thesis has identified some research gaps that still need to be explored. First, study 1 in the present project addresses a current *gap* in the literature, which is that many studies investigating young students' writing motivation base their findings on adult observations and may not include the students' own perspectives regarding motivation in their investigations (e.g., Baker & Lastrapes, 2019; Cheung & Jang, 2019; Cordero et al., 2018; Cremin et al., 2006; Daniels, 2004; Duran & Karatas, 2019; Fox, 2001). In addition, while previous literature reviews have explored writing motivation (Camacho et al., 2020; Ekholm et al., 2018; Klassen, 2002; Troia et al., 2012), none of them has synthesized findings with a specific focus on the students' own perspectives. In other words, existing reviews have synthesized findings from students' own viewpoints and the perspectives of others (e.g., teachers and researchers) as *unisonous*, despite numerous studies indicating differences between these two sources (e.g., Chohan, 2011; Jones et al., 2016; S. D. Miller & Meece, 1997; Paquette et al., 2013).

Second, given that motivation is an internal state, observation alone is not fully valid. To tap into the students' internal states, student self-reported data is also needed. However, to date there is no consensus on standard forms for collecting self-reported data on writing motivation with young students. A myriad of self-reports (including interviews, surveys, and questionnaires) addressing various motivation constructs and writing tasks has been used for this purpose (e.g., Abbott, 2000;

Bayat, 2016; Bradford et al., 2016; Jones et al., 2016; Kanala et al., 2013; Liao et al., 2018), but to date we lack an overview of these instruments. Such an overview is useful for creating a knowledge base and helping the field with identifying guidelines for designing valid instruments, which is the gap addressed in study 2.

The findings of study 2 not only answered questions regarding self-report design, but also opened new questions. An interesting finding from study 2 regarding the widespread use of pictorial supports (e.g., animals, faces, geometrical figures) used in motivation measures with young students, yet lack of sufficient rationales for this design, revealed a gap regarding the validity of these visual aids. This gap is addressed in study 3.

1.1.2 Empirical connections between the three studies

To address the research gaps described above, first, a systematic literature review was conducted. This resulted in an overview of the students' perspectives on writing motivation (study 1), and of the types of self-reports used to capture their voices (study 2). Thereafter, as mentioned above, findings from study 2 revealed a research gap concerning the need for investigating potential influences that the format of scales (i.e., type of visual aid) used for measuring writing motivation may have on young students' responses. To test this hypothesis, a quasi-experimental study was conducted with first and third graders in the Norwegian context.

Figure 2 illustrates how each study contributes to the overarching investigation of the present thesis, and the research questions investigated in each study. Each study led to a single article.

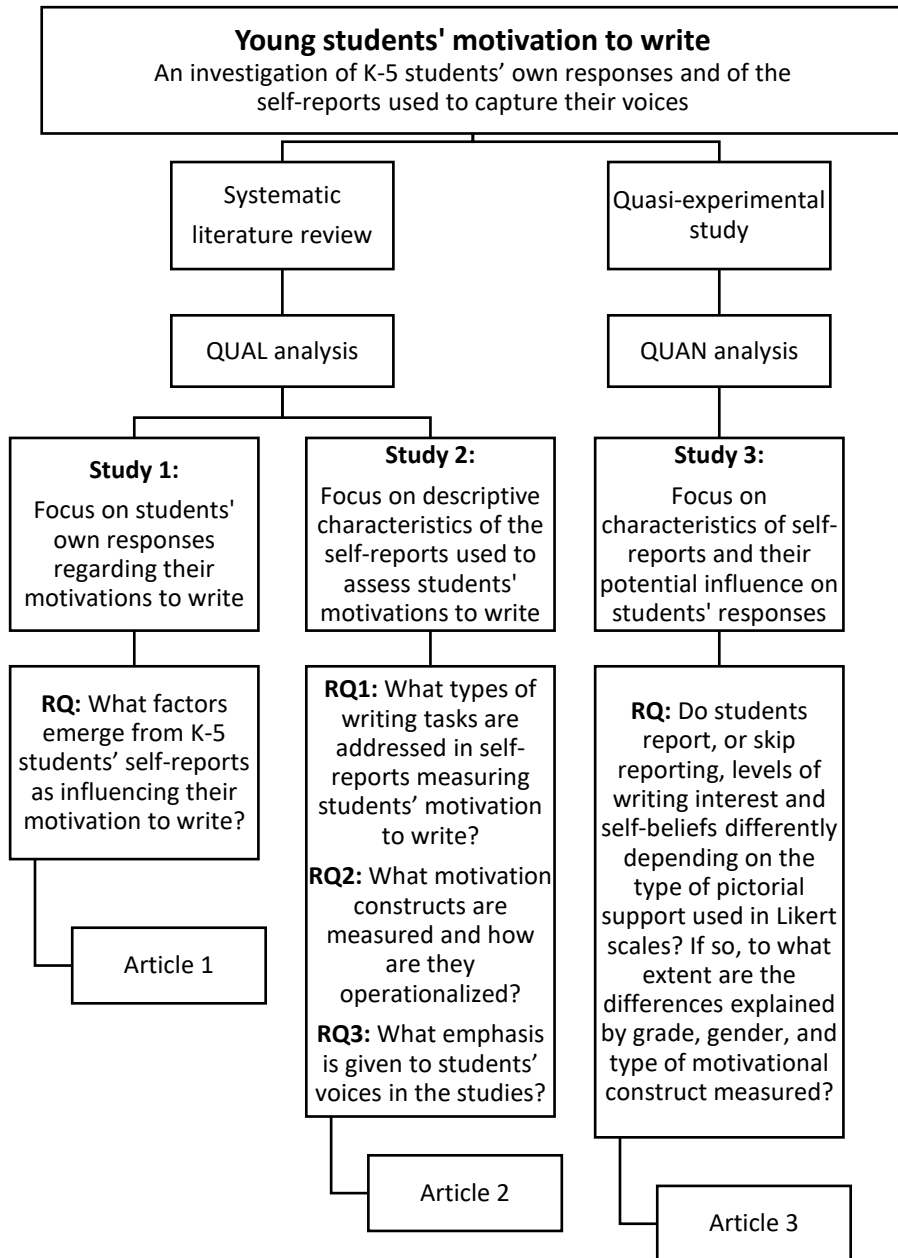


Figure 2 –Visual representation of the underlying studies of the present thesis.

1.1.3 Theoretical grounding of the studies

Due to the nature of the present project, commencing as a systematic literature review, it was essential to consider theory broadly and educate myself on the wide array of writing motivation theories and how they are connected (see section 2 for more detail). A fundamental aspect of doing a systematic review is acknowledging the theories in which each of the studies are grounded. Consequently, the present project requires theoretical breadth and open-mindedness. However, as a writing motivation researcher, it is also essential to identify my own theoretical orientations, which I briefly describe as follows.

According to Hodges (2017), writing theories can be categorized into four main theories: (1) Cognitive process theories, (2) Sociocultural theories, (3) Social cognitive theories, and (4) Ecological theories (see section 2.1 for more detail). My interest in students' perspectives and motivation for writing aligns me with social cognitive theory (Bandura, 1986, 1994) as my primary orientation. Second, as my focus was on writing motivation, it is important to acknowledge that the current schemas and models of writing motivation that informed this work (Boscolo & Gelati, 2018; Graham, 2018; Troia et al., 2012) draw upon broader motivational theories, including self-determination theory (Ryan & Deci, 2000), attribution theory (Weiner, 1986), goal-orientation theory (Dweck & Leggett, 1988; Elliott & Dweck, 1988), and expectancy-value theory (Wigfield & Eccles, 2000), as shown in more detail in section 2.

1.1.4 Contributions to the field of writing motivation

Ultimately, findings from the present thesis can inform both researchers and practitioners, by contributing with:

- 1) information on factors influencing K-5 students' motivation to write, and classroom practices that may facilitate the integration of these factors (article 1),

- 2) a descriptive overview of different types of self-report used for assessing writing motivation in K-5 settings (article 2),
- 3) statistical analysis investigating the potential influence of extraneous factors (e.g., different types of pictorial supports) on first and third graders' responses in self-reports (article 3),
- 4) and recommendations for designing writing motivation self-reports for K-5 students (articles 2 and 3).

The three articles included in the present thesis are as follows:

- Article 1 Alves-Wold, A., Walgermo, B. R., McTigue, E., & Uppstad, P. H. (under review). The ABCs of writing motivation: A systematic review of factors emerging from K-5 students' self-reports as influencing their motivation to write.
- Article 2 Alves-Wold, A., Walgermo, B. R., McTigue, E., & Uppstad, P. H. (2023). Assessing Writing Motivation: A Systematic Review of K-5 Students' Self-Reports. *Educational Psychology Review*, 35(1), 24. <https://doi.org/10.1007/s10648-023-09732-6>
- Article 3 Alves-Wold, A., Walgermo, B. R., & Foldnes, N. (2024). Assessing writing and spelling interest and self-beliefs: Does the type of pictorial support affect first and third graders' responses? *Assessing Writing*, 60, 100833. <https://doi.org/10.1016/j.asw.2024.100833>

1.2 Defining key terms of this thesis' narrative

The choice of the camera metaphor in Figure 1 is not arbitrary, but intentional, as it will be explained in this subsection. Here, I present the reasoning for the choice of four key terms that help unlocking this metaphor and that characterize the rationale of the present thesis. These are: (1) students' voices (2) self-report, (3) capture, and (4) portray.

As proposed by Gonzalez et al. (2017) in their review, the term *students' voices* can entail a myriad of connotations and conceptualizations. However, in essence, the term *voice* can be defined as “an expression of opinion, or the right to express your opinion” (*Definition of VOICE*, 2023). Accordingly, in the present study, the concept of *students' voices* is simply understood as student’s opinions, perspectives, or their right to express their opinions and perspectives. In addition, an important aspect of students’ voices in this thesis is related to *how* students can *express* their opinions. As discussed by Thomson (2011), in understanding ‘voice’, a narrow view of *expression* is often taken, in which voice is often mediated through words, connected to speech acts or events (e.g., speeches, conversations, or answering open-ended questions in interviews), and such a limited view undermines the fact that we all live in multi-mediated environments with diverse possibilities for and means of expression. Following this line of thought, I argue that expressing one’s own opinions, that is, one’s voices, or being given the right to do so, can also be possible through other types of *self-reported data*, like drawings, ratings, scores, or answers to close-ended questions. For example, when students rate how much they enjoy writing or reading specific genres, such ratings provide – even though limited – some insight into the students’ opinions regarding these genres.

This broad understanding of how individuals can express their voices through self-reports in research is derived from a simple contrast: research in which students are invited to express their opinions through any type of self-reported data (e.g., interviews, surveys, drawings) versus research in which students are *not* invited to express their opinions through any type of self-reported data. This contrast is supported by the following definition of *self-report*: “a report about one’s behavior provided especially by one who is a subject of research” (*Definition of SELF-REPORT*, 2023). The key point in this definition of self-report, and which is fundamental for the present thesis, is that opinions

regarding individuals are provided by the *individuals themselves*, in contrast to reports provided by others through, for instance, observation.

There are of course variations in how much room different types of self-reports may give for students to express their voices. For instance, longer interviews including various open-ended questions may give students more room to express their unique views compared to short surveys including perhaps only one close-ended question. However, it is worth noting that open-ended questions do not automatically provide a deeper understanding of students' opinions compared to close-ended questions. The level of insight obtained from self-reports depends on how the questions are formulated and the types of answers provided. In fact, in some cases, close-ended questions can offer a more comprehensive understanding of students' motivation profiles. For example, longer surveys that inquire about specific activities and preferences (e.g., Jones et al., 2016) can sometimes provide more detailed information compared to shorter surveys that only include a single open-ended question, which may not adequately address the topic under investigation. Moreover, even though open-ended questions may appear to offer students more freedom to express their own opinions, students may lack the vocabulary to do so or choose to provide brief responses (e.g., Snyders, 2014), which again limits the depth of insight into their opinions, compared to longer close-ended surveys that include detailed rating scales that may offer more thorough understandings (e.g., Merisuo-Storm, 2006).

Nevertheless, even though some types of self-reports may give more room for students to express their voices, we need to remember that the data captured through self-reports are somewhat limited snapshots of what the landscape of students' perspectives entail, as visually represented in Figure 3.

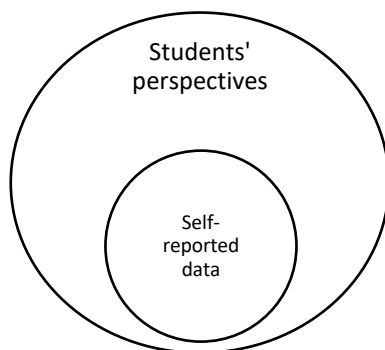


Figure 3 – Illustration of broad landscape of students’ perspectives and limited snapshot of it as captured through self-reports.

Such a *stratified representation* of students’ voices is in line with a central ontological proposition of *critical realism* (CR), in which reality is stratified and science sets out to uncover the essence of those things that may not be immediately evident or obvious to us through our everyday experiences (Bhaskar et al., 2017). In this endeavor, Hoddy (2019) remarks that CR shares features of a *realist ontology* combined with an *interpretive epistemology*. For instance, CR shares with positivism the idea that an objective reality exists and that causal explanations can be produced, but it differs in the degrees to which reality can be observed and in terms of what causation entails. In addition, although CR gives interpretive methods an important role in social research, it emphasizes specific qualifications that contrast to strong *social constructionist* traditions. In sum, CR acknowledges that while there might exist diverse ‘realities’ and modes of ‘knowing,’ certain descriptions and narratives of a particular phenomenon will more closely approximate reality than others (Hoddy, 2019). This leads us to the next two central terms to be defined in relation to this thesis’ narrative, namely the concept of *capturing* and *portraying* something.

The main idea behind the use of these two concepts stems from an analogy to that of *capturing* (taking) a limited/narrow picture (snapshot)

of a broad landscape and *portraying* (exhibiting/showing) that picture to others. In the present study, students' voices are *captured* through self-reports from different studies, i.e., limited snapshots of the landscape of students' perspectives are taken and these captured data are analyzed and *portrayed* in the studies' findings. First, snapshots of the students' perspectives are taken in each study that is included in the systematic literature review conducted for the present thesis, and these are portrayed in article 1 and 2. Then new snapshots are taken in the quasi-experimental study included in this thesis and these are portrayed in article 3. In this analogy, as different camera lenses and choices regarding where to focus on a broad landscape will result in different snapshots, so will various types of self-reports capture different angles of the students' perspectives – some narrower, some broader.

Finally, although similar pictures may be taken by different photographers, how each photographer portrays these pictures may be done very differently. Imagine an exhibition wall, given photographers' understanding of the significance of each snapshot, some may choose to foreground specific pictures and place them in bigger frames on the wall, whereas others might prefer to give the same pictures less focus and display them in smaller frames. The same is true for how researchers portray captured students' voices. As discussed especially in article 2 of the present thesis, when portraying students' perspectives in their studies' findings, some researchers give students' voices more emphasis, whereas others may give teachers' and/or researchers' perspectives more focus. Either way, these choices are a result of the researchers' understanding of which snapshots communicate best the areas they investigate, and which consequently deserve more attention. In the present thesis, this is also my intention, and I seek to present the most significant snapshots derived from the three conducted studies to represent findings regarding K-5 students' motivation to write, including highlights of both their perspectives and of the self-reports used to capture their voices.

1.3 Dissertation structure

In this dissertation, I present an overview of the present project and of the contributions derived from each of its three underlying studies, by providing insight into this project's theoretical and methodological bases, as well as its key findings. In chapter 2, I present the theoretical and empirical foundation for the present thesis, and in chapter 3, I provide an overview of the methods used in the three studies that compose this thesis. In chapter 4, the results derived from these three studies are summarized, and in chapter 5, the main findings of the present thesis are discussed, followed by recommendations regarding classroom implications. In chapter 6, philosophical reflections regarding different aspects of the present thesis are discussed, and in chapter 7, limitations of the present project and suggestions for further research are addressed.

2 Theoretical and Empirical Background

Although the present investigation is anchored within the specific domain of *writing*, consequently relying on theory and research derived from the field of writing, this thesis is also fundamentally supported by three other fields, namely *motivation*, *education*, and *assessment*. However, as the focus of the present thesis shifts from a broader perspective on students' writing motivation and classroom practices that support it (article 1) to a narrower viewpoint directed towards issues of writing motivation assessment (article 2) and self-report design (article 3), the contribution derived from each of these research fields also shifts. As illustrated in Figure 4, shifting the perspective from article 1 to 3, also represents a shift from more dependance on the fields of motivation and education to more dependance on the fields of motivation and assessment, in addition to looser ties to the writing domain.

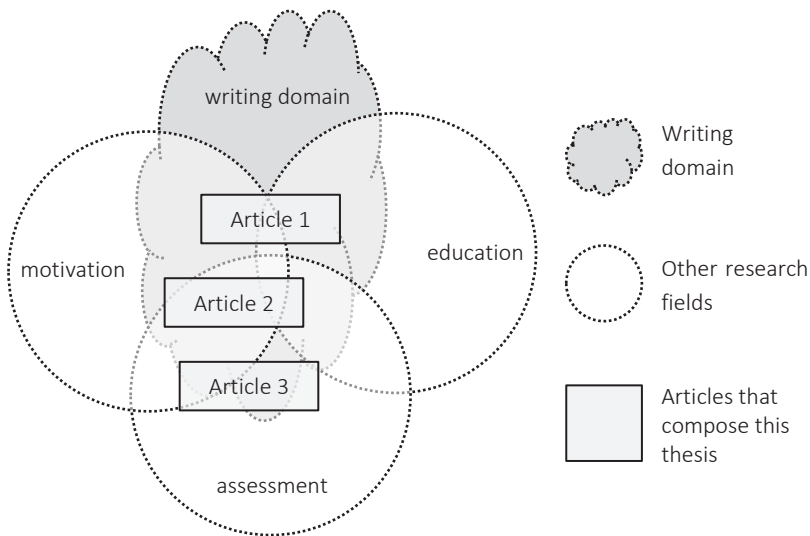


Figure 4 – Visual representation of the theoretical anchoring of the three articles that compose this thesis within the writing domain and the research fields of motivation, education, and assessment.

In this section, drawing on relevant research from these four fields, I will present the theoretical and empirical background for the present project.

2.1 Writing

In the same way that Wolf (2018, p. 18) refers to reading as “an unnatural cultural invention,” Olson (2009, p. 13) refers to writing as “a set of signs invented to represent speech.” The aspect of *invention* in these accounts of both reading and writing is crucial for understanding writing in the present thesis, as it implies that reading and writing are not *innate human abilities*, but *skills* that need to be *learned* and *developed*. However, this does not mean that writing, being a skill, does not require abilities. In fact, human abilities provide the foundation upon which skills are built, meaning that skills can be seen as the manifestation of abilities. For example, the skill of running is dependent on the ability of moving. Nevertheless, possessing an ability does not necessarily guarantee the development of a skill, as skill requires effort and practice for learning how to coordinate the necessary abilities effectively.

In historical terms, Tønnessen and Uppstad (2015) claim that skill is often seen as the performance and execution of mechanical or technical tasks in an automatic, virtually flawless manner, but relying on a combination of the behaviorist concept of *automaticity*, and the concept of *awareness* derived from cognitive psychology, the authors propose that every skill requires the ability to switch between automaticity and awareness. For instance, for an individual running in an even terrain in a straight path, the execution of this skill may be characterized by higher levels of automaticity, but if the path becomes more uneven and less predictable, more awareness is needed to avoid missteps or accidents. Finally, the authors remark that to improve skills – and implicitly the ability to switch effectively between automaticity and awareness – practice is necessary.

In neurological terms, Wolf (2018, pp. 17–18) explains that a basic human ability like *oral language* has specialized genes that unfold with minimal support, which results in our capacity to speak, comprehend, and think through words. However, in contrast to oral language, there is no ‘genetic blueprint’ for reading. This skill needs to be learned; and it can be manifested because of the flexibility of the human brain and its ability to connect different regions (e.g., language, vision, and cognition) to go beyond its original functions and develop completely new circuits that make the development of skills, like reading, writing, or dancing, possible.

A similar distinction between the development of oral language and reading is proposed for distinguishing speech and writing. Olson (2009) argues that although visual signs and any form of speaking are generally natural ways of communicating, writing is not an innate communicative ability like speech, as it evolved slowly throughout history and is usually acquired late by children, commonly requiring explicit instruction.

However, differently from Wolf (2018) and Olson (2009), I argue that although *language* is indeed a human ability, both oral and written language, as we know them today, are indeed inventions, since they are the result of evolutive creations and conventions that are constantly adapted to meet the needs of the individuals and societies that make use of these types of languages to communicate. For instance, no matter where in the world, newborn babies produce similar sounds with their vocal tract (i.e., cry) to signalize – or communicate – needs, like hunger or thirst. These are innate linguistic human abilities. Later, children learn to use words like ‘food’, ‘comida’ (Portuguese), or ‘mat’ (Norwegian), orally or in written form, to communicate similar needs. These are inventions that vary depending on cultural and societal conventions.

Following this line of thought, writing (like reading and speaking) is thus seen as a skill that needs to be *learned* and *developed*. Moreover, as remarked by many, this skill needs to be continually developed

throughout the lifespan (Bazerman et al., 2017; Camacho et al., 2021; Graham et al., 2022). However, as it will be discussed in subsection 2.1.1, given the magnitude of, among others, processes, mechanisms, purposes, tools, and contexts involved in the manifestation and execution of this skill, writing is also seen as a *social practice* (Deane, 2018; Graham, 2018). In addition, given the large semantic widening that this word has undergone (“Writing, n.,” n.d.), ‘writing’ can also designate, for instance, a problem-solving act (Camacho et al., 2021), a learning tool (Kieft et al., 2006), a social activity (Graham, 2018), a complex task (Graham et al., 2005), or a system of signs (Olson, 2009).

In the next subsection, I will present the four main theories of writing and present the rationale for choosing *social cognitive theory* to conceptualize writing in the present thesis. Then in section 2.1.2, I will limit the scope of writing in early literacy settings.

2.1.1 Main theories for conceptualizing writing

According to Hodges (2017), some theories of writing place greater emphasis on mechanics and form, while others emphasize creativity and sociability, and the focus on these different aspects of writing has given rise to four main theories of writing: (1) Cognitive process theory, (2) Sociocultural theory, (3) Social cognitive theory, and (4) Ecological theory. In this section, I will first present a brief overview of these four major theories of writing, then I will provide the rationale for this project’s theoretical anchoring, in which social cognitive theory (Bandura, 1986; Graham, 2018) and ecology theory (Barton, 2007) are combined to conceptualize writing.¹

Cognitive process theory was one of the first acknowledged theories to conceptualize writing and it focused on the importance of cognitive

¹ Parts of this theoretical rationale in subsection 2.1.1 were written in relation to a paper submission for a Ph.D. course on literacy at the University of Stavanger, but this text has never been published before.

processes (Hodges, 2017). In this theory, writing is seen as a complex system of interconnected cognitive processes, as proposed by Flower and Hayes (1981) in their model of writing. Prior to their model, writing was seen as a linear operation where writers moved forward through progressive steps in the process of writing. However, as remarked by Hodges (2017), Flower and Hayes' model of writing as an iterative activity aligns more closely with the actual processes involved in writing and composition, in which different steps of the writing process, such as drafting and editing, are continuously revisited by writers.

On the one hand, although cognitive process theory acknowledges the importance of mental operations in the individual's brain during the writing process, it focuses solely on individuals, and it neglects the influence of external factors, such as motivation and social interactions. *Sociocultural theory*, on the other hand, stems from the work of Vygotsky (1978) who emphasized the fundamental role of social interactions for children's learning, and who proposed that children learn through their interactions with more knowledgeable others. In sociocultural theory, writing is then seen as a social – collaborative – activity where, for instance, writers in the beginner level can learn from more experienced ones. In addition, through the lens of sociocultural theory, writing is not only seen as a subject that students need to excel in, but also as a tool for learning in other subjects, as when students write, for example, summaries, notes, or mind-maps (Hodges, 2017).

In sum, cognitive process theory thus focuses on individuals and the mental processes they undergo in the writing activity, whereas sociocultural theory focuses on the social aspects of the writing process. As suggested by Brand (1998), writing researchers thus claim that 'theoretical balance' is achieved through *social cognition*, where the latter term stands for the *individual* and the former for the *community*. That is, in *social cognitive theory* both individual cognitive processes and social interactions are accounted for. In line with such a social cognitive understanding of writing, Graham et al. (2013) define writing

as both a *cognitive activity* that requires the management of various elements related to the writer and the writing environment in a skillful manner, and a *social activity* influenced by purpose, audience, culture, history, and society. To depict such an understanding of writing, Graham (2018) suggests that a writing model that incorporates both of these strands is likely to yield a more comprehensive and richer understanding of what writing is. Such models have been proposed, among others, by Hayes (1996) and Graham (2018).

Finally, it is important to note that social interactions, both in sociocultural and social cognitive perspectives, can also be understood through *ecology theory*. Ecology, which stems from biology, is the study of the *interrelationship* of organisms and their environments. Thus, an ecological approach to writing looks at interaction from a two-way perspective. That is, it looks at how a human activity, such as writing, being part of an environment not only is influenced by the environment around it, but also influences it at the same time. According to Barton (2007), an ecological approach investigates the social and psychological context of human activities in a manner that gives room for change. The author then argues that rather than focusing on isolated skills underlying reading and writing, this approach advocates for a shift into the study of literacy – an array of social practices linked to specific symbol systems and their corresponding technology.

2.1.1.1 Seminal theory of writing influencing the present project

Given its all-encompassing status that acknowledges both *individual abilities* and *social interactions*, social cognitive theory is thus used to conceptualize writing in my project. Writing is then here defined as a skill that involves the effective use of language, textual conventions, and tools, in various social contexts for different purposes (Graham, 2018). This definition reflects the integration of individual cognitive abilities, like critical thinking, planning, and creative expression, with the influence of social interactions. As individuals observe and interact with

others, while developing and adapting their writing behaviors within varied social contexts, cognitive functions play a crucial role in shaping the quality and effectiveness of their written communication.

2.1.2 The scope of writing in early literacy settings

In addition to conceptualizing writing through a social cognitive perspective, my study is also guided by a conceptual framework that sees writing as a skill in which it is necessary to develop and coordinate specific component skills (Bingham et al., 2017). As remarked by Bingham et al. (2017), for more proficient writers, such as adults and adolescents, writing entails the coordination of both lower order skills, such as spelling and transcription, and higher order skills, such as ideation (e.g. text planning and revision).

Although writing is a highly complex task (Graham, Liu, et al., 2018; Hayes, 1996; Hayes & Flower, 1980; Nystrand, 1989), in educational settings, simple views of writing are often maintained focusing on component skills required in the writing activity. These are conventionally divided into three main components: *handwriting*, *spelling* and *composing* skills (Berninger et al., 1992; Bingham et al., 2017; Kaderavek et al., 2009), which will be briefly defined in the next paragraph.

Handwriting, which requires cognitive skills, and visual motor integration, refers to the ability of forming letters properly (Berninger et al., 1992, 2006). As remarked by Bingham et al. (2017), perfection is of course not expected in the earliest stages of writing development, but an initial awareness with regards to specific letter shapes or letter-like forms, linearity, and how forms are separated. Given the increase on the use of digital tools for writing, also in the alphabetization process, it is important to note that *typewriting* can also be juxtaposed with handwriting as a component skill. *Spelling* refers to one's ability to put letters together to form words and requires morphological, phonological

and orthographic knowledge (Kaderavek et al., 2009). For novice writers, orthographic knowledge is limited, thus so called ‘invented spellings’ (i.e., unconventional word spelling) are common. *Composing*, sometimes referred to as ‘writing for communication’ (Håland et al., 2019), requires higher order skills and is defined by Gerde et al. (2012, p. 351) as the “activity of expressing ideas, opinions and views in print.” Even in the lowest stages of writing development, as rudimentary as a composition can be, it requires thinking skills as a writer needs to create ideas and use his or her encoding abilities (i.e., handwriting/ typewriting and spelling skills) to transform these ideas into written text.

Following this three-pronged distinction among component skills, writing essentially requires one’s ability to (1) form letters (handwriting/ typewriting), (2) put them together into words (spelling), and (3) put words together into texts to convey meaning (composing). Thus, learning and practicing each of these three components is paramount for children’s writing development. However, as pointed out by Gerde et al. (2012), in early literacy settings, writing is often confounded with neat handwriting or penmanship, and sometimes too much focus is placed on handwriting and spelling tasks that are disconnected from composition practice. Many researchers have therefore recommended more focus on composition in early elementary education, and argued that composition makes writing meaningful (Bingham et al., 2017; Gerde et al., 2012; Håland et al., 2019). However, although fundamental, practicing the abovementioned component skills is not enough for children’s writing development, i.e., early literacy success is determined by both *skill* and *motivation* (Graham et al., 2007; McTigue et al., 2019), which will be introduced in the next section.

2.2 Motivation to write

In this section, I will define the motivation construct and present relevant motivation theories and studies within the writing domain, followed by a note on the relationship between *motivation* and *engagement*.

2.2.1 What is motivation?

Motivation, also referred to as drive (Troia et al., 2012), can be defined as any force whereby behavior is energized and directed, in which behavior derives its intensity, strength and persistence from energy, while purpose and goal-directedness are instilled by direction in behavior (Reeve, 2012). This means that motivation is not simply a matter of a present or absent force, like an ON/OFF switch, where one can only either be motivated or not, rather, it has different degrees of intensity. Motivation can therefore also be seen through a quantitative perspective (Boscolo & Gelati, 2018), where students can demonstrate higher or lower levels of motivation to write.

Motivation is the result of a complex process derived from various sources, such as goals, beliefs, values, needs and environmental circumstances (Reeve, 2012; Schunk & Mullen, 2012). This complex process has been studied through various influential motivation theories, in which each has shed light on important aspects and sources of motivation. Here, I will briefly present the core principles of five of these theories, because of their significance for motivation research in the writing domain, which will be discussed in the next section.

The first theory is *social cognitive theory* according to Bandura (1986, 1994, 1997). In this theory, *self-efficacy* is a central aspect in which the belief in one's ability to successfully execute and control behaviors influences motivation, performance, and resilience in the face of challenges. The second is *self-determination theory* (Ryan & Deci, 2000), which posits that human motivation is driven by three innate psychological needs. These are *autonomy*, *competence*, and *relatedness*, and individuals are more likely to be intrinsically motivated when these needs are satisfied. Third, *attribution theory* (Weiner, 1986), which seeks to understand how individuals interpret and *attribute causes* to events, often relying on *internal* or *external factors*, to make sense of their own and others' behaviors. Fourth, *goal-orientation theory* (Dweck &

Leggett, 1988; Elliott & Dweck, 1988), which suggests that individuals' motivation and behavior are influenced by their orientations towards *mastery/learning goals* (i.e., seeking knowledge and improvement) or *performance goals* (i.e., demonstrating competence and avoiding failure). Finally, *expectancy-value theory* (Wigfield & Eccles, 2000), which proposes that decisions related to achievement are driven by a combination of individuals' expectations of success and the subjective value they place in tasks in specific domains. For instance, children may be more inclined to engage in an activity when they expect to succeed in that activity and they see the value in that activity (Leaper, 2011).

Although, each of these theories has advanced our understanding of general elements of motivation, in accordance with other researchers, Wigfield (1997, p. 59) argues that “motivational constructs vary across domain and should be studied at that level.” Studying motivation within a domain-specific approach is crucial because it allows for a nuanced understanding of the unique factors influencing behavior within a particular context, which can lead to more targeted and effective interventions or strategies for fostering motivation in that specific domain. For example, although foundational components of motivation like value, goals, and interest may be found in different domains (e.g., writing, reading, and dancing), operationalizing these constructs and fostering motivation in distinct domains will take form differently, given the peculiarities and activities that characterize each domain. For the present thesis, investigations of motivation *within the specific domain of writing* are therefore of relevance, and these will be presented in the next section. However, despite particularities pertaining to the specific domain of writing, as it will become clear for the reader, the five seminal theories of motivation presented above have indeed impacted greatly and laid the foundation for studies of motivation within this domain.

2.2.2 Motivation within the writing domain

Following domain-specific views on motivation, many researchers have identified important aspects of motivation within the domain of writing (Graham, 2018; Hidi & Boscolo, 2008; Troia et al., 2012). In this section, I will present some of the most significant contributions, and which are of great relevance for the present thesis. First, I will introduce the work of Troia (2012), in which *four broad components* of motivation are proposed. Second, I will present the work of Graham (2018), where he suggests *seven sets of beliefs* that influence whether one engages with writing activities or not. Then, I will present the work of Boscolo and Gelati (2018), where *three main factors* affecting students' motivation to write are proposed. As it will be pinpointed, there is clear overlap among these three works, as they rely on the same seminal motivation theories presented above. Finally, the centrality of *beliefs* also becomes evident in all three propositions, which, I argue, places social cognitive theory (Bandura, 1986) in a privileged position, compared to other theories of motivation.

After reviewing motivation research in writing, Troia et al. (2012) propose that researchers have identified *four broad components* of motivation. First, *self-efficacy beliefs* (derived from the work of Bandura, 1986, 1994), also referred to as perceived competence (Hughes et al., 2011), in which positive levels of self-efficacy are often related with better writing outcomes. These beliefs play a crucial role for students' writing motivation, because their perceptions of competence regarding writing skills (e.g., spelling, planning, structuring texts) across various writing tasks (e.g., writing essays, creating posters) impact the quality of and persistence in their writing endeavors. Second, *goal orientations* (derived from the work of Dweck & Leggett, 1988; Elliott & Dweck, 1988; Harackiewicz et al., 2002), including 'mastery/learning goals' (aiming to improve one's writing skills/competence), and 'performance goals' (aiming to show superior writing ability than others, and/or to receive public appraisal). Although both types of goals may help students

persist with writing activities when faced with challenges, performance goals may also lead to maladaptive behaviors, like task avoidance (Troia et al., 2012). That is, instead of choosing tasks that they do not master in order to practice and learn (learning goals), students may choose to refrain from these tasks to avoid demonstrating incompetence (avoidance performance goals).

Third, *task interest* and *value*, which can influence the types of goals that individuals set for themselves. On the one hand, *interest* (derived from the work of Hidi, 1990; Hidi et al., 2002) can be divided between ‘personal interest’, derived from individual preferences, and ‘situational interest’, which arises from contextual characteristics of a task. For instance, a child may not be so interested in writing as a general activity (i.e., personal interest) but writing a persuasive letter to his parents to convince them of buying a toy may spark his interest for this specific writing task (i.e., situational interest). On the other hand, *value* (derived from the work of Eccles et al., 1983; Wigfield & Eccles, 2000), can be divided into ‘attainment value’ (e.g., how important a writing task is), ‘intrinsic value’ (e.g., how enjoyable a writing task is), ‘utility value’ (e.g., how connected the writing task is to one’s goals), and ‘cost’ (e.g., how much has to be given up to execute the writing task). According to Troia et al. (2012), task interest and value are closely related, but can operate independently because although a person may find a writing task (e.g., creating a poster) very interesting, little value may be assigned to this task, or a person may assign much value to a task (e.g., writing informational texts), but not be interested in it. However, they can also trigger each other. That is, the value assigned to a task, like the perceived importance of writing catchy titles, can spark one’s interest in that task, while genuine interest in a topic, like describing dinosaurs, can increase the perceived value in that writing task.

Finally, *outcome attributions* (derived from the work of Weiner, 1986) refer to an individual’s beliefs about the reasons for one’s success and failure in their writing endeavors. These beliefs can be influenced by

three dimensions: *controllability* (i.e., one's perception of how much control one has over a cause), *locus* (i.e., whether one attributes reasons for success to internal factors, like effort, or to external factors, such as luck), and *stability* (i.e., whether one sees the reasons for success as fixed or malleable). As remarked by Troia et al. (2012), students who struggle often attribute writing success or failure to factors outside of their voluntary control, instead of linking them to their own efforts (or lack thereof). Such a maladaptive pattern may lead to writing aversion and avoidance of tasks that require extensive writing, since these students see themselves as incapable of improving their writing achievements.

Troia et al. (2012) then propose a schema to portray the interrelationship between these four main components of motivation (e.g., interest) and associated constructs (e.g., personal interest). However, although their schema is an attempt to portray the interrelationship among motivation constructs, the authors note that many links, like the causal pathways between interest, self-efficacy, and value, are still not clear. Troia et al. (2012) then invite researchers to combine different research methods to help untangling some of the interrelationships among various theoretical components of motivation and their connections to writing outcomes. Therefore, it is important to acknowledge that although writing researchers have made progress in understanding motivation at the level of the writer, there is still much needing to be explored through research. Moreover, although the schema presented by Troia et al. (2012) portray theorized connections among *motivation components*, it does not link these to *writing mechanisms*. Thus, to get a better picture of how motivation constructs interact with other elements of the writing process, such as *knowledge* and *skills*, a broader schema including writing mechanisms is necessary.

In his revised writer(s)-within-community (WWC) model of writing, Graham (2018) seeks to depict such a complexity and his model includes *affective and social influences*, as well as *cognitive mechanisms in writing*. The author also points out that motivation resources are at play

at different dimensions, such as through a writer's purposes and actions, or in relation to physical and social environments that can affect motivation within a writing community. Graham (2018, p. 266) then proposes *seven broad sets of beliefs* that influence whether one engages in writing or not, and these will be briefly presented here.

The first set of beliefs includes judgments about the value of a writing task (including attainment, intrinsic, and utility value, as well as cost). The second set concerns whether one likes writing or not or how much writing is viewed as an attractive activity. In addition to pointing out the difference between personal and situational interest regarding writing activities, Graham (2018) also remarks that individuals might hold more or less positive attitudes towards writing in general or towards specific writing tasks or circumstances. The third set refers to perceptions of competence, including self-efficacy beliefs, and can be influenced by previous experiences, but also from beliefs regarding whether ability is fixed or malleable depending on one's effort. The fourth set concerns the reasons why one engages in writing. At one level, it can be because writing is experienced as enjoyable or inherently satisfying (intrinsic motivation), or because one wishes to be rewarded or avoid punishment (extrinsic motivation). At another level, reasons for engaging in writing can also be related to one's goal orientations (including mastery and performance goals).

The fifth set relates to one's beliefs about why one succeeds or fails, including the three dimensions of controllability, locus, and stability. The sixth set focuses on the beliefs one has about his or her writer identities. These identities can be multiple, depending for example on the writing context and purpose, and can also be influenced by other identities, like one's ethnicity, gender, and culture. Finally, the last set concerns beliefs individuals develop about writing communities, including views on the community's values, purposes, audiences, actions, and the tools utilized by the community. Views about a community's identity are also shaped

by an individual's interactions inside a community and one's social belonging and climate.

In addition to these *seven sets of beliefs* proposed by Graham's (2018), and the *four broad components* of motivation highlighted by Troia et al. (2012), Boscolo and Gelati (2018) suggest *three main factors* that influence students' motivation to write: (1) attractiveness and value of the task, (2) perceived writing competence, and (3) beliefs about writing.

As the reader can infer, there is evident overlap among the propositions of Troia et al. (2012), Graham (2018), and Boscolo and Gelati (2018), which is logical since these suggestions stem from the same influential motivation theories, as presented above. However, there are also significant differences among these three major contributions. For example, although all three works remark the important role of *attitude* as an affective influence, Troia et al. (2012) argue that attitude is not a core theoretical component of motivation but is rather an affective positioning towards an activity, instead of the desire to engage in it.

In fact, as remarked by Ekholm et al. (2018) in their review of writing attitudes, there is no consensus in the literature regarding the relationship between attitude and motivation, which has consequences for how both constructs are conceptualized and studied. Nevertheless, despite the lack of consensus on the conceptualization of these two constructs, positive levels of both motivation and attitude have been associated with desirable writing outcomes (Camacho et al., 2021; Ekholm et al., 2018; Klassen, 2002b; Troia et al., 2012). Consequently, many researchers have investigated and recommended practices that may impact positively students' writing motivation (and/or attitude), and a brief overview of such recommendations will be presented in section 2.4. However, before proceeding to classroom practice suggestions, some brief comments on the relationship between *motivation* and *engagement* will be presented, followed by a short presentation of issues related to the assessment of students' writing motivation through self-reports.

2.2.3 Motivation and engagement

The terms *motivation* and *engagement* are often used interchangeably, but as remarked by Martin et al. (2017) inadequately or incorrectly merging or distinguishing motivation and engagement can sustain theoretical uncertainty, pose validity issues for measurement and research, and establish an erroneous groundwork for instructional interventions. Providing a detailed account of the differences (and similarities) of these two constructs is beyond the scope of the present thesis, but a concise definition to draw a fuzzy contour around these constructs and give an indication of how they are understood in the current study is here presented. For this matter, in line with Martin et al. (2017, p. 150), motivation is seen as “the inclination, energy, emotion, and drive relevant to learning, working effectively, and achieving,” while “engagement is defined as the behaviors that reflect this inclination, energy, emotion, and drive.” To illustrate this point, a kindergartner who is *motivated* to learn how to write his own name, can be said to have the drive or will to *engage* with the necessary activities to perform this task. A behavior that could reflect this kindergartner’s drive (motivation) to learn, can for example be his eager practice of the letters necessary to write his name (engagement with the writing activity).

In this example, motivation precedes engagement. That is, motivation is the drive that leads to subsequent engagement. However, Martin et al. (2017) argue that although longitudinal data supports an operational sequencing in which motivation serves as a driving force for subsequent engagement, prior engagement also seems to account for considerable variation in subsequent motivation, suggesting rather a cyclical process, as illustrated in Figure 5. For instance, even though a student initially may be unmotivated to work with a specific task, when engaged with this task for example because of obligations, the student may become motivated to keep on working with this task, and even reengage with it in later occasions.

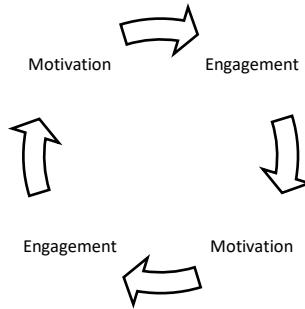


Figure 5 – Illustration of a cyclical process between motivation and engagement.

The abovementioned illustrations and concise definition of motivation and engagement practically equate engagement with *behavior*, but there are other aspects of engagement which are also important to remark. For instance, despite extensive debates on what constitutes *engagement* (Reschly & Christenson, 2012), there is general acceptance that, like motivation, engagement is a multifaceted construct, comprising at least three broad components: behavioral engagement, cognitive engagement, and emotional engagement (Pettersen et al., 2023; Toshalis & Nakkula, 2012). In educational settings, *behavioral engagement* often refers to observable behaviors like participation and involvement, whereas *cognitive engagement* concerns mental processes like self-regulation, choice of appropriate strategies for completing different tasks, and learning goals. *Emotional engagement*, in turn, concerns both positive (e.g., excitement, enjoyment) and negative (e.g., nervousness, frustration) emotions and reactions to, among others, tasks, classmates, and teachers (Pettersen et al., 2023).

Although each of these three components represent different potential manifestations of engagement, when imagining an *engaged student* in a classroom, it is common again to focus on behavior and picture someone who is active, participative, and involved in the school activities. In contrast, when picturing an *unengaged student*, it is easy to imagine someone who looks bored, disengaged from school activities, and who

may even be staring out of a window. That is, in everyday understandings of engagement, it is easier – and common – to focus on observable aspects of engagement, like behavior. However, as described above, engagement also encompasses less observable aspects, like mental processes and emotions. Thus, judging individuals' engagement based on observation alone can be misleading. For instance, a student staring out of a window during a writing task may seemingly look disengaged, even bored, but may in fact be very engaged cognitively, imagining characters and scenarios, and planning a plot. Consequently, to tap into the less observable aspects of engagement and of students' inner drives, observation alone is insufficient. Therefore, in the present thesis, the value of students' self-reports as a rich source of data for this purpose is highlighted.

2.3 Assessing writing motivation through students' self-reports

As mentioned in the introduction of this thesis, the utilization of self-reports as a research method for capturing the voices of young children has been debated due to concerns about reliability. However, many studies have highlighted the significance of children's self-reported data (Hall & Axelrod, 2014; Nolen, 2007; Zumbrunn et al., 2017). In addition, from the literature reviewed, Sturgess et al. (2002) have gathered compelling evidence supporting the validity, desirability, and benefits of self-reports with young children in various contexts. Studies have also shown that, despite potential vocabulary limitations, even children in primary grades can provide indications of their motivations to write (Graham, Berninger, et al., 2012; Mata, 2011; Paquette, 2008).

To develop practices that meet the students' needs and contribute positively to their writing motivation, we need to assess what the students' needs are, including for instance their motivation levels and beliefs. For this purpose, researchers have used various types of self-report methods, depending on the information they want to investigate.

For investigating students' levels of motivation *quantitatively*, methods like surveys (e.g., Bayat, 2016; Liao et al., 2018) and questionnaires (e.g., Jones et al., 2016; Kanala et al., 2013) are commonly used. These often comprise mostly close-ended questions and are administered with large samples. However, for *qualitative* investigations, methods that allow students higher degrees of openness in their answers are preferred, like interviews (e.g., Abbott, 2000; Beck & Fetherston, 2003), and written responses (e.g., Bradford et al., 2016; Erdoğan & Erdoğan, 2013) and these are often administered with smaller samples.

Given the young age of the students, adaptations regarding the language used in the self-reports and the format of the assessment have been implemented by researchers. For instance, given the young students' potential lack of vocabulary to express themselves through words, some other aesthetic forms of expression, like drawings (e.g., Zumbrunn et al., 2017), may be used to collect the students' perspectives. In addition, although Likert-scales in surveys conducted with adults may only display text options ranging from 'not at all' to 'very much', similar scales for children are often accompanied by some type of *pictorial support*, like smileys ranging from sad to happy (Chambers & Craig, 1998; Jones et al., 2016) to aid the students with providing their responses. However, as it will be discussed in study 3 of the present thesis, the potential influence of such pictorial supports on the students' responses is often not sufficiently addressed in the studies assessing students' writing motivation. For instance, students' preferences for choosing specific emojis, or desire to avoid other emojis, may pose a validity threat to these assessments. Moreover, as it will be discussed in study 2 of this thesis, there is still no consensus on standard guidelines for developing self-reports for assessing children's writing motivations, and to move the field forward, more attention needs to be drawn to these issues.

Finally, although necessary, it is important to remember that assessing students' writing motivation through self-reports in educational settings

is not a *goal*; it is a *means*. That is, the *goal* is to foster students' writing motivation, and using students' self-reports is a *means* of collecting data that can help researchers and practitioners get a better understanding of the students' needs with regards to their motivations to write. Then, based on findings derived from the self-reports, as well as from other data sources, like classroom observations and teacher/parent interviews, researchers and practitioners can design classroom practices that may impact positively students' writing motivation levels. In the next section, a brief overview of current classroom practice recommendations for fostering students' motivation to write will be presented.

2.4 Classroom practices for fostering writing motivation

Following the advancement of writing motivation research in educational settings, instructional recommendations tend to reflect the main elements proposed by these advancements. For instance, for each of the four theorized components of motivation proposed by Troia et al. (2012) (i.e., self-efficacy beliefs, goal orientations, interest and value, and outcome attributions), the authors also suggest a list of instructional recommendations. To enhance students' *self-efficacy beliefs*, teachers can provide sufficient scaffolding and give students opportunities to master challenging tasks. To encourage students to adopt *mastery goal orientations*, teachers can help students set concrete, proximal and ambitious goals for themselves, while focusing on personal mastery and improvement (Troia et al., 2012). To increase students' *interest* and *value*, teachers can for example help students' connect their personal experiences to what is learned and make explicit the value of what they are learning (Troia et al., 2012). Finally, to promote adaptive *outcome attributions*, the authors recommend fostering the belief that ability can be altered depending on the students' efforts.

Bruning and Horn (2000), in turn, propose *four clusters of conditions* that are necessary for developing writing motivation in the classroom:

(a) nurturing students' positive self-beliefs and their beliefs about the value of writing, (b) using authentic writing tasks and goals that foster student engagement, (c) creating a supportive context that helps students develop the necessary skills for engaging in and succeeding with writing tasks, and (d) creating a positive environment, where unnecessary stress and anxiety connected to writing activities are eliminated.

These clusters were indeed highlighted by Camacho et al. (2021), as aligning with the teaching practices identified in their review. However, the authors remark that not all motivation constructs investigated in the studies were changed by teaching practices, as some motivation constructs seemed more malleable (e.g., situational interest) than others (e.g., self-efficacy beliefs). In addition, Camacho et al. (2021) also bring attention to the *time issue*, arguing that some of the interventions investigated in the studies were not long enough, and in line with Klassen's (2002b) remarks, they posit that it is challenging to alter *deeply-rooted beliefs* about writing through short interventions.

Yet, this concept of 'deeply rooted beliefs' about writing, although oft repeated as an explanation when interventions are not successful, deserves further consideration for the age group addressed in the present review. In K-5 settings, given the young age of the students and their limited opportunities for developing *deeply* rooted beliefs, it is arguable how *deep* their beliefs about writing actually are. That is, even less malleable motivation constructs, such as self-beliefs, may be more malleable for students in this age group, or even less fixed, compared to older students.

In fact, in a meta-analysis targeting the effect of interventions on reading self-efficacy conducted by Unrau et al. (2018), grade level was found to be a significant moderator of effect size. Meaning that the lower the grade level was, the stronger were the effects of the interventions on students' reader self-efficacy. Even if self-efficacy for reading and writing are not directly comparable, results from Unrau et al.'s meta-

analysis (2018) illustrate that domain specific self-efficacy seems to be most malleable in the earlier grades. This finding is also in line with Bandura's (1997) proposition that self-efficacy is particularly vulnerable in early phases of skill development.

However, although this means that younger students may be more susceptible for modifying their beliefs, this does not mean that fostering students' positive beliefs about writing or about themselves is a quick or easy task, rather it takes time and intentionality. In fact, Boscolo and Gelati (2018) emphasize the fundamental role of the teacher in stimulating motivation to write, and argue that the strategies a teacher employs are decisive for helping students with instilling a sense of mastery in their writing capabilities and fostering a positive attitude towards writing. The authors then propose a summary of teacher 'moves' or strategies for helping students with valuing writing, like providing students with meaningful opportunities to write individually and collaboratively, and to compare literacies in an out of school. The authors also argue that, in exemplary practices that have the potential to foster writing motivation, teachers help students with seeing the value of writing, provide students with *attractive writing tasks*, and support them with the *management of writing skills* (Boscolo & Gelati, 2018).

The role of the teacher is also highlighted by Graham (2018), who argues that although each member of a writing community (e.g., writers, collaborators, and teachers) will have the same basic cognitive structure (control mechanisms, long-term memory and production processes), they also have to rely on other types of *knowledge, emotions and beliefs* to perform their roles. Teachers, for example, need knowledge about adequate teaching practices and how to organize the teaching environment in a way that caters for their students' needs and characteristics. In addition, as remarked by Graham (2018), teachers need to manage their own emotions and beliefs about writing, as these may also influence the students' own beliefs. For instance, Zumbrunn et

al. (2019) found a positive relationship between student-perceived teacher writing enjoyment and the students' own writing enjoyment.

2.5 Highlighting theoretical links to the present work

The schemas and models of writing motivation presented in this section served as the basis for significant aspects of the present project. First, the motivation constructs proposed in these works (e.g., self-efficacy, value, interest), and the writing tasks and skills highlighted as pertaining to earlier educational settings (e.g., handwriting, typing, spelling) were used for identifying relevant search terms for both *motivation* and *writing* in studies 1 and 2 (see fig. 1 on page 23 of article 2). Second, the motivation components derived from the three main contributions of motivation research within the writing domain presented in this section (Boscolo & Gelati, 2018; Graham, 2018; Troia et al., 2012) provided a starting point for coding data deductively in study 1, and the work of Troia et al. (2012) specifically served as a framework for sorting into five categories the constructs of motivation identified in study 2.

Study 3 differs in focus from the previous systematic literature review in study 1 and 2, as it focuses on a subskill of writing for early writers, namely spelling. The reason for focusing on spelling in study 3 is because this study is derived from a national literacy screening test that aims to identify students at risk of developing literacy difficulties, and spelling has demonstrated to be a powerful predictor. In fact, as remarked by Troia et al. (2012, p. 13) “lower level transcription skills such as spelling and handwriting, exert a powerful influence on how well students accomplish composing tasks when these skills are underdeveloped.” Returning to the theoretical constructs of motivation, in study 3 we chose to focus on *interest* and *self-beliefs*, given that underachieving students – the target group of screening tests – often also report lower levels of interest and uncalibrated self-beliefs (Graham et al., 1993; Klassen, 2002a).

3 Methods

The methods used in the present project have been well detailed in the three articles that compose this thesis, but in this chapter, I provide more justification for their selection and for methodological considerations that were not extensively covered in the articles.

3.1 *Developmental mixed methods in a multiple study program of inquiry*

The present project adopts a mixed-methods paradigm (Johnson et al., 2007), in which studies 1 and 2 rely on *qualitative* methods of research, while study 3 relies on *quantitative* methods, as shown in Figure 2 in the ‘Introduction’ section of this dissertation. In this manner, the present project is not defined with a particular mixed methods design (i.e., a single study), but in accordance with mixed methods definitions in which researchers can integrate or connect quantitative and qualitative data in either a single study or in related studies in a multiple study program of inquiry (see for example definitions provided by Creswell, and Johnson and Onwuegbuzie in Johnson et al., 2007, pp. 119–120). For the present project, as it will be explained in more detail below, an *intentional sequence* where quantitative methods are used as follow-up to explore previous qualitative results in a program of inquiry comprising multiple studies (Johnson & Onwuegbuzie, 2004; Plano Clark, 2017; Tashakkori & Teddlie, 2008) is adopted.

Qualitative and *quantitative* research methods are often presented as opposing paradigms and consequently incommensurate (Hitchcock & Onwuegbuzie, 2020), as according to Sale et al. (2002), each of these methods are based on a specific paradigm, that is, a structured set of assumptions related to the nature of reality (ontology), understanding that reality (epistemology), and the specific methods of gaining knowledge about that reality (methodology). On the one hand,

quantitative research stems from *positivism*, which is based on the ontological assumption that there exists only one truth, a reality that is objective and that exists independently of the perceptions of humans (Sale et al., 2002). Epistemologically, a phenomenon can be studied objectively without being influenced by the researcher, and common methods in this paradigm are, among others, randomized control trials and surveys. *Qualitative research*, on the other hand, stems from *interpretivism* and *constructivism*, which ontologically assumes that multiple realities or truths can emerge depending on individuals' constructions of reality (Sale et al., 2002). Epistemologically, findings are co-constructed as the researcher and the object of study interact, and common methods of investigation are, for instance, interviews and participant observation.

Despite debates regarding the incompatibility of these methods (Hitchcock & Onwuegbuzie, 2020), qualitative and quantitative methods of research are paradigmatically compatible due to their complementary nature in addressing research questions (Mertens & Hesse-Biber, 2012). When used together, qualitative and quantitative methods can provide a more comprehensive and robust understanding of research problems, as qualitative data can generate hypotheses for quantitative testing, and quantitative findings can be enriched and contextualized by qualitative insights. This integrated approach allows for triangulation, validation, and increased reliability of research findings, making qualitative and quantitative methods compatible within a paradigm like critical realism, which can be seen as a middle way between positivism and interpretivism (Zachariadis et al., 2013).

Such an integrative approach also seems to be the most appropriate method of research for investigating the research questions explored in this thesis, where qualitative data derived from study 2 generated hypotheses to be tested quantitatively in study 3. That is, in the present project, a sequential combination of qualitative methods followed by a quantitative inquiry, also referred to as 'QUAL → quan' design (Johnson

& Onwuegbuzie, 2004), was chosen for *developmental purposes*, in which one strand provided hypotheses to be tested sequentially in the other (Tashakkori & Teddlie, 2008).

It is also worth noting that these methods of inquiry have been used to generate findings that have practical applications for educational contexts, as outlined in section 1.1. Such an orientation is in line with a *pragmatic philosophy* – common in education research, in which research is oriented towards understanding real-world phenomena and solving practical problems (Johnson & Onwuegbuzie, 2004). However, as discussed here and in section 1.2, at a theoretical level, the present project tends towards a critical realist philosophy.

To investigate the research questions in studies 1 and 2, a systematic literature review was conducted, and results were qualitatively analyzed. In these studies, as described in more detail in articles 1 and 2, I was responsible for the literature search, the screening of the articles in all three phases of the process, the adaptation of the methodological quality score (MQS) for assessing the quality of the included studies, the coding of the data, the calculation of interrater reliability (after the second author of these studies had double coded the same data), and the analysis of the data. I was also responsible for providing a complete draft for article 1 and a semi-complete draft for article 2, to be revised in cooperation with my co-authors, and which I was responsible for finalizing after revision.

For study 3, a quasi-experimental study was conducted, and results were quantitatively analyzed. Data for this study was collected from a bigger project (coordinated by the second author of this article) in which a new national literacy screening test was piloted with first and third grade students in Norway. As mentioned earlier, the idea for this study was derived from a research gap identified in study 2, in which it was revealed a need for investigating potential influences of scale format on young students' responses. In this study, I collaborated with the second author of this article to plan the project design, research questions, and

to make choices regarding the theoretical and empirical foundation of the article, while the quantitative analyses in R (R Core Team, 2022) and description of analyses and results were executed in collaboration with the third author of this article. Finally, the methodological descriptions and discussion of findings was executed in collaboration with the second author of this article, and I was responsible for the progression, revision, and finalization of the article.

Given the weight that the systematic literature review conducted in studies 1 and 2 has for the present thesis' findings, this method deserves further attention with regards to its significance as a method of research inquiry. In the next section, I will therefore explain in more detail what this method entails, its epistemological value for evidence-based education, and why I chose it for the present project.

3.2 *Evidence-based education and systematic reviews: Why am I doing it?*

In education there is simply not enough evidence on the effects and effectiveness of what teachers do in classrooms to provide an evidence-based corpus of knowledge. The failure of educational researchers, with a few exceptions, to create a substantial body of knowledge equivalent to evidence-based medicine means that teaching is not – and never will be – a research-based profession unless there is major change in the kind of research that is done in education. (Hargreaves, 1996, p. 4)

Professor David Hargreaves' keynote speech at the Teacher Training Agency Annual Conference in 1996, stirred up a debate among the research community as he portrayed educational research and teaching practice in a negative light, compared to that of medicine (Hammersley, 1997; Hemsley-Brown & Sharp, 2003, p. 449; Norris, 1996). According to Hargreaves (1996), professional decisions in medicine are based on the best research evidence available, ensuring that patients get the best

possible treatment. However, he argued, teachers seldom rely on research to guide their practices and decisions about what works best for their students. He then called for better cooperation between researchers and teachers and more effective ways of disseminating findings from educational research to practitioners.

Although many have argued that Hargreaves' comparison is unwarranted (Biesta, 2007, 2010; Hammersley, 1997) and pointed out limitations regarding evidence-based approaches to educational research and practice (Chahal, 2017; Cowen, 2019; Mockler & Groundwater-Smith, 2017), evidence-based education has gained much attention and support (Davies, 1999; Gough, 2013; Hemsley-Brown & Sharp, 2003; Hunter, 2017; Slavin, 2002) and continues to increase in popularity. In addition, ways to facilitate research dissemination to practitioners have been proposed, such as systematic reviews (Davies, 2000).

In this section, I present some of the debate on evidence-based education and give the reasoning as to why I have to conduct a systematic review, which I argue can be a good way of communicating research findings in more feasible ways not only to educational researchers, but also to practitioners and policymakers.

3.2.1 *Evidence-based education*

The term *evidence-based education* (hereafter EBE) is relatively new, and concerns approaches to *research* (evidence-based educational research), *practice* (evidence-based educational practice) and *policy* (evidence-based educational policy). It is commonly referred to as the 'what works' agenda (Kvernbekk, 2017), and it basically entails that professional decisions about *what works best* in educational settings should be informed by the *best available evidence*. As remarked by Munro (2014), many might then wonder what is actually new about this term, as for instance policymakers have always relied on evidence – including scientific evidence – when making decisions. However, she

argues, what is new about this concept is that it not only encourages professionals to make use of evidence when making decisions, but it entails a shift from *ideologically* driven reasoning to *rational* decision-making. A shift that is quite clear in Hargreaves' call for the use of scientific research evidence to support and justify professional practices rather than basing them on tradition, prejudice, dogma and ideology (1996).

According to Davies (1999), the idea of EBE is inspired by the Master's program in Evidence-Based Health Care at the University of Oxford, where health professionals are offered the opportunity of further developing their professional skills while still working as practitioners. In this program, students are encouraged to learn by solving problems of a clinical and population-based nature that they bring themselves to the course (Davies, 1999), which is an attempt to close the gap between research and practice. That is, research might have greater *practical relevance* if it is in accordance with practitioners' needs. Therefore, evidence-based approaches are commonly referred to as the 'what works'-agenda, as through this perspective, researchers are encouraged to help practitioners find out *what helps solving the problems* they encounter in their fields.

As remarked by Kvernbekk (2017), looking at EBE through a 'what works'-perspective means that EBE is causal in nature, as it seeks to investigate and determine which actions might produce the best results. That is, by studying the effects of action X (e.g., repeated reading) on Y (e.g., text comprehension), one might conclude that X works. Such a simplistic view of EBE has indeed been criticized, as determining causal relations in the social domain is not as straightforward – if even possible, as argued by Biesta (2010). However, as proposed by Kvernbekk (2017), by interpreting X as an INUS condition (i.e., Insufficient, but Necessary part of an Unnecessary but Sufficient condition) introduces significant complexity to the straightforward X-Y scenario, potentially mitigating

some of the concerns raised by critics of evidence-based practice (hereafter EBP) who argue that EBP is intrinsically oversimplified.

In addition, the term ‘evidence-based’ has been criticized for being *deterministic*, that is, it suggests that professional decisions should be determined by research. However, even though alternative terms like *evidence-informed education* have been proposed (Nelson & Campbell, 2017), ‘evidence-based’ is still predominantly used in the literature. Moreover, as remarked by Davies (1999), the idea of EBE is not to replace professional judgement with evidence in the decision-making process, but to unite these two facets of knowledge to lay a solid ground for taking action. Indeed, even though evidence might suggest that ‘X works for Y’, professionals still need to judge whether X might also work for Y *in particular contexts*, which means that professional judgement is indispensable. In fact, former US assistant secretary of education Grover J. Whitehurst (2002) presents a model of EBE where he equates the importance of what he calls *professional wisdom* (i.e., individual experience and consensus) with that of *empirical evidence* (i.e., findings derived from scientific research and empirical studies), as shown below:

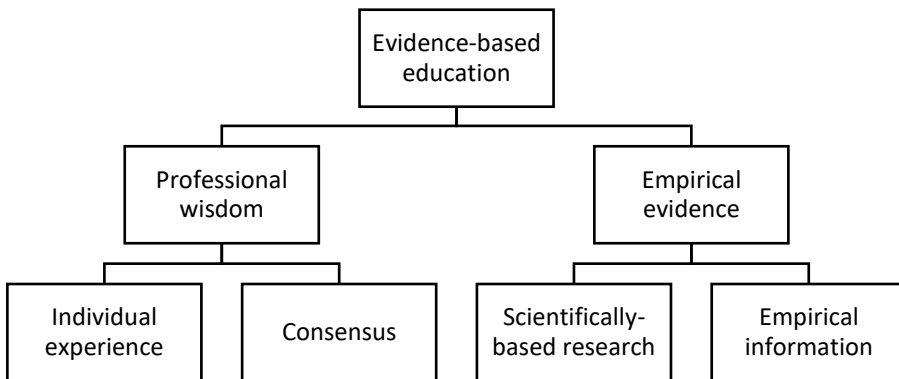


Figure 6 – Whitehurst’s model of evidence-based education.
Source: Adapted from Whitehurst (2002, p. 6). Adapted and reprinted with permission.

Whitehurst (2002, p. 3) then defines EBE as “the integration of professional wisdom with the best available empirical evidence in making decisions about how to deliver instruction.” He argues that, on the one hand, professional wisdom is necessary to adapt scientific evidence to local contexts and to act in areas where research evidence is absent or incomplete, whereas on the other hand, empirical evidence is necessary to resolve competing approaches, generate extensive cumulative knowledge and avoid bias (Whitehurst, 2002).

However, note that in Whitehurst’s definition of EBE, the term ‘*best* empirical evidence’ is used, which suggests that empirical evidence has different degrees of quality. In fact, to assist professionals in finding the *best* evidence available, various organizations that gather, review and judge the quality of research findings have been established. For instance, the U.S. Department of Education established in 2002 the ‘What Works Clearinghouse’ (WWC), an organization that reviews educational research, determines which studies meet rigorous standards, and summarizes findings. According to their website (WWC, n.d.), their goal is “to provide educators with the information they need to make evidence-based decisions” and by focusing on results from *high-quality research*, they set out to answer the question “What works in education?”

Usually, determining what counts as ‘high-quality research’ and ‘best evidence’ is done by rating the quality of the methods through which the evidence has been derived. Such an approach has led to the development of *hierarchies of evidence*, where randomized control trials (RCTs) and systematic reviews are commonly found on the top of most hierarchies, as findings derived from these methods of research are considered the gold standard for providing evidence that supports causal assertions (Munro, 2014). According to Munro (2014), this is based on the assumption that such research methods are considered superior because they are more likely to minimize bias and eliminate alternative reasonable interpretations of the results.

Thus, as proposed by Davies (1999), evidence operates at two levels: (1) *establishing* evidence (e.g., WWC’s aim at establishing ‘what works in education’) and (2) *using* evidence (e.g., teachers using the evidence established from WWC to make professional decisions about instruction). Indeed, as pointed out by Munro (2014), being designated as evidence is not an impartial, objective fact, but a result of an individual, in a specific context, determining its relevance as evidence for their own purposes. For example, findings from a reading intervention become evidence only when someone decides to rely on them to support their claims about the importance of repeated reading for struggling readers. Such remarks have led to some criticism on EBE approaches, claiming that evidence can be used in a *selective manner*, i.e., one can choose to focus on specific evidence to support one’s claims and downplay evidence from competing sources. In addition, many have shown concern about evidence being used in a *managerialist manner* to impose education policies on schools (Chahal, 2017; Lather, 2004; McKnight & Morgan, 2019). For instance, Lather (2004, p. 769) argues that in new managerialism there is a tendency for objectives to come first and evidence to be generated later to justify them. McKnight and Morgan (2019, p. 10) then question “How can educators avoid practitioner wisdom being obliterated by topdown managerialism?”

This is a paradox, as according to Cutspec (2004, p. 2), “[t]he need for evidence-based policies is based on the observation that politicians, government agencies, and others repeatedly impose policy on schools without, or with limited, evidence about the likely costs and benefits.” In this sense, evidence is thus seen as both the ‘means’ through which managerialist policies can be imposed on practitioners and the ‘antidote’ against it. Indeed, like evidence in law can be used by both prosecutors and defense attorneys, so can evidence be used by different stakeholders in education. However, for a fair debate to happen, it is crucial that evidence is made available to all parts. EBE is an attempt to make this

connection possible, and for EBE to be successful, *knowledge sharing* is key.

In fact, even though many of Hargreaves' strong claims (1996) about the need for EBE were heavily refuted, many agreed about the existence of issues regarding research dissemination. For instance, Cutspec (2004, p. 2) argues that findings from educational research are often presented in a way that is not user-friendly for teachers, and given the local knowledge that teachers have, they should be more active in the research process, by for example assisting in framing research questions and interpreting data, rather than being mere recipients of research findings. Moreover, as remarked by Davies (2000), disseminating educational research has also been problematic given that its findings lack cumulativeness. As a result, many ways of synthesizing research and disseminating its findings more effectively have been proposed. For instance, Hemsley-Brown and Sharp (2003, p. 449) point out that communication networks have been created to provide better links between researchers and practitioners, which consequently has led to a greater involvement of practitioners in the research process.

In addition, as pointed out by Gough et al. (2012), navigating the myriad of single studies and research reports in a quest for the best evidence would be too demanding and time-consuming – if at all possible – for practitioners, policymakers and people in general. Therefore, they rely on educational researchers to keep up to date with the growing literature, review it and make its findings available in more manageable formats.

Synthesizing research findings in this manner is usually done through systematic reviews; a method that according to Davies (2000) allows educationalists to move past the constraints of individual studies and to uncover both consistencies and variations in apparently similar investigations, consequently enabling some degree of knowledge accumulation in educational research.

Given that there was a research gap, in which it was revealed the need for a comprehensive overview (i.e., cumulative knowledge) on the students' own responses regarding their motivations to write, and of the types of self-reports used to capture their voices, I decided to conduct a systematic review on these topics. In addition, this method was appropriate for the present project, as it aligns with goals set for the project. First, given that systematic reviews have proven to be a promising way not only for synthesizing, but also for disseminating research findings, it aligned with the goal of providing useful overviews not only for researchers, but also for teachers, who are also an important audience to this work. For instance, in article 1, I aimed to organize the results of the synthesis in a manner that promotes communication with teachers, which is why the factors identified in study 1 were organized according to the first nine letters of the alphabet, as a mnemonic device. Second, being able to synthesize results from a significant number of studies, provided a useful overview for researchers developing similar types of self-reports. However, despite its strengths, this method of research also has its limitations, which will be discussed in the next section.

3.2.2 *Systematic reviews*

A *systematic review* is a relatively new method of synthesizing research evidence. It requires the use of explicit and systematic methods to identify, select, and critically evaluate research in order to answer a clearly formulated question. What is new about this type of review is the *transparency* of its methods. As remarked by Gough et al. (2012), a few decades ago, most reviews of research were not conducted in a systematic way, i.e., reviewers did not always attempt to identify all relevant research, check the quality and reliability of the reviewed studies, or report their conclusions in a transparent manner.

Traditionally, reviewers presented research findings by summarizing what was known about a topic, however, even though they usually gave

details about the studies they were reviewing, they often did not provide the criteria they used to identify and select studies, nor the reasoning for including and discussing some studies and not others.

If inclusion and exclusion criteria are not explicitly specified, it is not possible to judge the appropriateness of the reviewer's decisions or how rigorous and consistent they were. Consequently, it is also not possible to judge the significance of the review findings (Gough et al., 2012). Hence, the aim of conducting a review in a systematic manner is to present its methods in an explicit and transparent way, which will help the reader make sense of its findings. Indeed, as mentioned in the previous subsection, systematic reviews are seen as a promising method for reducing bias and eliminating alternative plausible explanations. However, as with any method of research it has its strengths and weaknesses, and these will be discussed below under three subsections: *methodology, population size and context, and time and effort.*

3.2.2.1 Methodology

A systematic review is usually conducted in three main phases: (1) clearly defining a research question, (2) identifying and selecting relevant studies based on clearly predefined inclusion and exclusion criteria, and (3) interpreting the data and synthesizing the findings. As described above, in this type of review, it is crucial to clearly state the methodology used for identifying, selecting, and evaluating research, which for studies 1 and 2 is presented in detail in the Methods-section of articles 1 and 2.

Being explicit about these methods is an important factor to reduce bias. That is, by looking for evidence to support one's claims, reviewers might actively search for studies that support their views. As remarked by Munro (2014), this does not necessarily indicate conscious misguidance, but it might stem from biases derived from the author's familiarity with certain studies or the usual tendency of scrutinizing more rigidly the

methods of a study in which one does not agree with the results and thereby identifies reasons for excluding those studies from the review process. Therefore, to avoid these types of biases, it is crucial in systematic reviews that researchers state explicitly their criteria for including and excluding studies.

Nevertheless, even though a systematic review clearly states its methodology for searching and evaluating research, some argue that the process still requires various judgements (Munro, 2014). As discussed below, judgements can in fact be necessary in all three phases of the review, which for simplification purposes have been entitled *research question*, *selection*, and *synthesis*.

Research question: Depending on the research question, different types of systematic reviews can be used. For instance, one can review literature *qualitatively* to synthesize research on a topic or one can *quantitatively* analyze the results from single studies and bring them together in so-called meta-analyses to determine the effect-sizes of different interventions. Gough et al. (2012) stress the importance of clearly formulating a research question, as the question serves as the guiding force for our actions, and all subsequent decisions and ponderations are dependent on the necessity of answering this question. Furthermore, judgements are made not only about the methodological quality of the studies reviewed (e.g., are they reliable?), but also about how relevant they are for answering the research question (Gough et al., 2012).

Selection: According to Pae (2015), the primary benefit of systematic reviews is that their findings are derived from systematic and thorough literature searches from *all resources that are available*, which minimizes selection bias. However, Brunton et al. (2017) remark that in practice, it is not possible to know for sure how many studies have addressed the topic one seeks to investigate; thus, it is not possible to be certain that *all* possible studies have been included in the review. In addition, identifying potential studies for a review, requires the definition

of *search terms*, the *sources* to be searched (e.g., electronically searching digital databases or manually searching journal publications), and the *scope* of the search, which also requires judgements. For instance, defining the search terms for a systematic review about writing motivation has consequences for what types of studies that will be located through the search. That is, if one looks for terms like ‘intrinsic motivation’, the searches will return studies that probably look at motivation in accordance with SDT (Deci & Ryan, 1985), whereas if one looks for terms like ‘self-efficacy’, studies that look at motivation from a Banduran perspective (Bandura, 1986) are more likely to be located. Therefore, in addition to clearly defining the methodology used to select studies, it is also fundamental to clearly present the conceptual and theoretical framework used to guide one’s understanding of ‘motivation’ to avoid excluding different perspectives on the topic, without explicitly accounting for them.

Synthesis: *synthesis* is defined as “the composition or combination of parts or elements so as to form a whole” (*Definition of SYNTHESIS*, 2023). Thus, as pinpointed by Thomas et al. (2012), synthesizing research is not simply listing the results from various studies, it also entails transforming the data extracted from primary studies to construct an interconnected whole, meaning that the result of a synthesis is more than just the sum of its parts, it requires the generation of new knowledge. However, generating new knowledge from research whose primary data one often does not have full access to can be problematic. Thus, in synthesizing the findings in studies 1 and 2, I constantly had to ponder if I had enough knowledge about the reviewed data and its sources to decide whether it was possible to draw well-informed conclusions from them or not. For instance, in cases where it was not explicit in study 2 whether self-reports were conducted individually or in group, or digitally or in paper form, I preferred to state that this information was not available in the reviewed studies, instead of making assumptions based on probability.

3.2.2.2 Population size and context

As Davies (2000) point out, it is often unclear how much that can be generalized from single studies, as they have constraints related to time, sample, and specific contextual factors that can potentially compromise their relevance and applicability to different contexts. However, as systematic reviews draw upon the findings from various relevant and reliable studies, rather than the results of a single study or a group of unsystematically selected studies, it increases the possibility of making better informed conclusions about the consistencies and variability found in apparently similar studies. In addition, by integrating multiple studies in systematic reviews, the reliability of the results is strengthened, as the population size increases and different contexts are considered (Munro, 2014). That is, if action X shows positive effects on Y at different contexts and with different populations, then the assumption that ‘X works’ is strengthened.

Nevertheless, some have criticized reviews for often including only few studies and ignoring much relevant research. However, as discussed by Gough et al. (2012), drawing such conclusions by only looking at the number of excluded studies in a review might be misleading for at least two reasons. First, many reviews seek to answer narrow research questions and thus narrowly define their boundaries. Second, electronic searches are imprecise, and therefore capture too many irrelevant studies that later need to be screened out by the reviewer. For instance, given that one of the search terms in my review was ‘composition’, the databases returned a significant number of studies related to music and body composition, which were far from related to the topic of investigation of my project. These issues also show the importance of explicitly presenting the eligibility criteria for the review, as these can account for the reasons as to why specific studies have been included or not.

3.2.2.3 Time and effort

Systematic reviews usually require considerably more time and effort than traditional reviews. According to Whitaker (2019), before engaging in a systematic review, it is important to be aware of at least three aspects: *time*, *resources* and the *amount of data*. He estimates that from conception to submission, a systematic review usually takes 12-24 months. In addition, it often requires a team of investigators. Whitaker (2019) recommends a team of 4 members, including a main and a second investigator, a librarian to assist in the literature search and someone competent in statistics (for quantitative reviews). Finally, Whitaker (2019) points out that very recent or highly specific topics often lack a sufficient amount of primary research data from which one can derive meaningful conclusions. Therefore, given the time and effort required to conduct a systematic review, questions that can generate more conclusive results are preferred, which is why I chose a broad topic in this project, which could generate rich data for producing significant findings for both researchers and practitioners.

3.2.3 Summary

EBE is a relatively new approach to educational research, practice, and policymaking, which has gained attention and support, as it promotes better cooperation among educational stakeholders for generating a more cumulative body of knowledge that can be used as evidence for professional decisions. As discussed in this section (3.2), this approach to education has been mainly criticized for being deterministic, that is, it suggests that professional decisions are to be determined by evidence from scientific research. However, as proposed by Whitehurst, knowledge derived from scientific research should be juxtaposed with knowledge derived from professional wisdom, meaning that for EBE to be successful, the integration of both sources of knowledge is thus needed. Furthermore, for educationalists to use evidence to support their professional decisions, it is essential that research findings are made

available, so that educational stakeholders can judge the relevance of the evidence – or lack of it – for their purposes.

Nevertheless, *availability* is not enough, research findings also need to be *accessible* and presented in more feasible ways, such as through systematic reviews. Therefore, as an educational researcher who sees the benefits of EBE in bringing together knowledge from research and practice, and the importance of communicating research findings in more transparent and manageable formats, I decided to conduct a systematic review. In addition, before conducting research in a topic, a researcher needs to get acquainted with the *state of the art* in that field, which is usually done through a literature review. Thus, as a review is needed anyway, why not do it systematically?

3.3 Studies 1 and 2: Systematic literature review

Given that one of the hallmarks of systematic literature reviews is the transparency of its methods, an extensive description of how data was collected and analyzed is provided in articles 1 and 2, but in this section, I provide additional considerations that may not have been addressed sufficiently in the articles.

3.3.1 Data collection

3.3.1.1 The exclusion of gray literature

The choice of inclusion/exclusion of gray literature comes with dilemmas on both sides, but I argue that for the specific context of the systematic review conduct in this project, peer-reviewed studies was a purposeful and scientifically sound solution.

First, following the protocols of recently published reviews (Camacho et al., 2021; Garcia et al., 2019; Kupers et al., 2019; D. M. Miller et al., 2018; Muenks & Miele, 2017; Perry-Hazan, 2021), I chose to only

include peer-reviewed studies, as a quality indicator, given that these studies have been rigorously scrutinized in peer-review processes, which is widely accepted as a methodology that can ensure a certain level of academic credibility (D. M. Miller et al., 2015). Second, this choice was especially important in this review since, like Camacho et al. (2021), we did not use our quality scoring as an exclusion criterion. Meaning that all identified studies that met our inclusion/exclusion criteria were kept in the synthesis. Therefore, setting a threshold of *at least* meeting the rigor of the peer-review process was an important step to ensure the quality of the included studies.

Third, as noted by Garcia et al. (2019) in their choice of only including peer-reviewed studies in their synthesis, many dissertations have been translated into different journal articles, which would lead to duplications that would be difficult to trace. Adding to this complexity, they remarked that the quality and range of the gray literature, including policy briefs, magazines, and education reports, were too extensive to be addressed in a systematic manner.

Finally, a broader inclusion of material would be more essential in an area with a highly limited research base, and/or for meta-analyses, where publication bias is explored quantitatively. The way the RQs are formulated in studies 1 and 2 rather requires a sampling method that leads to *saturation*. Thus, through purposeful sampling of studies from peer-reviewed journals that represent current trends and standards in writing motivation research, I contend that the studies reviewed offer the diversity necessary to explore the research questions addressed in studies 1 and 2. However, I recognize that while this strict criterion serves as a measure of quality, it may introduce bias. Therefore, an explicit remark regarding this choice was provided in the Limitations-section of articles 1 and 2.

3.3.1.2 The exclusion of studies solely focusing on L2 students and students with disabilities

First, it is important to clarify that L2 students and students with disabilities are not absent from the review's sample, since these students are indeed an integral part of K-5 mainstream classrooms, which are the settings investigated in the review. The studies excluded are only those that investigate samples that are *solely* composed of L2 learners and/or students with disabilities.

The reason why such studies were excluded is because they represent different samples with specific needs, and which merit their own reviews. In fact, specific studies and reviews aiming at investigating these particular student populations separately as distinguished samples are not uncommon (e.g., Acquah & Katz, 2020; Guthrie, 2004; Klassen, 2002a; Scammacca et al., 2015; Troia, 2009; Tyler & Chard, 2000).

In addition, a review aiming at investigating these specific populations would require different search protocols than the one elaborated for the review conducted in the present project, as search protocols are developed for *purposive searching* the specific populations and settings intended for the review (Brunton et al., 2017). For instance, when specifically investigating *students with learning disabilities*, Klassen (2002a) includes in his search protocol the specific term "learning disab*". Similarly, in their meta-analysis of *struggling readers*, Scammacca et al. (2015, p. 371) include specific terms associated with learning and reading difficulties like "reading difficult*, learning disab*, LD, mild handi*, mild disab* reading disab*, at-risk, high-risk, reading delay*, learning delay*, struggling reader, dyslex*". The same is true for studies investigating L2 settings, like Acquah and Katz (2020, p. 5) who include specific terms related to language as described in their Boolean search string: ("language acquisition" OR "language learning" OR "second-language" OR "foreign language" OR "language education" OR "language class" OR "L2"). Thus, any attempts at adding these

populations *post hoc* would be diverging from the search protocol developed for the review.

Moreover, like Camacho et al. (2021), who also excludes studies with samples that are *solely* composed of L2 learners and students with special needs, we wanted to examine practices in classrooms that are inclusive and heterogeneous. Therefore, we chose to focus on motivation to write in L1 in mainstream classroom settings, as, for instance, investigations of L2 writing and in settings like summer camps and/or specialized programs for students with learning disabilities would differ strongly from mainstream settings, which would be difficult to account for in a systematic manner. Nevertheless, as mentioned above, although differentiated programs are often offered to L2 students and students with special needs, these students would still spend the majority of their time in mainstream classrooms and are therefore also an integral part of the sample included in the review.

Although this exclusion criterion was necessary to keep the focus on the sample and settings investigated in the review, I acknowledge the importance of investigating factors that affect the motivations to write of different student populations and in diverse settings. Therefore, this limitation was explicitly addressed in article 1, and recommendations for such investigations were proposed in the ‘Future Directions’-section of the article.

3.3.2 *Data analysis*

3.3.2.1 **Clustering factors in study 1**

As described in detail in the ‘Methods’ section of article 1, to code the data about factors influencing students’ writing motivation both *deductive* and *inductive content analyses* (Moser & Korstjens, 2018) were conducted. Data was first coded deductively in a coding matrix according to components of writing motivation (e.g., interest, value,

beliefs) derived from the three main contributions of motivation research within the writing domain (Boscolo & Gelati, 2018; Graham, 2018; Troia et al., 2012), as presented in section 2.2.2. Then, to accommodate data that was coded inductively (e.g., choice, feedback, instructor), additional categories were added. Finally, following the strategies for coding and analyzing qualitative data suggested by Saldaña (2014), related concepts in the coded material were grouped together into clusters, which resulted in the suggested clusters of factors presented in study 1.

To better understand this process of clustering factors together, an important distinction between *motivation constructs* and *motivation factors* is necessary. *Motivation constructs* on the one hand refer to theoretical concepts or models used to understand motivation, like intrinsic motivation, extrinsic motivation, self-determination, interest, or value. These constructs provide frameworks for understanding the underlying mechanisms and processes that drive behavior. On the other hand, *motivation factors* are more specific elements or stimuli that influence an individual's motivation in particular situations or contexts. These could include factors like rewards, punishment, goals, autonomy, social influence, or personal values. In this sense, motivation factors are more concrete and observable than constructs and are often used to analyze or influence motivation in practical settings. In other words, while motivation constructs are theoretical frameworks that help explain motivation, motivation factors are more specific elements that may influence motivation in real-world scenarios. Given the focus of study 1, in which the aim was to identify factors that influence students' writing motivation in classroom contexts, only identifying theoretical motivation constructs was not sufficient, the identification of more concrete factors was necessary.

In this process, as it will be discussed in section 6.3, the *context* in which both motivation constructs (e.g., interest, value) and motivation factors (e.g., choice, feedback) were identified played an essential role for bringing 'concreteness' to these factors, as the contexts provided more

specific information about what the identified concepts entailed. For instance, as explained in section 3.3.3 in article 1, some studies focused on motivation constructs like value, self-concept, and writing enjoyment (e.g., Mata, 2011), while others focused on characteristics of writing tasks, like whether tasks were meaningful (e.g., Merisuo-Storm, 2006). Initially, these concepts seem unrelated, but after analyzing the context in which they were identified, similarities start to emerge. For example, *enjoyment* of writing tasks, as identified in Mata (2011), is indeed related with *characteristics* of writing tasks that are perceived as attractive to students, as discussed in Merisuo-Storm (2006). During the coding of the studies, many other instances of characteristics of writing tasks that are perceived as attractive (or not) by students were identified in other studies, like being *fun*, *engaging*, *interesting*, or *authentic* (e.g., Chen & Liu, 2019; Gallini & Zhang, 1997; Liao et al., 2018; Perry et al., 2003). This group of characteristics of tasks that students perceive as attractive, and which is in line with Boscolo and Gelati's (2018) suggestion of 'task attractiveness' as a motivation factor was thus grouped together in a cluster which I initially called *task attractiveness*, and which was later relabeled as *appeal of the task*, as it will be explained in the next section.

However, even though *appeal of the task* is a more concrete motivation factor that teachers and researchers can focus on when aiming to design motivating writing tasks, than the construct 'enjoyment' would be, it is still not concrete enough, as it does not unveil what this factor entails and what types of tasks that may be attractive. To address this limitation and present this factor as concretely as possible in article 1, I presented the underlying elements that were coded within this cluster (e.g., *enjoyable*, *fun*, *authentic*) with reference to the studies from which they were extracted. In addition, I provided examples of writing tasks or practices that incorporated these elements, like writing birthday cards or documenting real-world science to distant peers, also with reference to the studies which investigated these practices. The same levels of description were employed for each of the other eight clusters of

motivation factors identified in study 1, and these are presented in article 1 in sections 4.4.1 to 4.4.9. This deep level of description (i.e., thick description, as explained in section 6.3) was intentional in order to make clear for readers what was coded within each cluster and to provide them with concrete examples of how these factors were implemented in the included studies. In this way, it is easier for teachers and researchers to judge how the identified factors and practices may be relevant for and/or how they can be implemented in their own contexts.

Although, the description of all nine clusters of factors identified in study 1 follow this pattern, it is important to note that these clusters are linked to different components of writing communities (Graham, 2018). That is, while some motivation factors are linked to characteristics of *tasks*, others are more closely connected to characteristics of *students* and *teachers*, or of the *environments* in which students and teachers participate. For example, although the factor *appeal* is linked to characteristics of writing *tasks*, which would be found in the center of writing communities, *feedback* is more closely related to *students* and *teachers*, which represent a different component of writing communities. Based on Graham's depiction of writing communities (see Figure 1 on page 264 in Graham, 2018), a simplified version depicting a classroom writing community is provided below in Figure 7, showing the different components to which the nine identified clusters of motivation factors are related.

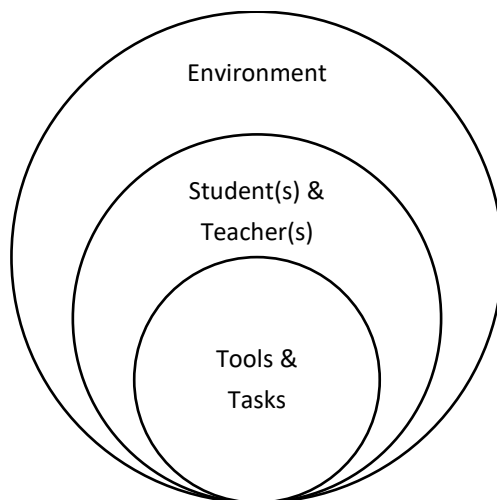


Figure 7 – Simplified version of the basic components of a classroom writing community.

Finally, it is important to notice that although clusters of factors are presented separately, they are most likely intertwined, as it will be discussed in section 5.1.3. This means that different motivation factors influence each other, as remarked by Troia et al. (2012), and that they may also be linked to different components of writing communities. For example, although the factor *difficulty of the task* is primarily related to characteristics of a task, these characteristics are perceived by students differently based on their ability levels, meaning that the influence that this factor plays on motivation is also related to characteristics of the students themselves, which is a different component of writing communities. Moreover, as remarked by Miller and Meece (1999), showing preference towards more challenging writing tasks is also dependent on the *environment* in which these tasks are presented, that is, on how accustomed students are with these types of tasks in their classrooms. Again, to make these connections clear for readers, detailed descriptions of such nuances are provided for each cluster of factors in article 1.

3.3.2.2 The ABC-organization in study 1

When grouping the coded concepts together into clusters, I noticed a pattern, namely that the first letters of some of the identified clusters could be organized alphabetically. At this stage, the preliminary clusters identified were: [B] beliefs (coded deductively based on the work of Graham, 2018; Troia et al., 2012), [C] choice (coded inductively), [E] environment (coded deductively based on the work of Graham, 2018; Troia et al., 2012), [F] facilitation (coded inductively), [F] feedback (coded inductively), [G] goals (coded deductively based on the work of Graham, 2018; Troia et al., 2012), [T] task attractiveness (coded deductively based on the work of Boscolo & Gelati, 2018) and [T] teacher (coded inductively).

As the readers can see, the alphabetical organization was not perfect for four reasons: (1) there was no [A], (2) there was a gap between [C] and [E], (3) [F] was repeated, and (4) [T] did not fit the organization. However, being so close to an alphabetical organization of the factors and knowing that such an organization (as a mnemonic device) would be valuable for communicating findings, and for helping teachers with a useful checklist for lesson planning, I decided to address this issue to see if it was possible to achieve a perfect organization.

First, solving the issue of the missing [A] was easy, as renaming the ‘task attractiveness’ category to *attractiveness of the task* solved the problem, and eliminated one of Ts. Then going back to the concept of task attractiveness proposed by Boscolo and Gelati (2018), the authors argue that adjectives like *authentic*, *interesting* and *challenging* are commonly used to describe how attractive a writing task is. However, although *authenticity* and *interestingness* of writing tasks are more intuitively connected to attractiveness, the *challenge* aspect is more peripheral. In addition, given students’ responses regarding the difficulty of writing, the ‘challenge’ component was moved to a separate category named *difficulty of the task*, which solved the second issue of the missing [D].

Since the ‘challenge’ aspect of the category ‘attractiveness of the task’ was moved to a new category, I then renamed the [A] category to ‘appeal of the task’, to mark this defragmentation from the original ‘task attractiveness’ concept proposed by Boscolo and Gelati (2018).

Next, addressing the repeated Fs, I started looking for synonyms of ‘facilitation’, where the closest suggestions to the concept I was trying to capture in this category, i.e., having available tools, strategies, or people that help in making the writing process easier, were ‘aid’, ‘help’, and ‘support’. Then the choice was easy; [H] help worked well as the next category after [G] goals, and the category ‘facilitation’ was then renamed accordingly.

Finally, I was left with the last category, namely ‘teacher’. Then, looking for synonyms, I came across three suggestions that were very similar to the concept of ‘teacher’. These were ‘educator’, ‘instructor’, and ‘tutor’. Again, the choice was easy, as ‘instructor’ was a perfect fit to complete the ABC-organization of the identified clusters of factors, which resulted in the following categories: [A] Appeal, [B] Beliefs, [C] Choice, [D] Difficulty, [E] Environment, [F] Feedback, [G] Goals, [H] Help, and [I] Instructor.

As remarked by Saldaña (2014, p. 587), “category construction is our best attempt to cluster the most seemingly alike things into the most seemingly appropriate groups,” which involves the three important processes of *reorganization*, *reordering*, and *labeling of categories*. All these three processes were involved in the ABC-organization of the identified clusters of factors in study 1, as transparently described here. In addition, to make clear for readers the details and nuances that these clusters entail and to illustrate the contexts in which they were identified, concrete details and examples of students’ voices (e.g., examples of utterances as expressed by the students themselves and of practices where factors are identified) are provided in article 1.

3.4 Study 3: Quasi-experimental study

Article 3 provides a detailed description of how data was collected and analyzed in study 3, while in this section, I provide some additional considerations related to the methods described in the article.

3.4.1 Data collection

Study 3 was conducted as part of the development of a new national literacy screening test in Norway for first and third graders, as requested by the Norwegian Directorate for Education and Training (NDET). The development of this screening test was administered by the Norwegian Centre for Reading Education and Research, which is the center that houses the present doctoral project, and the design of this test was coordinated by the second author of article 3. The primary goal of this test is to provide teachers with a tool to identify students who need extra support to develop adequate reading and writing skills. In such tests, students' literacy skills are commonly assessed. However, despite an increased focus on motivation in educational contexts, motivation is often not addressed in those same tests. This lack of motivation assessment in literacy screening tests is surprising, since underachieving students – the target group of these tests – often also report lower levels of interest (Graham et al., 1993) and optimistically uncalibrated levels of self-beliefs in their abilities (Klassen, 2002a). To address this gap, items measuring motivation constructs like *interest* and *self-beliefs* were added by the project coordinator to the test.

As noted in section 1.2, surveys with close-ended items for measuring students' writing motivation are not the optimal way of capturing students' unique views since open-ended questions are more suitable for this purpose. However, the primary objective of the national screening test presented here is not to collect students' unique views on writing motivation, but to get an indication of the levels of writing interest and self-beliefs of a large number of students – especially those at risk of

developing literacy difficulties, so that these students' needs can be better met and those who need can get help to strengthen their writing motivation. To generate such an overview of students' levels of motivation, quantitative investigations are commonly conducted, and as explained in section 2.3, surveys with close-ended questions are generally accepted instruments used for this purpose. Therefore, given the format of the national screening test described here and that this test approaches writing motivation from a quantitative standpoint, a survey with close-ended questions is used to get an indication of the students' levels of writing interest and self-beliefs.

In sequence, given the findings disclosed in study 2 regarding the lack of rationales for the choice of pictorial supports in scales assessing students' writing motivation, in collaboration with the project coordinator of the screening test, we decided to investigate this potential influence of scale format on the students' responses. Accordingly, we designed study 3, in which we compare two different types of pictorial support: *faces* and *circles* (see Figure 8).



Figure 8 – Two types of pictorial support: *faces* and *circles*.

The shaping of the scales was developed in cooperation with professional graphic designers, and we aimed to create scales that were quite neutral. For instance, we avoided adding too many colors to the scales, as in scales rendering from happy green to furious red faces. This choice was deliberate, as we wanted to avoid introducing too much *noise* (Kahneman et al., 2021) to the scales, so that differences derived from

the face-scale were primarily based on the face expressions, rather than, for example, color differences.

During the study, all participants responded to seven motivation items (see Table 1 in article 3) on two occasions within a two-week period. The items were included in both test booklets, with the order of the scales reversed, which characterize a *flipped design*, is shown in Figure 9. Specifically, for each of the seven items, approximately half of the students (1146 first graders and 832 third graders) were presented with the face-scale in booklet 1 and the circle-scale in booklet 2, while the remaining half (1051 first graders and 908 third graders) was presented with the opposite sequence. In line with this flipped design, all participating students completed their assigned test booklet 1 (wave 1) before moving on to their assigned test booklet 2 (wave 2).

The reason for using two waves of data collection, where each student responded to all motivation items in both waves 1 and 2, is because it presents a more robust and nuanced approach compared to a single wave analysis. The dual-wave methodology with a flipped design takes advantage of a nested structure within each student, which accounts for the matched pairs structure in the data. This design enhances the analytical power of the study as it allows for *within-subject comparisons*. Specifically, data from each student in wave 1, responding to the motivation items with a particular type of scale, can be directly contrasted with data from the same student in wave 2, where a different type of scale is employed. This paired comparison within individuals increases the precision of the analysis and offers a more comprehensive exploration of the impact of scale types on individual responses.

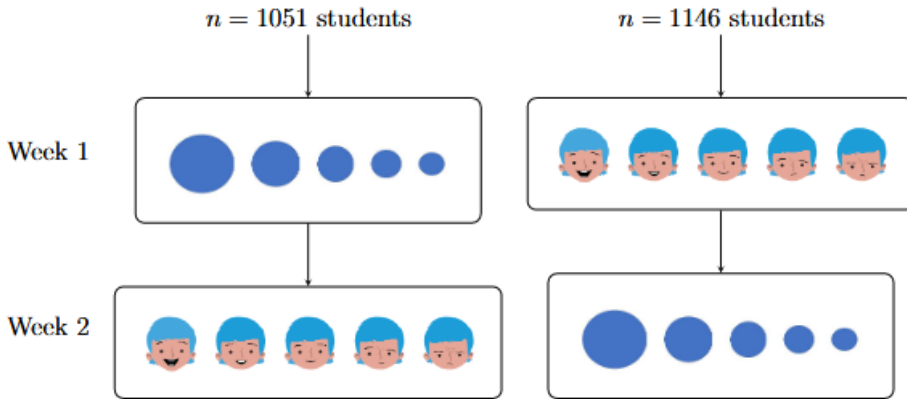


Figure 9 – Flipped design for grade 1.

3.4.2 Data analysis

In accordance with the design described in the previous section, study 3 set out to investigate whether the type of pictorial support used in items measuring students’ writing and spelling interest and self-beliefs affected the students’ responses. Given that NDET requests that the test design allows students to *skip* items to avoid obliging them to respond to tasks they find uncomfortable or do not wish to answer, our investigation also explored whether the type of pictorial support could be correlated with higher or lower levels of skipping. This investigation was guided by the following research question: *Do students report, or skip reporting, levels of writing interest and self-beliefs differently depending on the type of pictorial support used in the Likert scales? If so, to what extent are these differences explained by grade, gender, and type of motivational construct measured?*

To explore this research question, two hypotheses were formulated, and here I present their rationale in more detail. First, considering that students have the option to skip items, it was hypothesized that they may

refrain from answering items they find uncomfortable or lack sufficient motivation to address. Given that Bandura (2006) suggests that face-scales evoke a stronger emotional response from respondents compared to circle-scales, a higher rate of item skipping was expected when face-scales were used, since using this type of scale thus may be more emotionally costly for students.

Second, in line with Bandura's proposition (2006), we suggest that the extreme ends of the face-scale carry greater emotional weight than the more neutral options in the middle. Following this assumption on a greater emotional weight of face-scales compared to circle-scales, higher levels of avoidance of the extreme ends were expected for face-scales.

When comparing the face- and circle-scales presented in Figure 8, it was thus evaluated whether data supports the following two hypotheses:

*H*₁ Students' likelihood of skipping motivation items will be significantly higher for the face-scale compared to the circle-scale.

*H*₂ When not skipping, students will tend to use the scale extremes to a lesser degree with the face-scale compared to the circle-scale.

To analyze these two hypotheses, first descriptive statistics were considered, then statistical modeling and inferential statistics were conducted. Descriptive statistics were primarily presented visually through plot panels combining grade, gender, format, and wave. Subsequently, to conduct formal significance testing, we used mixed linear logistic regression to model item responses, and separate models were conducted for first and third grade. This choice was made because differences on what constitutes writing and spelling interest and self-beliefs for first and third graders are expected, given that the former has just been exposed to their first year of formal literacy instruction, compared to the latter, in which such literacy skills are expectedly more consolidated. Combining both grades in one model in the analysis would introduce cross-grade fixed effects, resulting in a more complex model

with potentially less reliable estimates. To further support this decision of running two separate models for grades 1 and 3, correlation matrices were computed for the 7 items in each grade, to test their equality, but this hypothesis was rejected (chi-square=220.9 with 21 dfs, $p=0$) by Steiger's test (Steiger, 1980).

Finally, the modeling and plotting processes were performed using the R software environment (R Core Team, 2022).

3.5 Ethical considerations

Although researchers should consider ethical issues when conducting systematic literature reviews, Suri (2020) argues that this is often not addressed explicitly by researchers. The author argues that, given that systematic reviews require substantial resources, researchers should analyze the cost-benefit of conducting the review by critically reflecting on the review's purpose and scope, along with its potential advantages for various stakeholders. As mentioned earlier, this was indeed an important consideration and goal of studies 1 and 2, where the massive data generated through the review was explored minutely to bring about knowledge that is relevant for various stakeholders, like researchers, teachers, and teacher educators. In addition, Suri (2020) remarks that the benefit of reviews is increased when reviews propose clear suggestions and recommendations for researchers and practitioners by offering an overview and critically examining current practices, which is true for both study 1 (e.g., a concise table with practice recommendations is provided) and 2 (e.g., an overview of current assessment practices is provided and the lack of students' voices in assessments of motivation is questioned). Finally, bringing students' voices to the foreground in studies 1 and 2 aligns with Suri's (2020, p. 43) invitation for ethical decisions in interpretive systematic reviews that prioritize the authentic representation of experiences and viewpoints from different groups, particularly those whose perspectives are typically underrepresented in the literature, as long as the available published literature permits.

With regards to study 3, one of the most important ethical issues to consider is related to the students' young age. As remarked by Felzmann (2009, p. 106), school research usually does not pose major physical risks, but social and psychological risks are not unusual. For instance, children might feel uncomfortable or isolated from their peers, if they are asked to leave their classrooms or groups to take part in individual research activities, or they might get tired from long interviews and extensive activities. However, during the collection of data for study 3, students were not asked to leave their classrooms. Data was collected as part of the students' completion of a national literacy screening test, which was administered by their teachers in the students' own classrooms. In addition, as explained in article 3, this test has been developed with the ambitious goal of fostering students' literacy interest and self-beliefs even during the assessment situation. For this aim, the length of the test has been reduced (from 60 to 30 minutes), and tasks in this test are designed to reflect high-quality and engaging literacy practices. The development of this test was requested by the Norwegian Directorate for Education and Training (NDET) to identify students at risk of developing reading and writing difficulties and although NDET recommends this test to all schools, participation is voluntary, which is in accordance with guidelines from the National Committee for Research Ethics in the Social Sciences and the Humanities (NESH, 2016).

As mentioned above, assessing the risk-benefit ratio of research before embarking on a project is an important ethical issue to consider. However, this consideration is not only connected to the early stages of research, with regards to the choice of investigation topic or methods for collecting research data. As remarked by NESH (2016, p. 39), it is an evaluation that also needs to be carried out with regards to the dissemination of research findings, so that "research cannot be interpreted tendentiously and misused in political, cultural, social and economic contexts." Therefore, I have strived to be very transparent with regards to how data has been collected and analyzed, so that findings are

open for scrutiny, which may leave less room for tendentious interpretations. Finally, to the best of my knowledge, for all three studies, there are no conflicts of interests, financial or otherwise.

3.6 Study trustworthiness

Despite debates on core quality criteria and standard terminology to address quality in mixed methods research, there is still no consensus among mixed methods scholars (Fàbregues & Molina-Azorín, 2017; Mertens & Hesse-Biber, 2012; Tashakkori & Teddlie, 2008). In their review of quality assessments in mixed methods research, Fàbregues and Molina-Azorín (2017) argue that some scholars have proposed the development of core quality criteria and the introduction of new terms like ‘inference quality’ (Tashakkori & Teddlie, 2008) to distinguish mixed methods from monomethod inquiries. However, the authors remark that other mixed methods scholars (e.g., Cheek, 2015) have argued that such standardized approaches are neither feasible nor desirable, since quality in mixed methods research is highly context-dependent. Consequently, depending on the significance of each of the strands in their mixed methods design, many researchers have rather chosen to use existing language derived from the most representative strand in their mixed design to address quality in their inquiries. In the same manner, given the weight of qualitative research methods for the present project, in which a ‘QUAL → quan’ design is employed, I thus adopt terminology derived from the qualitative strand to assess quality in this project, which is why the term ‘trustworthiness’ has been used to name this current section.

According to Korstjens and Moser (2018, p. 121), *trustworthiness* in qualitative research means simply posing the question “Can the findings be trusted?” To ensure trustworthiness in qualitative inquiries, the authors follow the terminology proposed by Lincoln and Guba (1985), and suggest four criteria – i.e., credibility, transferability, dependability,

and confirmability, and accompanying strategies (e.g., triangulation, member check) to tackle these criteria.

In the present project, I have used *triangulation* strategies to ensure that findings are *credible* (i.e., findings present plausible explanations). For instance, *investigator triangulation* was used in studies 1 and 2, in which an interrater double coded all 56 studies included in the systematic review, and the co-authors of articles 1 and 2 were available throughout the review process to discuss particularities of the included studies and peculiarities of the coding. The same strategy was used in study 3, in which methods for quantitatively analyzing and interpreting the data were discussed and agreed upon with the co-authors of article 3.

Second, to enable readers to assess *transferability* (i.e., the applicability of my findings to their own settings), *thick descriptions* were provided. For example, studies 1 and 2 provide transparent descriptions of inclusion and exclusion criteria. Additionally, article 1 provides detailed descriptions of the contexts and classroom practices in which factors were identified and examples of students' utterances describing what these factors entail, while article 2 provides a rich account of descriptive data regarding the types of self-report identified, which allows readers to judge the transferability of these findings to their own settings. In article 3, detailed descriptions of the context in which data was collected and of the design of the scale types investigated were also provided to facilitate readers' transferability judgements.

Third, to ensure *dependability* (i.e., consistency and stability of findings over time) and *confirmability* (i.e., neutrality, or the degree to which findings could be confirmed by other researchers), Korstjens and Moser (2018) propose the use of an *audit trail*, which include detailed accounts about the research process and decisions concerning it. Such detailed accounts have been provided in the three articles included in this thesis (see appendices 1-3) and in the present dissertation, which allows external auditors, such as peer-reviewers of the included articles,

evaluators of the present thesis, and other readers to assess the transparency of the present project's research path.

4 Results

As described in section 1.1, and reflected in the title of this thesis, the present project focuses on young students' motivation to write by exploring K-5 students' own responses and characteristics of the self-reports used to capture their voices. In this investigation, three studies were conducted, and results are presented in three separate articles (see appendices 1-3). Detailed results can be found in these articles, but in this section, I present a brief overview of results related to the research questions explored in each study.

4.1 Study 1

4.1.1 RQ: *What factors emerge from K-5 students' self-reports as influencing their motivation to write?*

A synthesis of findings from 56 empirical studies conducted in K-5 classrooms (1996-2020) identified 9 factors presented in article 1 as the *ABCs of Writing Motivation*, as these are organized in accordance with the first nine letters of the alphabet: [A] Appeal, [B] Beliefs, [C] Choice, [D] Difficulty, [E] Environment, [F] Feedback, [G] Goals, [H] Help, and [I] Instructor. These will be presented briefly here, and although they are presented separately, they are likely highly intertwined. For example, providing students with *choices* may be of little value, if the choices offered do not *appeal* to them.

The factor [A] *appeal of writing tasks* was derived from self-reports in more than half of the reviewed studies ($n = 30$), and often encompassed activities described as *fun* and *enjoyable*, (Chen & Liu, 2019; Liao et al., 2018), like play-based activities or apps where students could create interactive characters (Boscolo et al., 2012; Kanala et al., 2013; Sessions et al., 2016). *Authentic* and *meaningful* tasks like documenting real-

world science (Gallini & Zhang, 1997), or writing birthday cards (Perry et al., 2003) also emerged as appealing to the students, in contrast to activities referred to as *boring* or *disliked*, like preparing for mandated writing exams (Tunks, 2010), or summarizing texts (Kholisiyah et al., 2018).

Three types of [B] *beliefs* emerged as influencing the students' writing motivations in almost two thirds of the reviewed studies ($n = 37$). The first was *self-beliefs*, including *self-efficacy beliefs* (e.g. Grenner et al., 2020), and students' *writer identities* (e.g., Snyders, 2014). In sum, these studies indicated that students who have positive self-beliefs regarding their writing identities and abilities are often more motivated to write. The second type of belief referred to the students' *beliefs about writing*, in which seeing the *value* of writing and of the processes that writing requires (e.g., organization, revision), seemed beneficial for the students' motivation (e.g., Kim & Lorschach, 2005; Mata, 2011; Seban & Tavsanli, 2015). The third type of belief included in this category concerned the students' beliefs about why one succeeds or fails, in which positive motivational levels were associated with internal attributions for success, like effort (Truax, 2018).

The factor [C] *choice* emerged in one fourth of the studies ($n = 14$), revealing that students appreciated making choices about: *what* to write, *how* to write, *where* to write, and to *whom*. However, low-achieving students also indicated that choosing topics is hard (Seban & Tavsanli, 2015) and students appreciated getting help from their teachers for making choices when needed (Hall & Axelrod, 2014).

The [D] *difficulty of writing tasks* was referred to in more than a fifth of the reviewed studies ($n = 12$). Although some studies indicated demotivational features related to difficulty (e.g., Ihmeideh, 2015), with some students showing preference towards easier tasks (e.g., Boyacı & Güner, 2018), others indicated that challenging writing tasks were more motivating than tasks that were too easy (e.g., S. D. Miller & Meece,

1999). However, showing preference towards more challenging tasks seemed connected to the classroom environments the students belonged to, which leads us to the next factor identified in this review, namely *environment*.

The factor [E] *environment* emerged as influencing the students' motivation to write in approximately 40% of the studies ($n = 21$), in which three types of environments were identified. First, students indicated appreciating *physical environments* that were *quiet*, while *noise* and *distractions* were often referred to as demotivating (e.g., Hall & Axelrod, 2014; Paquette, 2008). Second, pleasant and supportive *social environments* that promote interaction, communication and collaboration were referred to as positive environments, while environments characterized by criticism and judgement were described by students as “uncomfortable writing environments” (Hall & Axelrod, 2014, p. 18). The third type of environment identified concerns the students' *psychological environment*, in which students indicated, for instance, that “writing is more enjoyable for them when they feel calm and relaxed,” while writing was experienced as “less enjoyable when they are in a bad mood, tired, or frustrated” (Zumbrunn et al., 2019, p. 10).

The type of [F] *feedback* that the students received from both their teachers and peers also seem to affect their writing motivation, and this factor was identified in a fifth of the studies. Often students indicated showing preference towards *positive responses* (e.g., Hall & Axelrod, 2014; Zumbrunn et al., 2019), while *negative feedback* that primarily focused on *mistakes* were referred to as demotivational (e.g., Truax, 2018). However, feedback does not have to be inherently positive for students to appreciate it, as students indicated that constructive and concrete feedback that they perceived as helping them grow as writers were also appreciated (e.g., Perry et al., 2003).

In more than a fourth of the studies ($n = 15$), the types of student writing [G] *goals* also emerged as influencing their motivations to write. Characteristics of both *mastery goals* in which students expressed a desire for personal mastery (e.g., Perry, 1998), and *performance goals* in which students expressed goals like getting better grades (e.g., Kim & Lorschach, 2005) or outperforming others (e.g., S. D. Miller & Meece, 1997) were identified in the studies as positive influences for motivation. However, *performance-avoidance goals*, like the desire to avoid failure seemed to play a negative role for the students' motivation (e.g., Perry, 1998).

The factor [H] *help* was identified in more than a third of the studies ($n = 19$). Given the complexity of writing, students indicated showing appreciation of *help from others*, like *teachers* and *peers*, (e.g., Hall & Axelrod, 2014; Lee & Enciso, 2017) but also of helpful *strategies* and *tools* that can facilitate the writing process (e.g., Andrzejczak et al., 2005; Beck & Fetherston, 2003; Chen & Liu, 2019). Although *facilitative affordances of digital tools* (e.g., word-processors that facilitate revision) were mentioned as positive for writing motivation (e.g., Beck & Fetherston, 2003), *technical problems*, like network issues and soft- and hardware malfunction were brought up as frustrating hinders (e.g., Li & Chu, 2018).

Given the crucial mediating role of the teacher in implementing the eight motivational factors described above in classroom practices, we turn to the last factor identified in this review, namely: the [I] *instructor*. Consequently, all studies – either directly or indirectly – lead to the teacher as a decisive element for the students' motivation to write. However, it should be noted that this category also encompasses other stakeholders than classroom teachers, since writing instruction can also be mediated by others like more skilled peers, invited visitors, parents, or researchers, which characterizes a writing community, as described by Graham (2018).

An overview of these nine factors, the motivational focus of each factor, and writing practice examples that may support those foci, is provided in article 1 in table form (see table 3 in article 1), which can be used by educators and researchers as a tool for planning writing lessons or interventions, either as source of ideas or as a checklist. In addition, to make transparent what the 9 identified factors entail, detailed examples of students' utterances and responses, and of the contexts in which these factors were identified are also provided in article 1 (see 'Results and Discussion' section in article 1).

Article 1 also provides a figure (see figure 2 in article 1) showing the overall prevalence of the identified factors among studies. This figure reveals differences on the emergence of factors depending on the type of methodology used in the studies (i.e., qualitative, quantitative, or mixed methods). For instance, although the factor *difficulty* emerges in 23% of the qualitative studies (3 out of 13) and in 32% of the mixed methods studies (8 out of 25), the same factor is present only in 6% of the quantitative ones (1 out of 18), which could be an indication that methodologies that include open-ended questions may give students more room to express their *unique views*, as described by Sturges et al. (Sturges et al., 2002). This issue regarding the methodologies used to capture the students' voices leads us to study 2, in which questions regarding the characteristics of self-reports used for assessing K-5 students' motivation to write are investigated.

4.2 Study 2

4.2.1 RQ1: What types of writing tasks are addressed in self-reports measuring students' motivation to write?

Different motivation constructs, which will be addressed in the next subsection, were assessed in the reviewed studies either towards writing

as a *general task* (e.g., “what makes you a good writer?” from Kim & Lorsbach, 2005) or towards writing operationalized as a *specific task* (e.g., writing blogs, revising paragraphs). Of the different types of writing tasks addressed in the studies, narrative writing/storytelling ($n = 17$) was the genre investigated most often, but genres like expository ($n = 5$), informational/explanatory ($n = 2$), and descriptive writing ($n = 1$) were also explored. However, it was not possible to calculate precise numbers of specific writing tasks, since authors use various terminology and categorizations to describe the types of writing investigated in their studies, with some providing more precise terminology like *letter writing* (Chohan, 2011), and others more general terms like *various writing activities* (Paquette, 2008).

Writing tasks and surveys were executed both on paper ($n = 12$), digitally ($n = 6$), or both ($n = 6$), but many studies did not explicitly indicate the type of technology used.

Different types of writing tasks were also explored in interventions in which students’ levels of motivation were measured before and after the intervention, including for instance writing workshops (Hertz & Heydenberk, 1997; Pollington et al., 2001), and the use of different digital tools (Kanala et al., 2013; Nair et al., 2013; Sessions et al., 2016).

Finally, in addition to assessing the students’ motivations to write, some studies ($n = 33$) also included measures for assessing students’ writing skills, which investigated aspects of writing like ideation (Schrodt et al., 2019), length of composition (Liao et al., 2018), and spelling accuracy (Jones et al., 2016).

4.2.2 RQ2: What motivation constructs are measured and how are they operationalized?

Following the terms used by the authors in each study, 32 motivational constructs were identified. These were organized according to Troia et

al.'s (2012) *theorized components* of motivation and *associated constructs* (as outlined in subsection 2.2.2 of this thesis) and presented in table format in article 2 (see table 4 in article 2). *Attitude* was the associated construct investigated most often ($n = 20$), followed by *self-efficacy* ($n = 11$), and *motivation* ($n = 9$). Corresponding totals were also found for the theorized component related to these associated constructs in which *interest and value* were investigated most often ($n = 39$), followed by *self-beliefs* ($n = 28$), and *motivation* ($n = 11$).

Self-reports investigating the identified motivation constructs were used 83 times in the 56 reviewed studies, and although authors used different terminology to describe the types of self-reports used, these can be divided into three main categories: (1) interviews ($n = 32$), (2) surveys and questionnaires ($n = 46$), and (3) alternative written responses ($n = 5$), like students' feedback on exit slips (Truax, 2018), or students' drawings and written explanations of a recent writing experience (Zumbrunn et al., 2017). Self-reports were administered both in groups ($n = 37$), and individually ($n = 23$), but in many cases this distinction was not specified ($n = 23$). Surveys and questionnaires were predominantly administered in whole-class groups ($n = 31$), while interviews were mostly performed individually ($n = 20$) or in focus-groups ($n = 2$).

Out of the 83 self-reports identified, almost half were derived from previously used motivational measures ($n = 41$), either without modifications ($n = 10$) or with adaptations ($n = 31$), while new measures were developed for the other half ($n = 42$). Previously used items were often used for surveys and questionnaires ($n = 34$), either with modifications ($n = 24$) or without ($n = 10$), while only less than 15% of the studies in which interviews were conducted ($n = 7$) indicated that these were derived from previously used measures. Researchers did not indicate whether alternative written responses were adaptations of previously used motivational measures. In addition, in some cases where previously used measures were adapted, researchers did not indicate explicitly what these modifications entail.

Given the young age of the students, different types of visual support were used in 19 of the studies to aid the students with providing their responses. These included the use of stuffed animals (Mata, 2011; Nolen, 2007; Schrodt et al., 2019), as well as pictorial supports, like geometrical figures ranging from smaller to larger (Hier & Mahony, 2018), or faces ranging from sad to happy (Paquette et al., 2013) indicating different responses on Likert-scales. However, the rationale for the choice of these types of visual support were often not explicitly addressed in the studies.

4.2.3 RQ3: *What emphasis is given to students' voices in the studies?*

Although all the 56 studies reviewed included at least one type of student self-report, the degree to which students' voices were emphasized by researchers when reporting findings varied greatly among the studies. This difference was largely due to the type of data collected in the studies. That is, in studies that relied on students' self-reports as their main source of data (e.g., Hall & Axelrod, 2014; Seban & Tavsanli, 2015), students' voices were given a predominant role. However, in studies combining students' self-reports with reports from researchers (e.g., Hall et al., 2017), teachers (e.g., Li & Chu, 2018), and parents (e.g., Teague et al., 2010), students' voices were juxtaposed with the perspectives of others. Moreover, when data derived from student-reported data diverged from teacher-reported data, students' voices were sometimes given less weight in the studies' findings (e.g., Chohan, 2011).

In studies deriving their findings from responses given from different sources (e.g., students, teachers, and parents), it was sometimes difficult to ascertain whose voices were being portrayed as results from each source were not presented separately (e.g., Hertz & Heydenberk, 1997). Moreover, researchers sometimes only provided examples of the items asked to students, but not of the students' responses. For example, in Lee and Enciso (2017), the authors mention that the open questions "What is

good writing?” and “What does a good writer do?” were also part of their survey, but examples of student responses are not presented in the study.

In addition to comparing the students’ viewpoints with that of others in their investigations, researchers also combined different types of students’ self-reports to capture different aspects of the students’ perspectives. For example, by combining quantitative data from students’ surveys with qualitative data from interviews (e.g., Truax, 2018). Combining these methods was often used for complementary purposes, where quantitative methods (e.g., Likert-scales), were used to measure *frequencies* and *levels* (e.g., *how often* students performed writing activities and *how much* they valued or enjoyed these activities), while qualitative methods (e.g., open-ended questions in surveys or interviews) were used to explore *reasons* for the students’ choices and *factors* influencing those choices. However, even though some studies only explore frequencies and levels of students’ writing motivation, some authors often also discuss reasons and factors influencing the students’ answers, without explicitly asking these questions to students.

These findings from study 2 raised questions regarding lack of rigor in the instrumentation used by researchers to capture K-5 students’ voices regarding their motivations to write, which leads us to study 3, where the use of pictorial support in Likert-scales measuring students’ writing interest and self-beliefs is investigated.

4.3 Study 3

4.3.1 RQ: Do students report, or skip reporting, levels of writing interest and self-beliefs differently depending on the type of pictorial support used in the Likert scales? If so, to what extent are the differences explained by grade, gender, and type of motivational construct measured?

Results from study 3 will be presented in relation to the two hypotheses introduced in section 3.4.2, and which are repeated here for facilitative purposes:

H_1 Students' likelihood of skipping motivation items will be significantly higher for the face-scale compared to the circle-scale.

H_2 When not skipping, students will tend to use the scale extremes to a lesser degree with the face-scale compared to the circle-scale.

4.3.1.1 Hypothesis 1

Statistically significant higher levels of skipping for face-scales were only found for first graders (for detailed statistics, see tables 5 and 6, and figure 5, in article 3). These results indicate that hypothesis 1 is confirmed for students in first grade, but not for third graders. Although similar skipping patterns were found for third graders, these were not statistically significant, which suggests that the face-scale format has less influence on the skipping behavior of third graders compared to their younger peers.

The skipping behavior of third graders seems, in turn, to be more related to the type of motivational construct measured. That is, although both first and third graders skip *task-specific* motivational items (e.g., spelling self-efficacy) less often compared to *general* items (e.g., writing self-

concept), statistically significant differences are only found for third graders.

4.3.1.2 Hypothesis 2

When face-scales were used, the likelihood of endorsing *maximum* values decreased significantly only for boys in third grade (for detailed statistics, see tables 7 and 8, and figure 6, in article 3). These results indicate that hypothesis 2 is confirmed with regards to maximum values only for boys in third grade. However, with regards to the other end of the scale, results show that the likelihood of endorsing *minimum* values decreased significantly for students in both grades (for detailed statistics, see tables 9 and 10, and figure 7, in article 3). These results indicate that hypothesis 2 is confirmed with regards to minimum values for both grades. In addition, statistically significant gender differences were found in both grades, indicating that girls are likely to avoid endorsing minimum values more often compared to boys.

Finally, statistically significant differences were found in both grades with relation to the type of motivational measure used. That is, students are more likely to endorse maximum values for items measuring task-specific self-beliefs compared to interest, while being less likely to endorse minimum values for the same items (see zigzag pattern in Figures 6 and 7 in article 3). These results suggest that students may find it less difficult to choose the negative end of scales when rating their *interest towards tasks*, but may find it more challenging to use the negative end of scales when rating *beliefs about their own competence*.

5 Discussion

In this chapter, findings regarding K-5 students' own responses on their motivations to write, and considerations about the design of self-reports for capturing young students' voices will be discussed. Implications of these findings for classroom practices will also be addressed.

5.1 Findings regarding the students' own responses on their motivations to write

5.1.1 Students' self-reports are a rich source of data

In study 1 of the present thesis, nine factors influencing K-5 students' writing motivation were identified: [A] Appeal, [B] Beliefs, [C] Choice, [D] Difficulty, [E] Environment, [F] Feedback, [G] Goals, [H] Help, and [I] Instructor. The goal of this investigation was to foreground the students' voices by identifying factors emerging from the students' own responses. In addition, the present project aimed at getting a picture of the needs of a specific age group, namely K-5 students.

Based on the massive data generated from the different ways through which researchers in the reviewed studies have investigated the students' perspectives on their writing motivations, the present thesis highlights the valuable insights we can gain from students' self-reports. By tapping into the students' own opinions and experiences, we can gather an abundant amount of data that can guide our quest in finding better ways to effectively address students' complex needs and potential maladaptive beliefs, and to promote more engaging writing practices. This is especially crucial for students in the K-5 age group, as their beliefs are less firmly established and more open to change compared to older students (Bandura, 1997; Unrau et al., 2018).

When we consider this finding in line of theory, it aligns with social cognitive theory, and the primordial role that *beliefs* play in this theory. To tap into such internal states, and get a gist of students' motivations to write, observation alone is not enough; self-reported data is also needed, and it is proven to be a very rich source of data.

However, although self-reported data can provide a glimpse into students' internal states, as discussed in section 1.2, some types of self-report, like close-ended surveys, may give students more limited opportunities for expressing their unique views. In fact, as remarked by Nolen (2007), much of the existing research on motivation takes an *epic* approach, in which children's perspectives are often studied within the framework delimitations of established motivational theories, either through experimental manipulations of variables or through surveys designed in accordance with specific theory-derived constructs. Although these contributions are necessary for getting a more detailed picture of motivation phenomena, Nolen (2007, p. 24) argues that by adopting an *emic* perspective that seeks to "capture the kid's eye view" through interviews with open-ended and broadly formulated questions, aspects of motivation not explicitly addressed in current theories could also be revealed and explored.

Therefore, as explained in more detail in section 3.3.2.1, both deductive and inductive methods were necessary for analyzing data in studies 1 and 2, and for designing the survey used in study 3. First, following an epic perspective, deductive analyses were used for sorting motivation factors (study 1) and constructs (study 2), according to motivation components derived from the three writing motivation research presented in section 2.2.2 (i.e., Boscolo & Gelati, 2018; Graham, 2018; Troia et al., 2012). Then, following an emic perspective, inductive analyses were used in study 1 to explore aspects of motivation not directly focused on in these theories (e.g., choice, feedback, instructor). Finally, given the purpose of study 3, in which the influence of different types of visual aids may have on young students' responses is investigated, an epic approach was used

to design a survey in accordance with specific theory-derived constructs, namely interest and self-beliefs.

5.1.2 *The decisive role of the teacher*

In the endeavor of fostering students' writing motivation in educational settings, teachers play a decisive role. That is, although students' beliefs, goals, and values are also influenced by other members of their *writing communities* (Graham, 2018), like parents and peers, in classroom settings, teachers are the authority that ultimately will be responsible for the day-to-day conduction of motivational writing practices. This finding aligns with Boscolo and Gelati's (2018) proposition on the fundamental role of the teacher in fostering writing motivation, as discussed in section 2.4.

In fact, although policymakers and researchers are also responsible for suggesting policies and designing tools and interventions that may support the learning environment, if teachers are not committed or do not engage with these propositions, recommended policies, tools, and interventions may not yield the expected results. For instance, in a schoolwide mailing program intervention, Chohan (2011) found a relation between the teachers' levels commitment and their students' levels of enjoyment of the program. Chohan (2011, p. 47) then argued that "unless all teachers are responsive to new initiatives, some teachers' efforts in implementing new strategies may be neutralized." Similarly, after investigating the effects of writing workshop and traditional writing instruction on fourth and fifth graders' self-perception, Pollington et al. (2001) concluded that children's writer self-perception is more significantly influenced by individual teachers than by specific approaches or strategies.

However, although teachers indeed strive to create supporting and motivating writing communities in their classrooms, some have struggled to achieve this goal (Korth et al., 2017). These findings

emphasize the significance of investing in teachers, whether through high-quality teacher education programs or ongoing professional development opportunities. In addition, to better help teachers with matters that are relevant for them, they should be given more room to take part in the development of policies and in the identification of research questions that concern their working environment.

5.1.3 Motivational factors are intertwined

As remarked in section 2.2.2, Troia et al. (2012) argue that motivation components are interrelated, and the authors invite researchers to experiment with diverse research methodologies to investigate and perhaps unravel some of these interconnections. In study 1, although motivational factors were organized within separate categories (according to Saldaña's (2014) processes for category construction), in line with Troia et al.'s (2012) proposition, I argue that these factors are notably intertwined. Such an assumption is in line with social cognitive theory (Bandura, 1986, 1994; Schunk et al., 2012; Zimmerman, 2000a, 2000b) and with social cognitive views of writing (Graham, 2018; Troia et al., 2012), as students' writing motivations are influenced by both individual characteristics (e.g., *choice*) and social interactions (e.g., *environment*).

During data analysis, some of these links among factors were indeed identified and to communicate them to readers, cross-references linking factors together were included in the 'Results and Discussion'-section of article 1. These are often given in parentheses in the form of notes like 'see also factor *X*.' For instance, under the presentation of the factor *beliefs*, process portfolios were suggested as classroom practice examples based on findings from Nicolaidou (2012), in which students' self-efficacy for essay writing increased over time after using this practice. However, given that the author remarked that these results were mediated by *feedback* received and given, the following note was added: 'see also factor *feedback*.'

Other links identified include, for example, connections between the categories: *difficulty* and *help*, *choice* and *help*, *environment* and *goals*, *environment* and *choice*, *goals* and *beliefs*, and *help* and *beliefs* (for details regarding these connections, see the presentation of factors in the ‘Results and Discussion’-section of article 1).

5.1.4 Age-appropriate implementations of factors

Findings from study 1 indicate that all nine factors identified in the students’ self-reports emerged in all grades, i.e., from kindergarten to fifth grade. However, what these factors entail and how they are materialized in *classroom practices* logically differ, depending on the students’ age. For example, writing tasks that are *appealing* and that have the appropriate level of *difficulty* for kindergarteners, like writing grocery shopping lists in a Dramatic Play Center (DPC) simulating a grocery shop (Ihmeideh, 2015), may not seem as appealing or sufficiently difficult for fifth graders.

However, practice examples from younger students are not by default unsuitable for older students and vice-versa, as practices can be adapted according to age-appropriate preferences and needs. For example, taking the *dramatic play* practice mentioned above, we can see how one practice can be adapted to different levels, so I will go into detail in this specific classroom practice for the purpose of differentiation. Taking this practice into consideration, it can be argued that even though fifth graders may not be so enthusiastic to write grocery shopping lists, DPCs may still be experienced as an engaging learning environment by these students. This is because DPCs seek to simulate *authentic scenarios* and *social contexts* in which writing takes place – i.e., writing communities (Graham, 2018), and in which students can practice these activities in a more controlled and less stressful environment (Skaftun, 2015). Nevertheless, depending on the age of the students, these scenarios will indeed greatly differ. For instance, while kindergarteners may write grocery shopping lists (e.g., Ihmeideh, 2015), third graders may write letters (e.g., Chohan, 2011),

and fifth graders may write scientific reports to distant peers (e.g., Gallini & Zhang, 1997). That is, the number of scenarios and activities that can be created and implemented with children in all grades is virtually endless. Here, teachers thus have plenty of room for being creative and setting up scenarios and activities that best suit their writing goals and that are age-appropriate for their students.

Based on this example, one can see that classroom practices can be adapted to age-specific needs and preferences, and that this type of differentiation presented in detail here can be used for other practices and contexts. Thus, to help teachers and researchers with finding other practice examples that can facilitate the implementation of the identified motivational factors and that can be adapted for students in K-5 grades, a table compiling the identified factors, their motivational focus, and a list of more than thirty accompanying writing practice examples is presented in article 1 (see Table 3 in article 1).

5.2 Findings regarding student self-reports design

5.2.1 The type of writing addressed matters

Findings from study 2 indicate a clear division between assessments of motivation towards *writing as a general task* (e.g., items investigating how much students like writing) and *writing as specific tasks* (e.g., items investigating students' motivations towards writing a story, spelling a word, revising a text). Notably, in investigations operationalizing writing as a specific task, *storytelling* emerged as the type of task addressed significantly most often in the studies. Similarly, Camacho et al. (2021) also found that narrative text/story was the genre most studied in their review of writing motivation in school. However, although storytelling tasks are common in K-5 grades, this overweight of storytelling tasks in measures of motivation is somewhat peculiar, since writing entails much

more than just this specific task. A question that thus remains unanswered is whether this overuse reflects (a) an overrepresentation of storytelling tasks in K-5 classroom practices, or (b) researchers' oversimplifications of writing in measures of writing motivation.

If the former is the case, more varied types of writing tasks and genres that better represent the myriad of writing activities found in writing communities (Graham, 2018) need to be introduced to K-5 classrooms. In fact, Hall et al. (2017) found that when students are encouraged to explore and create other types of writing, like informational texts, their confidence and interest in this genre increased. According to the authors, teachers have indicated using informational texts less often, because they believe that informational texts are too difficult for young students and that students also prefer narrative texts. However, the authors remark that there is strong evidence suggesting that young children can read and write informational texts, that they benefit from it, and that although some studies have indicated that children prefer narrative texts, a considerable body of research has shown that many children rather prefer informational texts. Thus, presenting young students with more varied writing tasks is beneficial not only to strengthen their competence in other genres, but also to foster their motivation towards a broader variety of writing activities.

If the latter is the case, then this overweight of storytelling tasks in assessments of writing motivation may be indicative of construct-underrepresentation in current investigations in school settings. In future investigations, researchers should therefore address more varied types of writing tasks, reflecting more realistically how multifaceted writing is.

5.2.2 The definition and operationalization of motivation constructs matter

Although theoretically anchoring investigations of writing motivation is crucial, findings from study 2 revealed that more than 30% of the

reviewed studies lack relevant references to motivation theory and/or research to properly *define* and *operationalize* the motivational constructs investigated in their studies. In addition, only a third of the reviewed studies explicitly explained in their methods section the rationale for their design/choice of motivation measure with reference to motivation theories.

Since motivation is a complex and multi-dimensional concept, as shown in section 2.2, this lack of clear definitions of the motivational constructs being investigated and the absence of sufficient rationales for how they are operationalized can have significant implications for the validity of measures, as it becomes challenging to determine which motivational components are being assessed and how. In essence, establishing theoretical clarity is crucial for creating a transparent connection between the motivational construct being investigated and the chosen method for measuring that construct. To enhance consistency and promote standardized definitions of motivational constructs, Camacho et al. (2021, p. 224) have compiled a list of definitions that can assist researchers in bringing clarity to their investigations of motivation. However, although the authors identified 24 constructs in their review, only the definitions of 11 of these constructs are given. Recognizing that researchers often use different nomenclatures to investigate similar constructs, and to support researchers with navigating this multitude of motivation constructs, the 32 motivation constructs identified in study 2 were sorted in a table (see table 4 in article 2) in accordance with the four main components of motivation proposed by Troia et al. (2012), as presented in section 2.2.2.

5.2.3 *Whose voices are being heard matters*

As described in section 4.2.3, studies weighed students' voices differently. Often, in studies relying primarily on student-reported data (e.g., Mata, 2011), students' voices were given more prominence. However, in studies including other data sources, like reports from

teachers (e.g., Chohan, 2011), parents (e.g., Teague et al., 2010), or researchers (e.g., Hall et al., 2017), students' perspectives sometimes received less attention compared to the other sources. Given this difference on the emphasis given to students' voices, when presenting findings regarding students' motivation that are derived from different sources of data, researchers should clearly delineate whose voices are being heard, and the significance of this contribution to their studies' findings. In addition, for clarity purposes, although different data sources can be combined and synthesized in the studies' findings, results derived from each data source should be presented separately, and examples of both the items/questions presented to the students and of their responses should be provided alongside the publication.

5.2.4 The type of pictorial support used matters

Despite the widespread use of various pictorial supports in measures of writing motivation with young students (Hier & Mahony, 2018; Mata, 2011; Nolen, 2007; Paquette et al., 2013; Schrodtt et al., 2019), the present thesis demonstrates that the type of pictorial support used in these measures may influence the students' responses. In fact, as remarked by Kahneman et al. (2021, p. 199), "the choice of a scale can make a large difference in the amount of noise in judgments, because ambiguous scales are noisy." Consequently, researchers should make deliberate choices when selecting appropriate scales for measuring writing motivation with young students. Regarding measures of writing and spelling interest and self-beliefs specifically, findings from study 3 in the present thesis indicate that using circles ranging from bigger to small (and from small to bigger) as pictorial support, instead of face expressions ranging from happy to sad, seem to be a more appropriate choice for measuring these constructs with young students. Findings indicate that circle-scales seemed to be the least noisy for both girls and boys, and specially for first graders who seemed to be more sensitive to face expressions in face-scales, compared to third graders. In sum, these

findings support the recommendations of Bandura (2006), in which circle-scales are preferred over face-scales in measures of self-efficacy, and further extends these recommendations to assessments of writing interest and self-concept. There is reason to assume that these recommendations may also apply to other motivational constructs, and perhaps to older students, but further research is needed to investigate these claims.

5.3 Implications for classroom practice

Previous research has provided recommendations for fostering writing motivation in educational settings (Bruning & Horn, 2000; Camacho et al., 2021; Troia et al., 2012), and the present thesis adds to the literature by providing an overview of factors that influence K-5 students' motivation to write. Unique for the present project is that these factors are presented as a mnemonic device, with the aim to communicate them in a way that would be more easily retained and disseminated, and consequently more likely to be translated into practice by educators. In this section, for an effectively implementation of these factors in classroom practices, I first highlight the importance of *teacher planning*. Then I discuss further implications with regards to two central elements of classroom writing communities (as shown in Figure 7), namely the types of *tasks* and *tools* used in K-5 writing instruction.

5.3.1 Teacher planning

School curricula, national tests, and teaching books often set premises for writing skills and genres that teachers should plan for in different grades, but guidelines for motivation aspects that also deserve attention in the classroom are often not as central in these documents. Such an imbalance may lead to more systematic teacher planning of curricula objectives and writing skills, but less systematicity with regards to motivation. However, evidence from the literature suggests that creating classroom practices that foster young writers' motivation is a complex

endeavor, and that to achieve this goal intentionality is required (Boscolo & Gelati, 2018; Bruning & Horn, 2000). Consequently, the creation and implementation of engaging writing practices requires systematicity not only in the planning of which *writing skills* and genres to practice, but also in the planning of which *motivation factors* to target.

As a first step towards more systematicity in this planning endeavor, the nine factors identified as the *ABCs of Writing Motivation* in the present thesis could serve as key concepts for how to plan for motivation within writing instruction. As mentioned in section 4.1.1, to facilitate this process, the identified factors are presented in table form as a checklist, including descriptions of the motivational focus of each factor and references to practice examples (see Table 3 in article 1). In addition, given that these factors are presented alphabetically as a mnemonic device, memorizing them becomes easier, which means that if teachers want to address these factors at any point, but do not have this checklist within reach, they may still remember the factors by heart.

Findings from the present project indicate that it is also important to assess how successful instructional efforts targeting writing motivation are. In fact, in one of the reviewed studies (Jones et al., 2016), researchers argued that teachers often use rainbow writing (the use of colorful pens to practice spelling) assuming that this artistic element is *appealing* to young students. However, by using a short questionnaire, Jones et al. (2016) found that students actually preferred the less colorful and more challenging activity of taking short quizzes and self-checking their answers. This example shows that it is important to test assumptions regarding students' preferences and needs, and as evidenced in the present thesis, a multitude of ways of capturing students' voices can be used for this purpose. For example, in addition to more traditional ways like surveys, questionnaires, and interviews, alternative instruments for inviting the students to express their views and which may be easier for teachers to implement in their everyday practices, like student-teacher conferences, written reflections, and even drawings, can also be used.

However, it is important to acknowledge that motivation assessments have the potential to both support and hinder students' writing identities (Walgermo & Uppstad, 2023), since experiences of mastery or failure in assessment situations contribute to the shaping of students' beliefs about writing and about themselves (Troia et al., 2012). Thus, mindful use is recommended.

5.3.2 Tasks

Writing is a complex skill that encompasses various forms, purposes, and audiences, as discussed in section 2.1. However, in educational settings, writing is often teacher-initiated and traditionally presented through a limited set of tasks (e.g., writing stories, practicing handwriting, spelling words) that do not mirror the myriad of authentic writing activities encountered in everyday life (e.g., online chatting, composing emails, or writing lists and invitations) and that characterize the different types, purposes, and conventions of writing communities (Graham, 2018). These limited school writing experiences can negatively influence the students' beliefs about themselves as writers and about the value of writing. Thus, to make writing more meaningful, teachers must provide students with varied opportunities to use writing for different purposes beyond typical school assignments, and present explicit connections between *school writing tasks* and *real-life writing situations*.

5.3.3 Tools

In addition to a myriad of tasks, authentic writing situations also include a multitude of tools, like pens, chalks, keyboards, markers, and cut-out letters. However, the range of writing tools typically found in traditional classrooms often do not reflect this vast spectrum of possibilities. Furthermore, Graham and Harris (2016) argue that despite technological advancements, many traditional elementary classrooms still employ pen and paper as their primary writing tools. The authors then pledge school administrators to update their classrooms with 21st-century writing tools.

Also Camacho et al. (2021) advocate for the utilization of and research on digital writing tools in classrooms, arguing that these tools have a promising effect on students' motivation to write.

However, findings from the present thesis suggest that the digitality quality of writing tools is not a motivational factor in itself, but rather affordances and features of these tools. For instance, in study 1, students reported positive levels of motivation when they could see the characters in their stories come to life both through digital platforms (Sessions et al., 2016), and non-digitally, when teaching artists invited to school performed their stories (Lee & Enciso, 2017). That is, rather than the digitality quality of the tool, the *appealing factor* of seeing characters come to life seems more influential for the students' motivation levels. Similarly, although *facilitative affordances* of digital writing tools, like spell-checkers and editing possibilities, influence positively students' writing motivation (e.g., Beck & Fetherston, 2003), *technical glitches*, like network issues and soft- and hardware malfunction, in turn, affect students' motivation negatively (e.g., Li & Chu, 2018). Hence, it is crucial for teachers to be purposeful in their choice of tools for writing instruction. Finally, as previously highlighted, to ensure the effectiveness of these selections in their efforts to create engaging writing practices, teachers can check with their students what their beliefs and preferences regarding these tools are.

6 Philosophical reflections

In this section, I address three important philosophical aspects of the present thesis that did not have space to be addressed in necessary length in the articles. First, according to Tønnessen (2011, p. 149), empirical research on writing often offer brief descriptions of skill and motivation, but these are rarely reflected upon through more precise and in-depth reflections or definitions, however, to move the field further, more fundamental considerations regarding these constructs are needed. Addressing Tønnessen's remarks, in section 6.1, I present some more elaborate philosophical considerations regarding these constructs, which are central for the present work. Based on Aristotle's propositions regarding actuality and potentiality, I will discuss ontological perspectives on the nature of writing skill and motivation as *potentialities*, and writing performance as *actuality*.

Second, although systematic reviews and meta-analyses are often used for identifying *fixed* rankings of motivation factors based on the prominence of these factors in the analyzed data, or differences in effect sizes, I will argue in section 6.2 that such rankings are rather *flexible*. Consequently, optimal conditions for fostering motivation in educational settings may vary depending on the *prioritization* of motivational factors and the *context* in which they are investigated. To explore this point, I will draw a parallel to Optimality Theory from theoretical linguistics, in which rankings of constraints to generate linguistic outputs are not fixed but vary according to – namely *prioritization* and *context*.

Finally, in this endeavor of unraveling potential different rankings of motivation factors depending on the contexts in which motivation phenomena are studied, I will argue in section 6.3 for the importance of providing *thick descriptions* of these contexts.

Given the foundational nature of the reflections presented in this section, they show themselves relevant also for other fields than writing, like reading, motivation, and education research.

6.1 *Actuality and potentiality: From potential skills to actual performances*

Following Aristotle's propositions regarding actuality and potentiality (Cohen & Reeve, 2021), some theoretical accounts addressing skill and motivation explain the nature of these concepts in terms of *potentiality* (Tønnessen, 2011; Walgermo, 2018), and refer to both skill and motivation as potentials in nature. In Aristotelian philosophy, potentials (also known as potentialities), refer to the inherent capacities or abilities that entities or substances possess to undergo change or develop into certain states or actualities (Cohen & Reeve, 2021). That is, *potentiality* represents what an entity can become or the possibilities inherent in its nature, while *actuality* refers to a potential that has been realized. For instance, a seed has the potential to grow into a tree because it possesses the inherent capacity or potentiality to develop and actualize its nature as a tree. The potentiality of the seed is fulfilled through a process of *change*, where it undergoes growth, drawing nourishment from the soil and sunlight, until it reaches its actualized form as a fully grown tree.

Following this distinction between potentiality and actuality, Tønnessen (2011) remarks the importance of distinguishing between *skill* and *performance*, where the former is seen as a potential, and the latter as the actualization of the former. Further, the author argues that "even though a skill changes over time, it is more stable than the actualizations of it," since actualizations are highly influenced by other factors (potentials), such as motivation and attention (Tønnessen, 2011, p. 155). Applying this thinking to writing motivation, and in line with social cognitive theory, one can then argue that *actual writing performances* are realizations of *potential writing skills*, and that these realizations are influenced by both individual and environmental factors, like cognitive

abilities and social interactions. In addition, given that potentials, like skills, can be influenced by other potentials, like motivation, and vice-versa, Walgermo (2018) argues that we may thus conceptualize the dynamics among them in biological terms, likening them to organisms found in symbiotic and dynamic relationships, in which mutual benefits are often derived from such interactions. In fact, different studies have found such mutual benefit in which the development of writing skills and increase in performance influence positively the development of motivation and vice-versa (Graham, Daley, et al., 2018; Graham et al., 2017; Hall et al., 2017; Pajares et al., 2001; Wijekumar et al., 2019).

However, although actualities and potentialities are often dichotomously divided, in which actualizations are seen as realizations of potentials, such realizations do not preclude the existence of potentialities, as both actuality and potentiality are *ways of being* (Witt, 2018). That is, once potentialities are realized, they do not consequently cease to be. In fact, Walgermo (2018, p. 67) describes potentialities as “dynamic entities that are never fully evolved; they always have the possibility for improvement or degeneration.” To illustrate this point, going back to our seed (potentiality) and tree (actuality) example, even if an actual tree is seen as the realization of a seed’s growth potentiality, potentials for growth or deterioration are still at play and do not cease to be.

Such a dynamic understanding of potentiality has important implications for educational settings. That is, skills and motivations, being understood as potentialities, are dynamic and prone to change. Thus, students’ skills and motivations should not be seen as static entities, but as dynamic potentials that according to Walgermo (2018) can be nurtured and actualized for any student at any stage. However, given that potentials do not self-actualize, meaning that they need to be realized (i.e., made actual) by some outside force (Cohen & Reeve, 2021), it is important to remember that although skills and motivation may influence each other’s development in symbiotic terms (Walgermo, 2018), they do not self-

actualize. Rather, skills need to be learned and practiced, and motivation need to be fostered.

This remark has at least two implications for educational contexts. First, in the same way that a seed does not simply self-develop into a robust tree, rather its growth is dependent on factors like being placed in fertile soil, and having enough sunlight, water, and nourishment (i.e., the realization of its full potential is dependent on optimal conditions), so do students need to forge their skills and motivations under optimal conditions to develop their full potential as writers. Second, in the same way that a gardener needs to know about what factors that influence a seed's growth in order to create optimal conditions that support this process, so do educational stakeholders need to know about what factors that influence students' growth into successful writers to create the necessary optimal conditions to help them.

In the present thesis, some of the factors that may contribute to the creation of optimal conditions for fostering writing motivation and consequently writing performance have been identified according to the students' own responses. However, in the next section, relying on Optimality Theory from the field of theoretical linguistics, I will argue that although these factors may be the same across different contexts, optimal conditions will be realized differently depending on the context.

6.2 *Optimal conditions vary: A thought experiment based on Optimality Theory*

In his comprehensive article introducing Optimality Theory (OT), Archangeli (1999) explains that in his presentation of OT, he seeks to help readers with understanding more complex work in the model, and possibly even inspire them to venture into experimenting with the optimality paradigm in other areas beyond the field of theoretical linguistics. In this section, this is exactly what I aim to do, that is, to experiment with OT in a field outside of theoretical linguistics. To

achieve this goal, I will first give a brief presentation of OT. Then – as a thought experiment – I will argue that a similar way of thinking, in which different constraint rankings lead to different optimal outputs, can be a fruitful way of looking at how different rankings of motivational factors can lead to different realizations of optimal conditions.

Relying on OT in the present thesis may come as a surprise to the reader, since OT stems from *generative* perspectives on language acquisition (Chomsky, 1959), which generally falls into formal, rather than functional linguistics. To distinguish these two, Van Valin Jr. (2003) argues that *formal linguistics* is characterized by *theory-internal criteria* for explaining linguistic phenomena with regards to rules, constructs, or principles that entail aspects derived solely from that domain. That is, syntactic phenomena are thus explained in terms of syntax, phonological phenomena are explained in terms of phonology, and so on. *Functional linguistics*, on the other hand, is characterized by *external explanatory criteria*, involving factors external to the studied domain, which can be internally or externally related to language itself (Van Valin Jr, 2003). To exemplify this point, Van Vallin Jr. (2003) argues that using phonetics to explain phonological phenomena illustrates the use of *language-internal*, but *domain-external* criteria in the explanation. Conversely, employing aspects of the human perceptual system to explain phonological phenomena illustrates the application of *language-external*, and *domain external* criteria to explain linguistic phenomena. Following this distinction on the use of external factors for explaining linguistic phenomena, even though OT is often portrayed as a product of formal understandings of linguistics, as remarked by other researchers (Boersma, 1997; Nakamura, 1999; Uppstad, 2005), OT can also be seen as a functionalist approach for explaining linguistic phenomena. Thus, although OT will be presented here as indeed stemming from formalist views on language, its relevance in the present thought experiment is in accordance with OTs potential for also providing functional explanations.

6.2.1 *A brief overview of Optimality Theory*

Although languages worldwide differ significantly on how they are realized, the Chomskyan view that humans are genetically predisposed to learn language suggests that in an abstract level, languages are indeed fundamentally similar (Chomsky, 1959), since they stem from the same set of *language universals*, which are underlying characteristics or patterns found in all human languages across different cultures and regions. Based on this assumption, Archangeli (1999) explains that to understand variations among different languages, linguists then try to systematize these differences by developing a formal schema of human linguistic capabilities that delineates the elemental similarities while accommodating the numerous variations inherent in natural language.

Such a model was introduced by Prince and Smolensky in the field of phonology (i.e., the study of sounds) in the beginning of the 1990s, and soon became a groundbreaking alternative model for explaining the organization of sound systems in human languages (Archangeli, 1999). This model is called Optimality Theory and is described as a linguistic framework that explains language patterns as the result of conflicting constraints, where the optimal output is determined by a hierarchy of constraints that prioritize certain linguistic factors over others. As remarked by Archangeli (1999), the assumption that universal constraints are violable is central for OT. That is, while *universal constraints* offer a way to describe cross-linguistic similarities, variations among languages are expressed through unique degrees of *constraint violation*. To permit such violations and to determine which constraints that are violable in different languages, OT proposes *constraint ranking*, in which violating constraints that are ranked lower is permitted to fulfill constraints that are ranked higher (Archangeli, 1999).

To demonstrate how OT works, I will provide a simplified example comparing how different phonological constraint rankings lead to

different pronunciation outputs. For example, the word ‘car’ is usually pronounced as /kar/ in American English (AE), but as /ka/ in British English (BE). That is, while the consonant sound at the end of the syllable, also known as *coda*, is kept in AE, this sound is deleted in BE. To explain this difference, OT uses the same universal constraints for both languages, but with different rankings, which will result in different outputs for each language. This process is usually illustrated in an OT tableau that is structured as shown in Table 1, in which (a) constraints are presented hierarchically from left to right, (b) asterisks denote violations, (c) exclamation marks denote fatal violations (i.e., violation that is worse than another candidate on the highest ranked constraint distinguishing them), and (d) a pointing finger indicates the optimal output (i.e., the output that is most *loyal* to the proposed constraint ranking).

Table 1 – OT tableau structure

	Input	Constraint 1	Constraint 2
a. ☞	Candidate A		*
b.	Candidate B	*!	

Note. This table illustrates that candidate A is better than B, since A violates a lower-ranked constraint, while B violates a higher-ranked constraint.

Applying OT to explain the different pronunciations of ‘car’ in AE and BE, universal constraints like NOCODA (i.e., syllables must have no codas), and MAX-IO (maximize input-output correspondence) can be used, but with different rankings², as shown in Table 2 and Table 3. That is, in AE, the MAX-IO constraint is ranked highest, prioritizing input-output correspondence. Consequently, the coda sound /r/ is kept to satisfy this higher-ranked constraint, as violating NOCODA (the lower-ranked constraint) is more acceptable than violating the higher-ranked constraint MAX-IO. For BE, the situation is the opposite, as the deletion

² Note that these rankings are simplified and do not encompass the full complexity of constraint interactions in OT analyses of AE and BE.

of /r/ in ‘car’ is allowed to satisfy the higher-ranked constraint NoCODA, rather than the lower-ranked constraint MAX-IO.

Table 2 – OT ranking for American English

	/kar/	MAX-IO	NoCODA
a.	/kar/		*
b.	/ka/	*!	

Table 3 – OT ranking for British English

	/kar/	NoCODA	MAX-IO
a.	/kar/	*!	
b.	/ka/		*

6.2.2 *Experimenting with motivation factors as OT constraints*

Now that OT has been briefly introduced, how can this theory be relevant for the present thesis? In essence, the way that OT systematizes linguistic constraints within various rankings to explain differences across languages can provide structural guidelines for how motivational factors could be systematized across different educational contexts. For example, similarly to how linguists attempt to uncover universal constraints in OT, motivation researchers may uncover (universal) motivational factors. In addition, as explained above, linguists try to systematize universal constraints within rankings that prioritize certain linguistic factors over others to explain how these hierarchies transform inputs into distinct optimal outputs in different languages. Similarly, potential universal factors could be systematized within different hierarchies to explain how optimal conditions for achieving writing objectives will vary across educational contexts based on how motivational factors are prioritized in these contexts. To exemplify this point of view, the structure of a hypothetical OT tableau for explaining how different rankings of motivational factors may result in distinct

optimal writing practices (i.e., practices that are most ‘loyal’ to the proposed factor ranking) could look like the one suggested in Table 4.

Table 4 – Hypothetical OT tableaux structure for optimal writing practices

Writing objective	Factor 1	Factor 2
a. ✎ Writing practice A		*
b. Writing practice B	*!	

To illustrate how this structure could be used with the motivational factors identified in the present thesis (see article 1 for more details about these factors), a comparison between two different rankings is provided in Table 5 and Table 6. In both examples, the writing objective is the same (i.e., practice spelling), but the educational context in Table 5 (context 1) is a group of kindergarteners practicing the spelling of fruit names, while the context in Table 6 (context 2) is a group of kindergarteners practicing words from a list for a spelling contest. On the one hand, for students in context 1, the appeal of the task may be ranked higher than for example being allowed to choose which words to spell or achieving the goal of spelling these words correctly, since invented spellings are not uncommon at this age. In this context, a grocery shop DPC seems better than oral dictate, because even though it violates the *choice* factor, practice b violates a higher-ranked constraint, namely *appeal*. For students in context 2, on the other hand, the *goal* of being well prepared for the upcoming spelling contest and being able to choose which words to practice (e.g., the words they struggle the most with) may be ranked higher than the appeal of the task. This difference in the ranking of the factors may help explaining why in this context a seemingly unappealing task, like oral dictate, may still be experienced as more motivational for these students than a seemingly more appealing alternative.

Table 5 – Hypothetical OT ranking for kindergarteners practicing spelling

	Practice spelling	Appeal	Goal	Choice
a.	Dramatic Play Center			*
b.	Oral dictate	*!		

Table 6 – Hypothetical OT ranking for kindergarteners practicing spelling for a spelling contest

	Practice spelling	Goal	Choice	Appeal
a.	Dramatic Play Center	*!	*	
b.	Oral dictate			*

In sum, although the same motivational factors may be at play in various contexts, different rankings of these factors allow variances among educational contexts – and even among students – on how optimal conditions are realized.

Such a way of thinking expands current views on the interrelationship of motivation factors, as it adds a layer to our understandings of *competing factors*, by suggesting different factor rankings depending on the context in which factors interact. That is, even though meta-analyzes may reveal larger effect sizes for specific motivation factors (e.g., *appeal*) compared to others (e.g., *choice*), these differences in effect sizes should not be seen as universal rankings of factors that, if followed, will lead to optimal conditions for motivation; rather rankings of factors and consequently the realization of optimal conditions may vary, depending on the prioritization of factors in different contexts. Thus, in the same way that thorough descriptions of the linguistic environments in which constraints interact (e.g., at coda position, before a front vowel, between consonants) are necessary for understanding how they are ranked differently, thorough descriptions of the *contexts* in which motivation factors interact are also indispensable, which will be discussed in the next section.

6.3 The importance of thick descriptions

Some theories of motivation have been criticized for being based on a naïve understanding of human behavior (Clark, 1987; Graham, 1991; Weiner, 2018) and have been likened to *folk psychology* (also known as *commonsense psychology*), meaning that many of the predictions derived from these theories resemble mere common knowledge. For instance, Weiner (2018, p. 12) argues that some critics of attribution theory have dismissed it as *mere common sense*, labeling it “bubba (grandmother) psychology,” i.e., even a layperson, who may happen to be a grandmother, would be able to make the predictions derived from attribution theory. However, the author contends that although predictions from laypersons and experts may be similar, laypersons often do not realize how mechanisms leading to these predictions constitute the same theoretical system. Weiner (2018) then remarks that while laymen are not concerned with the deep structure involving the elemental laws of emotion and motivation, this is of primordial importance for psychologists attempting to establish a concise and broadly applicable conceptual framework.

According to Clark (1987), such a ubiquitous conceptual framework, often likened to folk psychology, is the desire/belief framework. This theoretical system stems from the *rationalist approach* for explaining human behavior as rational in essence. That is, explaining human actions (in contrast to the movements of planets or amoeba) requires the provision of their rationale; and providing a rationale, according to this approach, means showing how individuals’ actions were the rational thing to do based on the individuals’ *beliefs* and desires (Fay, 1996). For example, a man wants to lose weight (desire) and believes that if he quits eating candy, he will lose weight (belief), then he quits eating candy (action).

However, an important distinction in the understanding of human behavior is the difference between *movement* and *action*, in which the

former refers to an involuntary physical movement, while the latter concerns something that an agent does, performed for a purpose. To capture differences between movement and action, Fay (1996) refers to philosopher Gilbert Ryle's remarks on the difference between *thin* and *thick* descriptions. To illustrate this point, Ryle describes the same physical movement performed by two boys: the contraction of their eyelids. However, for one boy, this contraction was a twitch, and for the other a wink. The *twitch* was not an action, it was an *involuntary bodily movement*, while the *wink* was an *action*, something that the boy performed deliberately, with an intention, for instance to convey a message. According to Ryle, describing the boys' behavior with regards to the physical movements alone is a thin description that does not account for the meaning of these movements. To describe this movement as an action, a thick description is needed, which involves references not only to the physical movements of the individual, but also of his intentions and the social norms that give meaning to it (Fay, 1996).

Thick descriptions thus add important layers for our understanding of human behavior, especially since not all action can be explained with reference to desires and beliefs, given that these may contradict. For instance, in the candy example above, despite the man's desires and beliefs, he may still keep eating candy. That is, his behavior contradicts his desires and beliefs, which is often referred to as *weakness of will* (Haas, 2018), meaning that even though an agent believes that something should not be done, the agent does it anyway. To understand his behavior, other factors need to be considered, like emotions, habits, traditions, and motivations. Thus, to capture a better understanding of the factors influencing individuals' behaviors, more detailed pictures of the contexts in which actions take place, that is, thick descriptions of these contexts, is needed.

This importance of sufficient descriptions of context is also highlighted by Ratcliffe (2006) in his critique of folk psychology, where he argues that explanation and interpretation of human actions are indeed highly

influenced by *situational factors*, that is, the contexts in which behavior takes place. The author then suggests that to properly interpret and explain behavior, layers of situational context need to be progressively added. For instance, as shown in study 1 of the present project, only providing information about students' preferences towards performance feedback would not be sufficient, since it was found that the appreciation of this type of feedback was dependent on the type of *environment* to which students belonged (see factor *feedback* in article 1). In addition, differences regarding students' motivations may also be related to their cultures, which may prioritize motivational factors differently. For example, although one culture (or classroom) may value performance goals more than being given choices, other cultures may value mastery goals and choice more than performance goals, which, in turn, influence the individuals' desires and beliefs in these environments. Therefore, providing sufficient descriptions of the environments in which motivation factors are studied is essential. In fact, even in machine learning studies, the importance of sufficient descriptions of situational factors is emphasized. For instance, in a study conducted by Rashkin (2018), a machine learning model was tested to track the mental state of story characters and to predict the characters' motivations and emotions. In this study, when the model is tested without context, it predicts the same emotional reactions for different characters, but with context, it predicts correct emotions for each character.

Thus, although the present project advocates for the importance of self-reported data in motivation research, it does not aim to undermine the importance of observations, as they are necessary for providing a more detailed picture of the contexts in which motivation phenomena is investigated. However, the present thesis remarks that for tapping into students' inner drives, observation *alone* is not sufficient. That is, inviting students to express their own opinions through self-reported data is also necessary.

7 Limitations and future directions

In addition to the limitations of the three studies that compose this thesis, which are presented in the three articles that accompany this dissertation (see appendices 1-3), the present project also presents two important limitations, which will be addressed in this section.

The first limitation concerns the main source of data used in this project, namely students' self-reported data. Although the present thesis shows that students' self-reports is a rich source of data for getting an insight into the students' motivations to write, this data is not triangulated with other sources, like classroom observations, or teacher data, in the present project. Triangulating student self-reported data with classroom observations and teacher data in studies of student motivation is important for at least three reasons. First, as discussed in the present thesis, student-reported data can be influenced by various factors, thus triangulating it with other sources of data can enhance the validity of findings. Each data source may offer unique information that contributes to a more holistic understanding of student motivation. For example, while student-reported data can provide insight into their perceptions, beliefs, and attitudes towards writing, classroom observations can help researchers situate student-reported data within the broader classroom context through observations of students' and teachers' behaviors, interactions, and levels of engagement. Teacher data can also provide additional information on the teachers' beliefs, instructional strategies, and classroom dynamics.

Second, by triangulating data sources, discrepancies may be identified. For example, a student may report high levels of motivation and engagement in class, but classroom observations and teacher data may indicate a lack of participation or interest. Exploring these patterns can help researchers gain a deeper understanding of factors influencing motivation and can inform the development of interventions. Third,

triangulating data with classroom observations and teacher data enhances the practical applications of research on student motivation as it helps researchers with identifying effective instructional practices and strategies that promote motivation. These findings can inform educators and help them create learning environments that foster student motivation and engagement. In addition, it results in a more holistic view of educational contexts and provides a more robust basis for decision-making and for creating supportive learning environments that cater not only for the needs of students, but also for the needs of different educational stakeholders.

Knowing the value of triangulating different data sources, it is important to note that the original plan for this research project indeed included classroom observations, but due to COVID-19 restrictions, these could not be pursued. However, this is the direction that I see myself taking in the future. In addition, although I have not conducted classroom observations or collected teacher data myself in the present project, some of the studies included in the systematic review conducted for studies 1 and 2 have indeed triangulated student-reported data with other data sources. Therefore, when synthesizing findings in articles 1 and 2, I was able to provide thick descriptions of the classroom contexts in which factors affecting students' motivations were identified (article 1) and discuss discrepancies among the students' views on their motivations to write and the views of others (article 2). Consequently, discussions of consistencies and discrepancies among different data sources, which contributes to a more holistic view of students' writing motivation, are indeed an integral part of the present thesis. However, more studies comparing different data sources in investigations of writing motivation are needed for systematically mapping consistencies and discrepancies among these sources of data.

The second limitation is related to the focus of the present project being primarily on writing *motivation*, and not including additional analyses of writing *performance*. However, as explained in the introduction of this

thesis – and emphasized throughout – to help students perform successfully in writing, targeting their skill levels is not enough, strengthening their motivation levels is also necessary. In fact, many researchers have found positive associations between writing motivation and performance (Graham, Daley, et al., 2018; Graham et al., 2017; Pajares et al., 2001; Wijekumar et al., 2019), and in their review of writing motivation in school, Camacho et al. (2021, p. 235) remark that “findings from regression analyses and structural equation models indicated that motivational constructs predicted or contributed (both directly and indirectly) to students’ writing performance.” Given this positive relationship between writing motivation and performance, findings from the present project can therefore ultimately also contribute to students’ writing performance. To achieve this goal, the present project has sought to contribute knowledge on factors influencing students’ writing motivation and classroom practices that facilitate the implementation of these factors (study 1), and knowledge on valid methods for capturing the students’ voices on their writing motivations (studies 2 and 3), which gives different educational stakeholders the possibility of mapping students’ complex needs and thereby addressing them more appropriately.

Departing from the findings of the present thesis, four main directions for future research are identified. First, given the various interactions among motivation factors (e.g., *constructive feedback* is appreciated by students but mostly in *environments* that value this type of feedback), future classroom investigations should observe and interpret how exemplary teachers implement these factors in their classrooms. For instance, by measuring students’ writing motivation and identifying classrooms that are highly motivated, researchers can observe and analyze how teachers in these classrooms create the conditions for motivational writing instruction, and if any of the identified factors in study 1 are more decisive than others. Second, the present thesis investigates L1 writing in mainstream classrooms, but studies looking at

other distinct student samples like L2 writers or students with special needs, as well as other educational settings, like private writing courses, specialized courses for struggling writers, or summer camps, would add important layers to the present findings. In addition, investigating if the identified factors in study 1 also apply to other student samples and/or educational settings would be a valuable addition to accounts of writing motivation in educational settings. Third, given the rapid advancement of AI tools for generating texts, there is a timely need to investigate how the affordances of digital tools may foster writing motivation, and whether some of the identified factors in this thesis may also be realized analogously. Finally, given the variation amongst motivation measures identified in the present thesis, future research should aim to validate, quantify, and organize methods for assessing students' motivations to write, especially with young children.

References

- Abbott, J. A. (2000). “Blinking Out” and “Having the Touch”: Two Fifth-Grade Boys Talk about Flow Experiences in Writing. *Written Communication*, 17(1), 53–92. <https://doi.org/10.1177/0741088300017001003>
- Acquah, E. O., & Katz, H. T. (2020). Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review. *Computers & Education*, 143, 103667. <https://doi.org/10.1016/j.compedu.2019.103667>
- Akyol, H., & Aktaş, N. (2018). The Relationship between Fourth-Grade Primary School Students’ Story-writing Skills and Their Motivation to Write. *Universal Journal of Educational Research*, 6(12), 2772–2779. <https://doi.org/10.13189/ujer.2018.061211>
- Alexander, P. A., & Fox, E. (2004). A Historical Perspective on Reading Research and Practice. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical Models and Processes of Reading* (5th ed., Vol. 1, pp. 33–68). International Reading Association. <https://doi.org/10.1598/0872075028.2>

- Andrzejczak, N., Trainin, G., & Poldberg, M. (2005). From Image to Text: Using Images in the Writing Process. *International Journal of Education & the Arts*, 6(12), 1–17.
- Archangeli, D. B. (1999). Introducing optimality theory. *Annual Review of Anthropology*, 28(1), 531–552.
- Baker, S. F., & Lastrapes, R. E. (2019). The writing performance of elementary students using a digital writing application: Results of a teacher–librarian collaboration. *Interactive Technology and Smart Education*, 16(4), 343–362.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1994). Self-efficacy. In *Encyclopedia of Human Behavior* (Vol. 4, pp. 71–81). Academic Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, 5(1), 307–337.
- Barton, D. (2007). *Literacy: An introduction to the ecology of written language* (2nd ed). Blackwell Pub.

- Bayat, S. (2016). The Effectiveness of the Creative Writing Instruction Program Based on Speaking Activities (CWIPSA). *International Electronic Journal of Elementary Education*, 8(4), 617–628.
- Bazerman, C., Graham, S., Applebee, A. N., Matsuda, P. K., Berninger, V. W., Murphy, S., Brandt, D., Rowe, D. W., & Schleppegrell, M. (2017). Taking the Long View on Writing Development. *Research in the Teaching of English*, 51(3), 351–360.
- Beck, N., & Fetherston, T. (2003). The Effects of Incorporating a Word Processor into a Year Three Writing Program. *Information Technology in Childhood Education Annual*, 2003(1), 139–161.
- Berninger, V. W., Abbott, R. D., Jones, J., Wolf, B. J., Gould, L., Anderson-Youngstrom, M., Shimada, S., & Apel, K. (2006). Early Development of Language by Hand: Composing, Reading, Listening, and Speaking Connections; Three Letter-Writing Modes; and Fast Mapping in Spelling. *Developmental Neuropsychology*, 29(1), 61–92.
https://doi.org/10.1207/s15326942dn2901_5
- Berninger, V. W., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in

- beginning writing. *Reading and Writing*, 4(3), 257–280.
<https://doi.org/10.1007/BF01027151>
- Bhaskar, R., Danermark, B., & Price, L. (2017). *Interdisciplinarity and Wellbeing: A Critical Realist General Theory of Interdisciplinarity* (1st ed.). Routledge.
<https://doi.org/10.4324/9781315177298>
- Biesta, G. (2007). Why “What Works” Won’t Work: Evidence-Based Practice and the Democratic Deficit in Educational Research. *Educational Theory*, 57(1), 1–22. <https://doi.org/10.1111/j.1741-5446.2006.00241.x>
- Biesta, G. (2010). Why ‘What Works’ Still Won’t Work: From Evidence-Based Education to Value-Based Education. *Studies in Philosophy and Education*, 29(5), 491–503.
<https://doi.org/10.1007/s11217-010-9191-x>
- Bingham, G. E., Quinn, M. F., & Gerde, H. K. (2017). Examining early childhood teachers’ writing practices: Associations between pedagogical supports and children’s writing skills. *Early Childhood Research Quarterly*, 39, 35–46.
<https://doi.org/10.1016/j.ecresq.2017.01.002>

- Boersma, P. (1997). Functional optimality theory. *Proceedings of the Institute of Phonetic Sciences of University of Amsterdam*, 21, 37–42.
- Boscolo, P., & Gelati, C. (2018). Motivating Writers. In S. Graham, C. A. MacArthur, & M. Hebert (Eds.), *Best Practices in Writing Instruction, Third Edition*. Guilford Publications.
<http://ebookcentral.proquest.com/lib/uisbib/detail.action?docID=5603731>
- Boscolo, P., Gelati, C., & Galvan, N. (2012). Teaching Elementary School Students to Play With Meanings and Genre. *Reading & Writing Quarterly*, 28(1), 29–50.
<https://doi.org/10.1080/10573569.2012.632730>
- Boyacı, S. D. B., & Güner, M. (2018). The Impact of Authentic Material Use on Development of the Reading Comprehension, Writing Skills and Motivation in Language Course. *International Journal of Instruction*, 11(2), 351–368.
<https://doi.org/10.12973/iji.2018.11224a>
- Bradford, K. L., Newland, A. C., Rule, A. C., & Montgomery, S. E. (2016). Rubrics as a Tool in Writing Instruction: Effects on the

- Opinion Essays of First and Second Graders. *Early Childhood Education Journal*, 44(5), 463–472.
<https://doi.org/10.1007/s10643-015-0727-0>
- Brand, A. (1998). Values: Understanding writing through brain biology. *Rhetoric Review*, 16(2), 290–309.
<https://doi.org/10.1080/07350199809389097>
- Bruning, R., & Horn, C. (2000). Developing Motivation to Write. *Educational Psychologist*, 35(1), 25–37.
https://doi.org/10.1207/S15326985EP3501_4
- Brunton, G., Stansfield, C., Caird, J., & Thomas, J. (2017). Finding relevant studies. In D. Gough, J. Thomas, & S. Oliver (Eds.), *An introduction to systematic reviews* (pp. 93–122). SAGE Publications Ltd.
- Camacho, A., Alves, R. 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>
- Chahal, D. (2017). Evidence-based practice and its discontents in Academic Language and Learning. *Pedagogy, Culture &*

- Society*, 25(4), 601–617.
<https://doi.org/10.1080/14681366.2017.1312495>
- Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain*, 78(1), 27–37.
[https://doi.org/10.1016/S0304-3959\(98\)00112-2](https://doi.org/10.1016/S0304-3959(98)00112-2)
- Cheek, J. (2015). *It depends: Possible impacts of moving the field of mixed methods research toward best practice guidelines.*
- Chen, Z.-H., & Liu, W.-Y. (2019). A six-stage story structure approach for elementary students' story production: Quality, interest, and attitude. *Computer Assisted Language Learning*.
<https://doi.org/10.1080/09588221.2019.1688354>
- Cheung, Y. L., & Jang, H. (2019). Teaching writing to elementary school students: An activity theory perspective. *Irish Educational Studies*, 38(4), 451–466.
- Chohan, S. K. (2011). Any Letter for me? Relationships Between an Elementary School Letter Writing Program and Student Attitudes, Literacy Achievement, and Friendship Culture. *Early Childhood Education Journal*, 39(1), 39–50.
<https://doi.org/10.1007/s10643-010-0438-5>

- Chomsky, N. (1959). A review of BF Skinner's Verbal behavior. *Language*, 35(1), 26–58. <https://doi.org/10.2307/411334>
- Clark, A. (1987). From Folk Psychology to Naive Psychology. *Cognitive Science*, 11(2), 139–154. https://doi.org/10.1207/s15516709cog1102_1
- Cohen, S. M., & Reeve, C. D. C. (2021). Aristotle's Metaphysics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2021/entries/aristotle-metaphysics/>
- Cook, D. A., & Artino, A. R. (2016). Motivation to learn: An overview of contemporary theories. *Medical Education*, 50(10), 997–1014. <https://doi.org/10.1111/medu.13074>
- Cordero, K., Nussbaum, M., Ibaseta, V., Otaíza, M. J., & Chiuminatto, P. (2018). Read, write, touch: Co-construction and multiliteracies in a third-grade digital writing exercise. *Journal of Computer Assisted Learning*, 34(2), 162–173.
- Cowen, N. (2019). For whom does “what works” work? The political economy of evidence-based education. *Educational Research*

- and Evaluation*, 25(1–2), 81–98.
<https://doi.org/10.1080/13803611.2019.1617991>
- Cremin, T., Gooch, K., Blakemore, L., Goff, E., & Macdonald, R. (2006). Connecting drama and writing: Seizing the moment to write. *Research in Drama Education*, 11(3), 273–291.
- Cutspec, P. (2004). Bridging the research-to-practice-gap: Evidence-based education. *Centerscope: Evidence-Based Approaches to Early Childhood Development*, 2(2), 1–8.
- Daniels, A. (2004). Composition instruction: Using technology to motivate students to write. *Information Technology in Childhood Education Annual, 2004*(1), 157–177.
- Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies*, 47(2), 108–121.
<https://doi.org/10.1111/1467-8527.00106>
- Davies, P. (2000). The Relevance of Systematic Reviews to Educational Policy and Practice. *Oxford Review of Education*, 26(3–4), 365–378. <https://doi.org/10.1080/713688543>
- Deane, P. (2018). The Challenges of Writing in School: Conceptualizing Writing Development Within a Sociocognitive Framework.

References

- Educational Psychologist*, 53(4), 280–300.
<https://doi.org/10.1080/00461520.2018.1513844>
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- Definition of SELF-REPORT*. (2023, February 26).
<https://www.merriam-webster.com/dictionary/self-report>
- Definition of SYNTHESIS*. (2023, February 6). <https://www.merriam-webster.com/dictionary/synthesis>
- Definition of VOICE*. (2023, March 1).
<https://dictionary.cambridge.org/dictionary/english/voice>
- Duran, E., & Karatas, A. (2019). Elimination of Writing Difficulty in Primary School: An Action Research. *International Journal of Progressive Education*, 15(5), 288–300.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256.
<https://doi.org/10.1037/0033-295X.95.2.256>
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., Midgley, C., & Spence, J. T. (1983). Expectancies,

- values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146).
- Ekholm, E., Zumbunn, 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>
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54(1), 5–12. <https://doi.org/10.1037/0022-3514.54.1.5>
- Erdoğan, T., & Erdoğan, Ö. (2013). A Metaphor Analysis of the Fifth Grade Students' Perceptions about Writing. *The Asia-Pacific Education Researcher*, 22(4), 347–355. <https://doi.org/10.1007/s40299-012-0014-4>
- Fàbregues, S., & Molina-Azorín, J. F. (2017). Addressing quality in mixed methods research: A review and recommendations for a future agenda. *Quality & Quantity*, 51(6), 2847–2863. <https://doi.org/10.1007/s11135-016-0449-4>
- Fay, B. (1996). *Contemporary philosophy of social science: A multicultural approach* (Vol. 1). Blackwell Oxford.

- Felzmann, H. (2009). Ethical Issues in School-Based Research. *Research Ethics*, 5(3), 104–109. <https://doi.org/10.1177/174701610900500304>
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365. <https://doi.org/10.2307/356600>
- Fox, R. (2001). Helping young writers at the point of writing. *Language and Education*, 15(1), 1–13.
- Gallini, J. K., & Zhang, Y.-L. (1997). Socio-Cognitive Constructs and Characteristics of Classroom Communities: An Exploration of Relationships. *Journal of Educational Computing Research*, 17(4), 321–339. <https://doi.org/10.2190/Y71J-T3EJ-VFTJ-P6WG>
- Garcia, G. A., Núñez, A.-M., & Sansone, V. A. (2019). Toward a Multidimensional Conceptual Framework for Understanding “Servingness” in Hispanic-Serving Institutions: A Synthesis of the Research. *Review of Educational Research*, 89(5), 745–784. <https://doi.org/10.3102/0034654319864591>

- Gerde, H. K., Bingham, G. E., & Wasik, B. A. (2012). Writing in Early Childhood Classrooms: Guidance for Best Practices. *Early Childhood Education Journal*, 40(6), 351–359. <https://doi.org/10.1007/s10643-012-0531-z>
- Gonzalez, T. E., Hernandez-Saca, D. I., & Artiles, A. J. (2017). In search of voice: Theory and methods in K-12 student voice research in the US, 1990–2010. *Educational Review*, 69(4), 451–473. <https://doi.org/10.1080/00131911.2016.1231661>
- Gough, D. (2013). Knowledge Mobilisation in Education in England. In B. Levin, J. Qi, H. Edelstein, & J. Sohn (Eds.), *The impact of research in education* (pp. 65–84). Policy Press. <https://doi.org/10.1332/policypress/9781447306207.003.0004>
- Gough, D., Oliver, S., & Thomas, J. (2012). Introducing systematic reviews. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 1–16). SAGE.
- Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review*, 3, 5–39.

- 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., Berninger, V., & Abbott, R. (2012). Are Attitudes Toward Writing and Reading Separable Constructs? A Study With Primary Grade Children. *Reading & Writing Quarterly*, 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., Bollinger, A., Olson, C. B., D’Aoust, C., Macarthur, C., McCutchen, D., & Olinghouse, N. G. (2012). *Teaching Elementary School Students to Be Effective Writers: A Practice Guide*. NCEE 2012-4058.
- Graham, S., Daley, S. G., Aitken, A. 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., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26*(1), 1–15. <https://doi.org/10.1007/s11145-012-9395-2>
- Graham, S., & Harris, K. R. (2016). A Path to Better Writing: Evidence-Based Practices in the Classroom. *The Reading Teacher, 69*(4), 359–365. <https://doi.org/10.1002/trtr.1432>
- 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., Kiuahara, S. A., Harris, K. R., & 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. <https://doi.org/10.1086/693009>
- Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for Writing: A Meta-Analysis of the Impact of Reading Interventions

- on Writing. *Review of Educational Research*, 88(2), 243–284.
<https://doi.org/10.3102/0034654317746927>
- Graham, S., Schwartz, S. S., & MacArthur, C. A. (1993). Knowledge of writing and the composing process, attitude toward writing, and self-efficacy for students with and without learning disabilities. *Journal of Learning Disabilities*, 26(4), 237–249.
- Graham, S., Tavsanlı, O. F., & Kaldırım, A. (2022). Improving Writing Skills of Students in Turkey: A Meta-analysis of Writing Interventions. *Educational Psychology Review*, 34(2), 889–934.
<https://doi.org/10.1007/s10648-021-09639-0>
- Grenner, E., Johansson, V., van de Weijer, J., & Sahlén, B. (2020). Effects of intervention on self-efficacy and text quality in elementary school students' narrative writing. *Logopedics Phoniatics Vocology*, 1–10.
<https://doi.org/10.1080/14015439.2019.1709539>
- Griffo, V. B., Mada, C. L., Pearson, P. D., & Raphael, T. E. (2015). Current issues and best practices in literacy instruction. In L. M. Morrow & L. B. Gambrell (Eds.), *Best practices in literacy instruction* (Sixth Edition, pp. 37–60). The Guilford Press.

- Guthrie, J. T. (2004). Differentiating Instruction for Struggling Readers Within the CORI Classroom. In J. T. Guthrie & Wigfield (Eds.), *Motivating reading comprehension: Concept-oriented reading instruction* (pp. 173–193). Routledge.
- Guthrie, J. T., & Cox, K. E. (2001). Classroom Conditions for Motivation and Engagement in Reading. *Educational Psychology Review*, *13*(3), 283–302. <https://doi.org/10.1023/A:1016627907001>
- Guthrie, J. T., Wigfield, A., & Perencevich, K. C. (Eds.). (2004). *Motivating Reading Comprehension: Concept-Oriented Reading Instruction*. Routledge. <https://doi.org/10.4324/9781410610126>
- Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional Contexts for Engagement and Achievement in Reading. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 601–634). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_29
- Haas, J. (2018). An empirical solution to the puzzle of weakness of will. *Synthese*, *195*(12), 5175–5195. <https://doi.org/10.1007/s11229-018-1712-0>

- Håland, A., Hoem, T. F., & McTigue, E. M. (2019). Writing in First Grade: The Quantity and Quality of Practices in Norwegian Classrooms. *Early Childhood Education Journal*, *47*(1), 63–74. <https://doi.org/10.1007/s10643-018-0908-8>
- Hall, A. H., & Axelrod, Y. (2014). “I Am Kind of a Good Writer and Kind of Not”: Examining Students’ Writing Attitudes. *Journal of Research in Education*, *24*(2), 34–50.
- Hall, A. H., Boyer, D. M., & Beschorner, E. A. (2017). Examining Kindergarten Students’ Use of and Interest in Informational Text. *Early Childhood Education Journal*, *45*(5), 703–711.
- Hammersley, M. (1997). Educational Research and Teaching: A response to David Hargreaves’ TTA lecture. *British Educational Research Journal*, *23*(2), 141–161. <https://doi.org/10.1080/0141192970230203>
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, *94*(3), 638–645. <https://doi.org/10.1037/0022-0663.94.3.638>

- Hargreaves, D. H. (1996). *Teaching as a Research-based Profession: Possibilities and Prospects*. Teacher Training Agency.
- Hayes, J. R. (1996). A New Framework for Understanding Cognition and Affect in Writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences and Applications* (pp. 1–27). Erlbaum.
- Hayes, J. R., & Flower, L. (1980). Identifying the Organization of Writing Processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing: An Interdisciplinary Approach* (pp. 3–30). Lawrence Erlbaum.
- Hemsley-Brown, J., & Sharp, C. (2003). The Use of Research to Improve Professional Practice: A systematic review of the literature. *Oxford Review of Education*, 29(4), 449–471.
<https://doi.org/10.1080/0305498032000153025>
- Hertz, M., & Heydenberk, W. (1997). A Kindergarten Writing Workshop: How Kindergarten Students Grow as Writers. *Reading Horizons*, 37(3), 203–214.
- Hidi, S. (1990). Interest and Its Contribution as a Mental Resource for Learning. *Review of Educational Research*, 60(4), 549–571.

- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: An intervention study. *Learning and Instruction, 12*(4), 429–446. [https://doi.org/10.1016/S0959-4752\(01\)00009-3](https://doi.org/10.1016/S0959-4752(01)00009-3)
- Hidi, S., & Boscolo, P. (2008). Motivation and Writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144–157). Guilford Press.
- Hier, B. O., & Mahony, K. E. (2018). The contribution of mastery experiences, performance feedback, and task effort to elementary-aged students' self-efficacy in writing. *School Psychology Quarterly, 33*(3), 408–418. <https://doi.org/10.1037/spq0000226>
- Hitchcock, J. H., & Onwuegbuzie, A. J. (2020). Developing Mixed Methods Crossover Analysis Approaches. *Journal of Mixed Methods Research, 14*(1), 63–83. <https://doi.org/10.1177/1558689819841782>
- Hoddy, E. T. (2019). Critical realism in empirical research: Employing techniques from grounded theory methodology. *International*

- Journal of Social Research Methodology*, 22(1), 111–124.
<https://doi.org/10.1080/13645579.2018.1503400>
- Hodges, T. S. (2017). Theoretically Speaking: An Examination of Four Theories and How They Support Writing in the Classroom. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 90(4), 139–146.
<https://doi.org/10.1080/00098655.2017.1326228>
- Hughes, A., Galbraith, D., & White, D. (2011). Perceived Competence: A Common Core for Self-Efficacy and Self-Concept? *Journal of Personality Assessment*, 93(3), 278–289.
<https://doi.org/10.1080/00223891.2011.559390>
- Hunter, W. J. (2017). Evidence-Based Teaching in the 21st Century. *Canadian Journal of Education*, 40(2), 1–6.
- Ihmeideh, F. (2015). The impact of dramatic play centre on promoting the development of children's early writing skills. *European Early Childhood Education Research Journal*, 23(2), 250–263.
<https://doi.org/10.1080/1350293X.2014.970848>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational*

- Researcher*, 33(7), 14–26.
<https://doi.org/10.3102/0013189X033007014>
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2), 112–133.
<https://doi.org/10.1177/1558689806298224>
- Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2016). Beyond the Rainbow: Retrieval Practice Leads to Better Spelling than does Rainbow Writing. *Educational Psychology Review*, 28(2), 385–400.
<https://doi.org/10.1007/s10648-015-9330-6>
- Kaderavek, J. N., Cabell, S. Q., & Justice, L. M. (2009). Early writing and spelling development. In P. M. Rhyner (Ed.), *Emergent literacy and language development: Promoting learning in early childhood* (pp. 104–152). Guilford Press.
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Hachette UK.
- Kanala, S., Nousiainen, T., & Kankaanranta, M. (2013). Using a mobile application to support children’s writing motivation. *Interactive*

- Technology and Smart Education*, 10(1), 4–14.
<https://doi.org/10.1108/17415651311326419>
- Kholisiyah, R. N., Rukayah, & Indriayu, M. (2018). Achievement Motivation Analysis of Outstanding Students in Learning Writing at Primary Schools. *International Journal of Educational Methodology*, 4(3), 133–139.
<https://doi.org/10.12973/ijem.4.3.133>
- Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2006). Writing as a learning tool: Testing the role of students' writing strategies. *European Journal of Psychology of Education*, XXI, 17–34.
- Kim, J.-A., & Lorschach, A. W. (2005). Writing Self-Efficacy in Young Children: Issues for the Early Grades Environment. *Learning Environments Research*, 8(2), 157–175.
<https://doi.org/10.1007/s10984-005-7248-5>
- Klassen, R. (2002a). A Question of Calibration: A Review of the Self-Efficacy Beliefs of Students with Learning Disabilities. *Learning Disability Quarterly*, 25(2), 88–102.
<https://doi.org/10.2307/1511276>

- Klassen, R. (2002b). Writing in Early Adolescence: A Review of the Role of Self-Efficacy Beliefs. *Educational Psychology Review*, 31.
- Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, 24(1), 120–124. <https://doi.org/10.1080/13814788.2017.1375092>
- Korth, B. B., Wimmer, J. J., Wilcox, B., Morrison, T. G., Harward, S., Peterson, N., Simmerman, S., & Pierce, L. (2017). Practices and Challenges of Writing Instruction in K-2 Classrooms: A Case Study of Five Primary Grade Teachers. *Early Childhood Education Journal*, 45(2), 237–249. <https://doi.org/10.1007/s10643-016-0774-1>
- Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2019). Children’s Creativity: A Theoretical Framework and Systematic Review. *Review of Educational Research*, 89(1), 93–124. <https://doi.org/10.3102/0034654318815707>

- Kvernbekk, T. (2017). Evidence-Based Educational Practice. In *Oxford Research Encyclopedia of Education*. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780190264093.013.187>
- Lather, P. (2004). Scientific research in education: A critical perspective1. *British Educational Research Journal*, 30(6), 759–772. <https://doi.org/10.1080/0141192042000279486>
- Leaper, C. (2011). More similarities than differences in contemporary theories of social development?: A plea for theory bridging. *Advances in Child Development and Behavior*, 40, 337–378.
- Lee, B. K., & Enciso, P. (2017). The Big Glamorous Monster (or Lady Gaga’s Adventures at Sea): Improving Student Writing Through Dramatic Approaches in Schools. *Journal of Literacy Research*, 49(2), 157–180. <https://doi.org/10.1177/1086296X17699856>
- Li, X., & Chu, S. K. W. (2018). Using design-based research methodology to develop a pedagogy for teaching and learning of Chinese writing with wiki among Chinese upper primary school students. *Computers & Education*, 126, 359–375.
<https://doi.org/10.1016/j.compedu.2018.06.009>

- Liao, C. C. Y., Chang, W.-C., & Chan, T.-W. (2018). The effects of participation, performance, and interest in a game-based writing environment. *Journal of Computer Assisted Learning*, *34*(3), 211–222. <https://doi.org/10.1111/jcal.12233>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
- Martin, A. J., Ginns, P., & Papworth, B. (2017). Motivation and engagement: Same or different? Does it matter? *Learning and Individual Differences*, *55*, 150–162. <https://doi.org/10.1016/j.lindif.2017.03.013>
- Mata, L. (2011). Motivation for Reading and Writing in Kindergarten Children. *Reading Psychology*, *32*(3), 272–299. <https://doi.org/10.1080/02702711.2010.545268>
- McBride, C., Snow, C. E., Kucirkova, N., & Grøver, V. (2017). Old and new: Reflecting on the enduring key issues in early literacy. In N. Kucirkova, C. E. Snow, V. Grøver, & C. McBride (Eds.), *The Routledge international handbook of early literacy education: A Contemporary Guide to Literacy Teaching and Interventions in a Global Context* (pp. 373–382). Routledge.

- McKnight, L., & Morgan, A. (2019). A broken paradigm? What education needs to learn from evidence-based medicine. *Journal of Education Policy*, 1–17. <https://doi.org/10.1080/02680939.2019.1578902>
- McTigue, E. M., Solheim, O. J., Walgermo, B., Frijters, J., & Foldnes, N. (2019). How can we determine students' motivation for reading before formal instruction? Results from a self-beliefs and interest scale validation. *Early Childhood Research Quarterly*, 48, 122–133. <https://doi.org/10.1016/j.ecresq.2018.12.013>
- Merisuo-Storm, T. (2006). Girls and Boys Like to Read and Write Different Texts. *Scandinavian Journal of Educational Research*, 50(2), 111–125. <https://doi.org/10.1080/00313830600576039>
- Mertens, D. M., & Hesse-Biber, S. (2012). Triangulation and Mixed Methods Research: Provocative Positions. *Journal of Mixed Methods Research*, 6(2), 75–79. <https://doi.org/10.1177/1558689812437100>
- Miller, D. M., McTigue, E. M., & Scott, C. E. (2015). The Quality of Recent Studies in Content-Area Writing in Secondary

- Classrooms. *Literacy Research: Theory, Method, and Practice*, 64(1), 461–477. <https://doi.org/10.1177/2381336915617602>
- Miller, D. M., Scott, C. E., & McTigue, E. M. (2018). Writing in the Secondary-Level Disciplines: A Systematic Review of Context, Cognition, and Content. *Educational Psychology Review*, 30(1), 83–120. <https://doi.org/10.1007/s10648-016-9393-z>
- Miller, S. D., & Meece, J. L. (1997). Enhancing Elementary Students' Motivation to Read and Write: A Classroom Intervention Study. *The Journal of Educational Research*, 90(5), 286–299. <https://doi.org/10.1080/00220671.1997.10544585>
- Miller, S. D., & Meece, J. L. (1999). Third graders' motivational preferences for reading and writing tasks. *Elementary School Journal*, 100(1), 19.
- Mockler, N., & Groundwater-Smith, S. (2017). Teacher Research: A Knowledge-Producing Profession? In P. Grootenboer, C. Edwards-Groves, & S. Choy (Eds.), *Practice Theory Perspectives on Pedagogy and Education* (pp. 215–230). Springer Singapore. https://doi.org/10.1007/978-981-10-3130-4_11

- Moser, A., & Korstjens, I. (2018). Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. *European Journal of General Practice, 24*(1), 9–18. <https://doi.org/10.1080/13814788.2017.1375091>
- Muenks, K., & Miele, D. B. (2017). Students' Thinking About Effort and Ability: The Role of Developmental, Contextual, and Individual Difference Factors. *Review of Educational Research, 87*(4), 707–735. <https://doi.org/10.3102/0034654316689328>
- Munro, E. (2014). Evidence-Based Policy. In N. Cartwright & E. Montuschi (Eds.), *Philosophy of social science: A new introduction* (First edition, pp. 48–67). Oxford University Press.
- Nair, S. S., Tay, L. Y., & Koh, J. H. L. (2013). Students' motivation and teachers' teaching practices towards the use of blogs for writing of online journals. *Educational Media International, 50*(2), 108–119. <https://doi.org/10.1080/09523987.2013.795351>
- Nakamura, W. (1999). Functional optimality theory. *Functionalism and Formalism, 2*, 253–276.
- Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: Meanings and applications. *Educational Research,*

59(2), 127–135.

<https://doi.org/10.1080/00131881.2017.1314115>

NESH - The National Committee for Research Ethics in the Social Sciences and the Humanities. (2016). *Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology*.

https://www.etikkom.no/globalassets/documents/english-publications/60127_fek_guidelines_nesh_digital_corr.pdf

Nicolaidou, I. (2012). Can process portfolios affect students' writing self-efficacy? *International Journal of Educational Research*, 56, 10–22. <https://doi.org/10.1016/j.ijer.2012.08.002>

Nolen, S. B. (2007). Young Children's Motivation to Read and Write: Development in Social Contexts. *Cognition and Instruction*, 25(2–3), 219–270. <https://doi.org/10.1080/07370000701301174>

Norris, N. (1996). Professor Hargreaves, the TTA and evidence-based practice. *British Educational Research Association Newsletter, Research Intelligence, Number 57*, July 2-4.

Nystrand, M. (1989). A social-interactive model of writing. *Written Communication*, 6, 66–85.

- Olson, D. R. (2009). The History of Writing. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (1st ed, pp. 6–16). SAGE Publications Ltd.
- Pae, C.-U. (2015). Why Systematic Review rather than Narrative Review? *Psychiatry Investigation*, 12(3), 417–419. <https://doi.org/10.4306/pi.2015.12.3.417>
- Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development*, 33(4), 214–221.
- Paquette, K. R. (2008). STUDY OF ELEMENTARY STUDENTS' ATTITUDES ABOUT WRITING AFTER A CROSS-AGE TUTORING EXPERIENCE. *Reading Improvement*, 45(4), 181–191.
- Paquette, K. R., Byrd, C. M., & Dragich, D. (2013). AN INVESTIGATION OF IMPLEMENTING DROP EVERYTHING AND WRITE (DEW) IN EARLY CHILDHOOD/ELEMENTARY CLASSROOMS. *Education*, 133(3), 356–369.

- Perry, N. E. (1998). Young Children's Self-Regulated Learning and Contexts that Support It. *Journal of Educational Psychology*, *90*(4), 715–729.
- Perry, N. E., Nordby, C. J., & VandeKamp, K. O. (2003). Promoting Self-Regulated Reading and Writing at Home and School. *Elementary School Journal*, *103*(4), 317–338.
- Perry-Hazan, L. (2021). Students' Perceptions of Their Rights in School: A Systematic Review of the International Literature. *Review of Educational Research*, *91*(6), 919–957.
<https://doi.org/10.3102/00346543211031642>
- Pettersen, E. B., Ertesvåg, S. K., Pöysä, S., Vaaland, G. S., & Virtanen, T. E. (2023). Students' situational engagement and its association with overall engagement: The application of the InSitu instrument in the context of a Norwegian lower secondary school. *Scandinavian Journal of Educational Research*, 1–14.
<https://doi.org/10.1080/00313831.2023.2175245>
- Plano Clark, V. L. (2017). Mixed methods research. *The Journal of Positive Psychology*, *12*(3), 305–306.
<https://doi.org/10.1080/17439760.2016.1262619>

- Pollington, M. F., Wilcox, B., & Morrison, T. G. (2001). SELF-PERCEPTION IN WRITING: THE EFFECTS OF WRITING WORKSHOP AND TRADITIONAL INSTRUCTION ON INTERMEDIATE GRADE STUDENTS. *Reading Psychology*, 22(4), 249–265. <https://doi.org/10.1080/02702710127640>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rashkin, H., Bosselut, A., Sap, M., Knight, K., & Choi, Y. (2018). *Modeling Naive Psychology of Characters in Simple Commonsense Stories*. <https://doi.org/10.48550/ARXIV.1805.06533>
- Ratcliffe, M. (2006). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. Springer.
- Read, S. (2005). First and Second Graders Writing Informational Text. *The Reading Teacher*, 59(1), 36–44. <https://doi.org/10.1598/RT.59.1.4>
- Reeve, J. (2012). A Self-determination Theory Perspective on Student Engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie

- (Eds.), *Handbook of Research on Student Engagement* (pp. 149–172). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_7
- Reschly, A. L., & Christenson, S. L. (2012). Jingle, Jangle, and Conceptual Haziness: Evolution and Future Directions of the Engagement Construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 3–19). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_1
- 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. <https://doi.org/10.1037/0003-066X.55.1.68>
- Saldaña, J. (2014). Coding and Analysis Strategies. In P. Leavy (Ed.), *The Oxford Handbook of Qualitative Research* (pp. 580–598). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199811755.013.001>
- Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-

- Methods Research. *Quality and Quantity*, 36(1), 43–53.
<https://doi.org/10.1023/A:1014301607592>
- Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A Meta-Analysis of Interventions for Struggling Readers in Grades 4–12: 1980–2011. *Journal of Learning Disabilities*, 48(4), 369–390. <https://doi.org/10.1177/0022219413504995>
- Schrodt, K. E., Elleman, A. M., FitzPatrick, E. R., Hasty, M. M., Kim, J. K., Tharp, T. J., & Rector, H. (2019). An Examination of Mindset Instruction, Self-Regulation, and Writer’s Workshop on Kindergarteners’ Writing Performance and Motivation: A Mixed-Methods Study. *Reading & Writing Quarterly*, 35(5), 427–444. <https://doi.org/10.1080/10573569.2019.1577778>
- Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2012). *Motivation in education: Theory, research, and applications*. Pearson Higher Ed.
- Schunk, D. H., & Mullen, C. A. (2012). Self-Efficacy as an Engaged Learner. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 219–235). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_10

- Seban, D., & Tavsanli, Ö. F. (2015). Children's Sense of Being a Writer: Identity Construction in Second Grade Writers Workshop. *International Electronic Journal of Elementary Education*, 7(2), 217–234.
- Sessions, L., Kang, M. O., & Womack, S. (2016). The Neglected “R”: Improving Writing Instruction Through iPad Apps. *TechTrends*, 60(3), 218–225. <https://doi.org/10.1007/s11528-016-0041-8>
- Skaftun, A. (2015). Leseopplæring og fagenes literacy. *Nordic Journal of Literacy Research*, 1(0). <https://doi.org/10.17585/njlr.v1.170>
- Skar, G. B., Graham, S., & Rijlaarsdam, G. (2022). Formative writing assessment for change – introduction to the special issue. *Assessment in Education: Principles, Policy & Practice*, 29(2), 121–126. <https://doi.org/10.1080/0969594X.2022.2089488>
- Slavin, R. E. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher*, 31(7), 15–21. <https://doi.org/10.3102/0013189X031007015>
- Snyders, C. S. B. (2014). ‘I Wish We Could Make Books All Day!’ An Observational Study of Kindergarten Children During Writing

- Workshop. *Early Childhood Education Journal*, 42(6), 405–414.
<https://doi.org/10.1007/s10643-013-0625-2>
- Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, 15(3), 335–352.
- Sturgess, J., Rodger, S., & Ozanne, A. (2002). A Review of the Use of Self-Report Assessment with Young Children. *British Journal of Occupational Therapy*, 65(3), 108–116.
<https://doi.org/10.1177/030802260206500302>
- Suri, H. (2020). Ethical Considerations of Conducting Systematic Reviews in Educational Research. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic Reviews in Educational Research* (pp. 41–54). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27602-7_3
- Tashakkori, A., & Teddlie, C. (2008). Quality of inferences in mixed methods research: Calling for an integrative framework. *Advances in Mixed Methods Research*, 53(7), 101–119.

- Teague, B. L., Smith, P. H., & Jiménez, R. T. (2010). Learning to Write in a Mexican School. *Journal of Language and Literacy Education*, 6(1), 1–19.
- Thomas, J., Harden, A., & Newman, M. (2012). Synthesis: Combining results. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 179–226). SAGE.
- Thomson, P. (2011). Coming to terms with ‘voice’. In G. Czerniawski & W. Kidd (Eds.), *The student voice handbook: Bridging the academic/practitioner divide* (pp. 19–30).
- Tønnessen, F. E. (2011). What are skills? Some fundamental reflections. *L1-Educational Studies in Language and Literature*, 149–158.
- Tønnessen, F. E., & Uppstad, P. H. (2015). *Can we read letters? Reflections on fundamental issues in reading and dyslexia research*. Sense Publishers.
- Toshalis, E., & Nakkula, M. J. (2012). *Toshalis, E., & Nakkula, M. J. (2012). Motivation, engagement, and student voice. Boston, MA: Jobs for the Future. Jobs for the Future.*
- Troia, G. A. (Ed.). (2009). *Instruction and assessment for struggling writers: Evidence-based practices*. Guilford Press.

- Troia, G. A. (2014). *Evidence-Based Practices for Writing Instruction (CEEDAR Document No. IC-5)*. Gainesville: Collaboration for Effective Educator, Development, Accountability, and Reform Center, University of Florida. Retrieved from https://cedar.education.ufl.edu/wp-content/uploads/2014/09/IC-5_FINAL_08-31-14.pdf.
- 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>
- Truax, M. L. (2018). The Impact of Teacher Language and Growth Mindset Feedback on Writing Motivation. *Literacy Research and Instruction*, 57(2), 135–157. <https://doi.org/10.1080/19388071.2017.1340529>
- Tunks, K. W. (2010). Attitudes Toward Writing in Fifth Grade Students Preparing for State-Mandated Writing Exam. *International Journal of Applied Educational Studies*, 8(1), 1–9.

- Tyler, B.-J., & Chard, D. (2000). Using Readers Theatre To Foster Fluency in Struggling Readers: A Twist on the Repeated Reading Strategy. *Reading and Writing Quarterly*, 16(2), 163–168.
- UNESCO (Ed.). (2017). *Fostering a culture of reading and writing: Examples of dynamic literate environments*. UNESCO Institute for Lifelong Learning.
- Unrau, N. J., Rueda, R., Son, E., Polanin, J. R., Lundeen, R. J., & Muraszewski, A. K. (2018). Can Reading Self-Efficacy Be Modified? A Meta-Analysis of the Impact of Interventions on Reading Self-Efficacy. *Review of Educational Research*, 88(2), 167–204. <https://doi.org/10.3102/0034654317743199>
- Uppstad, P. H. (2005). Optimality Theory: Rethinking phonology for the study of literacy. *Working Papers/Lund University, Department of Linguistics and Phonetics*, 51, 209–224.
- Van Valin Jr, R. D. (2003). Functional linguistics. In M. Aronoff & J. Rees-Miller (Eds.), *The handbook of linguistics* (pp. 319–336). Blackwell.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

- Walgermo, B. R. (2018). *Motivation for Reading within the First Year of Formal Reading Instruction* [Doctoral dissertation].
- Walgermo, B. R., & Uppstad, P. H. (2023). Enhancing Students' Identities as Readers and Writers through Assessment. In *Becoming Readers and Writers* (pp. 32–48). Routledge.
- Weiner, B. (1986). An Attributional Theory of Achievement Motivation and Emotion. In B. Weiner, *An Attributional Theory of Motivation and Emotion* (pp. 159–190). Springer US.
https://doi.org/10.1007/978-1-4612-4948-1_6
- Weiner, B. (2018). The legacy of an attribution approach to motivation and emotion: A no-crisis zone. *Motivation Science*, 4(1), 4–14.
<https://doi.org/10.1037/mot0000082>
- Whitaker, E. (2019). *UCSF Guides: Systematic Review: More about pros and cons of doing a systematic review*.
<https://guides.ucsf.edu/c.php?g=375744&p=3151000>
- Whitehurst, G. J. (2002). *Evidence-based Education (EBE)* [Archived presentation slides]. https://ies.ed.gov/director/pdf/2002_10.pdf

- Wigfield, A. (1997). Reading motivation: A domain-specific approach to motivation. *Educational Psychologist*, 32(2), 59–68.
https://doi.org/10.1207/s15326985ep3202_1
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
<https://doi.org/10.1006/ceps.1999.1015>
- Wijekumar, K., Graham, S., Harris, K. R., Lei, P.-W., Barkel, A., Aitken, A., Ray, A., & Houston, J. (2019). The roles of writing knowledge, motivation, strategic behaviors, and skills in predicting elementary students’ persuasive writing from source material. *Reading and Writing*, 32, 1431–1457.
- Witt, C. (2018). *Ways of Being: Potentiality and Actuality in Aristotle’s Metaphysics*. Cornell University Press.
- Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. Harper.
- Writing, n. (n.d.). In *OED Online*. Oxford University Press. Retrieved February 17, 2023, from <https://www.oed.com/view/Entry/230775>

- WWC. (n.d.). Retrieved June 8, 2023, from <https://ies.ed.gov/ncee/wwc/>
- Wyatt-Smith, C., & Jackson, C. (2016). NAPLAN data on writing: A picture of accelerating negative change. *Australian Journal of Language & Literacy*, 39(3), 233–244.
- Zachariadis, M., Scott, S., & Barrett, M. (2013). Methodological Implications of Critical Realism for Mixed-Methods Research. *MIS Quarterly*, 37(3), 855–879. JSTOR.
- Zimmerman, B. J. (2000a). Attaining Self-Regulation. In *Handbook of Self-Regulation* (pp. 13–39). Elsevier.
<https://doi.org/10.1016/B978-012109890-2/50031-7>
- Zimmerman, B. J. (2000b). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*, 25(1), 82–91.
<https://doi.org/10.1006/ceps.1999.1016>
- Zumbrunn, S., Ekholm, E., Stringer, J. K., McKnight, K., & DeBusk-Lane, M. (2017). Student Experiences With Writing: Taking the Temperature of the Classroom. *The Reading Teacher*, 70(6), 667–677. <https://doi.org/10.1002/trtr.1574>
- Zumbrunn, S., Marrs, S., Broda, M., Ekholm, E., DeBusk-Lane, M., & Jackson, L. (2019). Toward a More Complete Understanding of

References

Writing Enjoyment: A Mixed Methods Study of Elementary Students. *AERA Open*, 5(2), 233285841985079. <https://doi.org/10.1177/2332858419850792>

Appendices

Appendix 1 – Article 1

Alves-Wold, A., Walgermo, B. R., McTigue, E., & Uppstad, P. H. (under review). The ABCs of writing motivation: A systematic review of factors emerging from K-5 students' self-reports as influencing their motivation to write.

This paper is not included in the repository because it has not yet been published.

Appendix 2 – Article 2

Alves-Wold, A., Walgermo, B. R., McTigue, E., & Uppstad, P. H. (2023). Assessing Writing Motivation: A Systematic Review of K-5 Students' Self-Reports. *Educational Psychology Review*, 35(1), 24. <https://doi.org/10.1007/s10648-023-09732-6>



Assessing Writing Motivation: a Systematic Review of K-5 Students' Self-Reports

Aline Alves-Wold¹ · Bente Rigmor Walgermo¹ · Erin McTigue¹ · Per Henning Uppstad¹

Accepted: 28 November 2022 / Published online: 23 February 2023
© The Author(s) 2023

Abstract

For attaining success in writing, motivation is essential. Crucially, instruction is dependent on knowing the student's capabilities and inner drives. To date, research on writing has yet to establish a consistent framework for assessing writing motivation, and often fails to acknowledge students' self-reports, rather favoring evaluations of students' writing motivation made by others, such as teachers and researchers. This limbo state originates partly from a general skepticism towards the trustworthiness of elementary students' self-reports. Nonetheless, the validity of such self-reports has been acknowledged in adjacent fields, such as reading. Aiming to establish a knowledge base from studies addressing students' voices, the present study adopts the method of a systematic review and investigates how writing motivation has been assessed in empirical studies (1996–2020) through K-5 students' self-reports. Of the 7047 studies identified through database search, 56 met the inclusion criteria and are examined in this review. Results indicate that (a) *storytelling* is the genre most used to operationalize writing in the investigations, (b) surveys and interview questions measuring students' *attitude* towards writing are the most common type of self-report used, and (c) students' voices are weighted differently across the studies. Findings suggest that future research should (1) work to counteract existing biases in writing tasks, (2) provide a rationale for their choice/design of measure of motivation, and (3) report clearly whose voices are being heard (e.g., students', teachers', or researchers') and the appropriateness of this choice regarding study purpose, design, and findings.

Keywords Writing · Motivation · Elementary education · Systematic review · Measures · Self-report

✉ Aline Alves-Wold
aline.alves-wold@uis.no

Extended author information available on the last page of the article

Introduction

Successful instruction is dependent on knowing the students' capabilities. As the Danish philosopher Kierkegaard stated: "If one is truly to succeed in leading a person to a specific place, one must first and foremost take care to find him where he is and begin there. This is the secret in the entire art of helping" (Kierkegaard, 1859/1998, p. 45). Accordingly, to gain insight into individuals' inner states, it is essential to obtain their own perspectives; therefore, the students' voices must be heard. Facing 20–30 students in a classroom, teachers have limited possibilities of tapping their students' inner states through observation alone. Instead, teachers should aim at obtaining students' own perspectives as a primary source.

In assessment theory, the search for the optimal source of information is a crucial issue for validity. However, in research on writing motivation, discussions of optimal sources for obtaining information on students' motivation to write remain absent. For instance, recent reviews on writing motivation (Camacho et al., 2021; Ekholm et al., 2018), have analyzed and summarized findings regarding students' motivations to write without clearly delineating whether these findings were based on teachers' and researcher's evaluations of the children's inner states or formulated by the children themselves. Foregrounding the importance of listening to students' own perspectives to advance our understanding of writing motivation, we acknowledge that, in this quest, valid instruments to capture their voices are needed. In the present review, to contribute with knowledge on current uses of students' self-reports for assessing writing motivation, we analyze self-reports used in empirical studies, and provide an overview of how aspects of both writing and motivation were addressed in these self-reports, and how students' voices were emphasized in these studies.

Developing Writing Motivation in Early Elementary Settings

Before proceeding further, we must pause and address the consequences of learning to write in the early school grades. Hardly, a controversial point—there is disciplinary consensus that *proficiency* with the written word is essential for both school and life success, and early experiences with writing can predispose children to either seek out or avoid writing (NELP, 2008). Where the disciplinary consensus begins to falter is in the degree of recognizing the role of *motivation* in writing attainment. Like many current researchers (Boscolo & Gelati, 2018; Camacho et al., 2021; Ekholm et al., 2018; Klassen, 2002), we take the stance that developing writing skills requires much persistence, therefore writing research and instruction cannot only focus on *skills* but must also continually consider *motivation*. Specifically, to support young writers' motivation, we need to communicate value about writing and gain knowledge about their envisioned goals, interests, and self-beliefs. Yet, only recently, questions regarding the role of motivation for writing development and how to promote it through classroom

practices have begun to reach the center stage of writing research, which leads us back to the importance of finding out *where students are and meeting them there*. However, at this point, there is very little consensus on assessments of writing and of writing motivation in particular—especially for younger writers.

This overall lack of research about writing motivation in early stages of education is problematic when one considers the particular importance of initial writing experiences for motivation. In general, success builds beliefs in one's efficacy, while failure undermines such self-beliefs (Zimmerman, 2000). These mechanisms are particularly evident in early phases of skill development where failure typically occurs before a sense of efficacy has been firmly established (Bandura, 1995). This implies that children in their first years in school have writer self-beliefs that are particularly malleable and dynamic (Unrau et al., 2018). Consequently, the first years in school represent both great opportunities and potential threats to writing development.

Additionally, a second problematic trend is that although children often arrive at school with *intrinsic motivation* to write, as formal instruction progresses, students tend to shift in orientation to *extrinsic motivation*—such as grades (Boscolo & Gelati, 2007). Instruction that was attuned to motivation (i.e., informed by assessments of motivation) would ideally maintain or strengthen intrinsic motivation. However, as Troia et al. (2013) remark, unlike in reading research, there is no systematic research to document how, how much, or why writing motivation may diminish over time. We argue that this lack of knowledge is indicative of the fact that writing motivation is rarely assessed in schools. For instance, recent large-scale studies in England (Dockrell et al., 2016), the USA (Brindle et al., 2016), and the Netherlands (Rietdijk et al., 2018) have worked to document the common instructional practices for writing in elementary grades, and despite cataloging a wide variety of practices and assessments, none of the studies documented efforts to assess students' motivation.

Due to the limited research in writing motivation, we can consider the related field of reading research, for potential insight as to immediate needs and directions for writing motivation. For example, we now have meta-analyses that document bi-directional relationships between early reading skill and motivation (e.g., Toste et al., 2020), but this research is still to be expanded for writing before firm conclusions can be drawn. We also need to acknowledge that such advances in reading research have been obtained—to some extent—by the presence of validated models of assessment for elementary (e.g., MRQ, in Wigfield & Guthrie, 1997) and early childhood (e.g., SELM, McTigue et al., 2019).

Yet, recognizing that motivation is contextual (Troia et al., 2012), we cannot simply transpose knowledge from the domain of reading motivation to writing motivation. As early as first grade, attitudes towards writing form a unique construct compared to attitudes for reading (Graham et al., 2012a, 2012b). In fact, motivation to write may be even more important for literacy attainment than reading motivation because, simply put, writing is harder than reading because it is a production task demanding a complicated series of decisions and actions (Møller et al., 2022). As Bruning and Horn (2000, p. 26) aptly describe: “Students need to be motivated to enter, persist, and succeed in this ill-defined problem space we call writing.”

In line with such a domain-specific view of motivation, Troia et al., (2012, p. 7) have reviewed motivation research in the specific domain of writing and argue that *four broad components of motivation* have been identified: (1) *self-efficacy beliefs* (Bandura, 1986, 1994), (2) *goal orientations* (Dweck & Leggett, 1988; Elliott & Dweck, 1988; Harackiewicz et al., 2002), (3) *task interest* (Hidi, 1990; Hidi et al., 2002) and value (Eccles et al., 1983; Wigfield & Eccles, 2000), and (4) *outcome attributions* (Schunk, 1994; Weiner, 1986). In their review, a schema is also proposed, portraying the interrelationship between these four motivational components and associated constructs, such as domain self-concept and task utility (Troia et al., 2012, p. 11). However, although this schema portrays how motivation constructs are interrelated, Troia et al., (2012, p. 11) point out that some links, such as the causal pathways between self-efficacy, interest, and value, are still unclear. To help untangle these causal connections, those authors then invite researchers to test these connections by combining different research methods, such as classroom observation and students' self-reports (p. 18).

Quality of Self-Reports Measuring Writing Motivation

Over the past decade, there has been an enlarged focus on the documentation of the quality of assessments overall in education (e.g., Arnesen et al., 2019; Evers et al., 2013), but the positive effects of this focus have not appreciably impacted the assessment of writing motivation. Furthermore, being that motivation is an internal state of mind, assessments must include self-reports and not only rely on others' (e.g., teachers or parents) interpretations of behavior in order to most validly capture motivation. Yet, among existing approaches, there is a large variation regarding the extent to which students' self-reports are considered. In addition, as our focus includes writing in K-5 grades, there are additional challenges to consider when measuring young children's motivation (McTigue et al., 2019), as they may not be able to communicate their thoughts and feelings as well as older learners. Therefore, attempts at measurement are often compared with the ambition of hitting a moving target.

These circumstances touch upon central validity issues, and in particular *construct validity*, often referred to as the core of validity, concerned with measuring the construct in question as accurately as possible. Discussions on validity today, however, more commonly target stakeholders' (here both the researchers' and the teachers') interpretations and use of test scores (Kane, 2006, 2017; Lane et al., 2016, p. xv) in parallel with more common aspects of validity expressed in psychometric terms and standards (e.g., reliability). This practice embraces a unitary view of validity—in which no aspect of the validation process in principle is superior to the other. However, elaborating on the unitary view of validity, Kane (2013) states that potential scenarios for interpretation and use of test scores should be highlighted systematically even before the development of a test starts. That is, considering validity only during the interpretation phase is too late. In other words, valid test development starts by considering potential scenarios for the interpretation, using test scores, and evaluating limitations and weaknesses that may threaten valid interpretation and use of the test.

In the present study, the focus on students' voices acknowledges the appropriateness of argument-based validity in important ways. First, it values students' own perspectives on their motivational states as mostly relevant to both the use and interpretation of these measures. Second, it emphasizes the measurement itself, meaning that it must be understandable to its users, including teachers, students, parents, and policymakers.

Scope of the Present Review

Although different literature reviews about students' writing motivation have recently been published (Camacho et al., 2021; Klassen, 2002; Troia et al., 2012), to our knowledge, no study focused on the measures used to capture students' voices on this matter. Although Ekholm et al. (2018) address conceptualization and measurement issues in their review, they focus exclusively on writing attitudes, whereas we approach motivation more broadly. In addition, we bring attention to the necessity for children's voices to be given primacy and not their behaviors (which are interpreted through others) or the voices of adults. Indeed, as a crucial dimension of the United Nations Convention on the Rights of the Child (UNCRC), it is stated that if we want to know what is actually in the interest of the child, it is logical to listen to him or her (UNCRC, 1989; article 12:1). This emphasis on listening to the students' own perspectives is a radical position assuming that assessment of motivation to write should be built first and foremost on the students' voices—the optimal and primary source of information of their inner motivational drives.

First, we address the phrase “motivation to write,” acknowledging the impossibility to focus on *motivation* without also addressing what these drives are *directed towards*, namely *writing*. Walgermo et al. (2018) name the relation between motivation and the *target of motivation* in ecological terms as being symbiotic, and Walgermo and Uppstad (in press) state that “Unlike reading or writing skills, which are often studied for their own sake, motivation is a potential that is most typically investigated in relation to other potentials, like reading and writing.” In line with this remark, we first investigate the types of writing tasks addressed in self-reports measuring students' motivation to write.

Next, as remarked by Camacho et al., (2021, p. 234), an array of motivation-related constructs has often been presented in writing research without being explicitly defined, which, according to the authors, leads to “conceptualization issues and terminological overlaps.” However, we argue that not clearly defining motivational constructs in a study has consequences, not only for the conceptualization of these constructs but also for their measurement. That is, if a construct such as ‘self-concept’ is not clearly defined in a study, how can it be accurately measured to portray the students' voices? Thus, in the present review, we seek to identify what motivation constructs are measured and how they are operationalized in studies investigating early elementary students' motivation to write.

Finally, given the claimed inattention of what is the primary source of information when investigating students' motivation to write, there would likely be an

expectation of a large variance in how students' voices are actually valued and weighted across studies, so we investigate how much primacy students' voices are given in the identified studies.

Within this scope, we wish to identify strengths and weaknesses in current measurement to inform future research, and these areas of timely inquiry are addressed in the following three research questions:

- 1) What types of writing tasks are addressed in self-reports measuring students' motivation to write?
- 2) What motivation constructs are measured and how are they operationalized?
- 3) What emphasis is given to students' voices in the studies?

Method

The present study has adopted the methodology of a systematic review (Gough et al., 2012), which is usually conducted through five steps: (1) framing the research question(s) that will guide the review (as presented in the previous section), (2) identifying relevant work through systematic literature search and pre-defined selection criteria, (3) assessing the quality of the studies identified, (4) summarizing the evidence from the selected studies, and (5) discussing the findings (Khan et al., 2003). Where applicable, the present review follows guidelines from the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) (Moher et al., 2009).

Identifying Studies

Rationale for the Review Timeframe

The present review includes studies that were published between January 1, 1996 and April 1, 2020. The year of 1996 has been chosen as the starting point of the review for two reasons. First, according to Hidi and Boscolo (2008, p. 144), the impressive body of research on motivation that advanced during the 1980s only impacted writing research much later, when writing researchers demonstrated that writing is a complex task requiring not only the coordination and development of cognitive skills, but also *affective components*. Hayes' revised framework, published in 1996, reflects this new conceptualization of writing, where affective components are given a much more prominent role. Second, as proposed by Alexander and Fox (2004, p. 50), the period from 1996 onwards is seen as the "Era of Engaged Learning," a period characterized by a shift in the way the literacy community perceived and investigated learners and learning, and in which researchers began to look at motivational components, such as goals and interests, as critical factors for learning development. Thus, given the significance of the year 1996 for writing motivation research in educational settings, it has been chosen as the starting point for the current review. This being said, any cut-off dates are likely to be more or less indicative

rather than rigid, as the educational research literature tend to have few clear-cut joints. We acknowledge that there were studies on writing motivation before 1996—as will be the case for any time frames set—but the chosen year is here supported by an indication that something culturally changed, which marked a “key turning point” in writing motivation research.

Systematic Literature Search

A thorough search of the literature was conducted using four different databases: ERIC, Academic Search Premier, PsycINFO (1806–present), and Web of Science. The scope of this review—as formulated in the RQs above—addresses three main topics, namely: *writing*, *motivation*, and *early elementary education* (including K-5 grades). From each of these three topics, related search terms were added, providing a total of 49 search terms (see Fig. 1). Figure 1 portrays the area of interest for the present study in visually representing the intercept of the three main topics—with related terms—addressed in the present review.

The initial literature search returned 12,839 records. Thereafter, depending on their availability within each database, limiters matching some of the inclusion criteria discussed below were applied, and a total of 7047 studies were retrieved for screening. These studies were then exported to EPPI-Reviewer, a software tool for research synthesis, where 1252 duplicates were removed. This resulted in a total of 5795 studies that moved to the screening stage of this review, as summarized in Table 1.

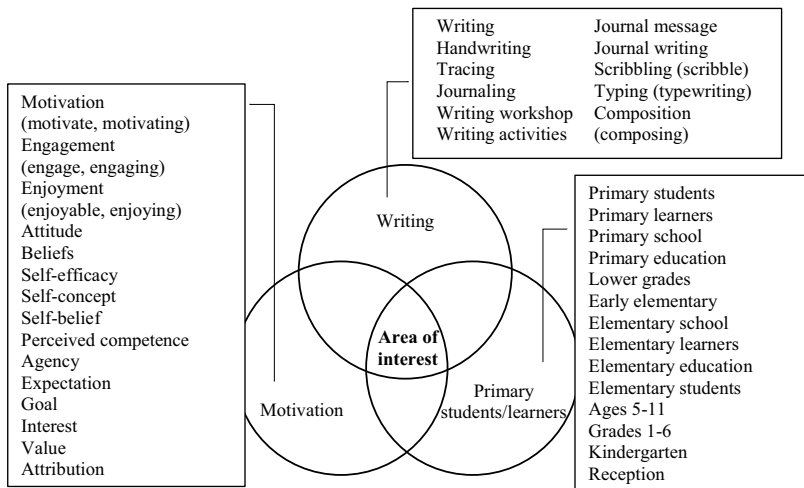


Fig. 1 Diagram of search terms clusters

Table 1 Total of records retrieved for screening

Database	Initial search	Limiters applied	Records retrieved for screening
PsycINFO (1806–present)	1244	Peer-reviewed, 1996–2020	551
ERIC	7419	Peer-reviewed, 1996–2020	2766
Academic Search Premier	1882	Scholarly, 1996–2020	1542
Web of Science	2294	1996–2020	2188
	12,839		7047
		Duplicates excluded	1252
		Total for screening	5795

Selection Criteria

In EPPI-Reviewer, the remaining 5795 studies were screened manually, first on title and abstract, and then on full text. Following the lead of Miller et al., (2018, p. 89), the inclusionary criteria for selecting studies in both phases were divided into four categories as shown below:

- a) Publication: articles were written in English and published between January 1996 and April 2020
- b) Research: studies were empirical and peer-reviewed
- c) Topic: studies focused on students’ writing motivation
- d) Participants: studies focused on K-5 students in mainstream classrooms (studies that focused solely on second language learners or students with special needs were excluded)

In phase 1 of the screening process, 5795 studies were screened on title and abstract and based on these criteria, 5434 studies were excluded in this phase. Then in phase 2, 361 studies were screened on full text; and based on the same criteria, 267 studies were excluded in this phase, which resulted in a total of 94 studies. Similarly to Hakimi et al. (2021), these stages were carried out by the first author, but the other authors were available to discuss abstracts and titles that were ambiguous or presented uncertainty.

As the process of a systematic literature review (SLR) is recursive, it allows researchers to adjust the procedures to maintain a focus on the research questions. At this stage, we noticed that in some studies, writing motivation was not directly investigated; it was rather used as a post hoc explanation of why students behaved in specific ways. Consequently, to ensure that the selected studies were directly focused on writing motivation, and the students’ perspectives were included, a third screening phase was added, incorporating the following two eligibility criteria: (1) the study has to include at least one research question about writing motivation (either explicitly stated in question format, or implicitly in terms of goals

and purposes of the study), and (2) the study has to include at least one type of *student self-report* on writing motivation (e.g., survey, questionnaire, interview).

Articles meeting these criteria, but solely focusing on instrument development and validity were excluded. In this work, we are synthesizing research regarding students' motivation in the classroom (i.e., the focus of our research questions is on classroom research), and validation research is focused more narrowly on establishing credibility of a measure which is fundamentally different. Student's motivation will be measured in such work but only for the purpose of establishing reliability and validity of scales. In addition, to avoid overemphasis of particular studies, in cases where multiple articles reported on the same data, the latest version of the study was kept, and previous duplicate studies were excluded (e.g., Li & Chu, 2018).

In phase 3 of the screening process, 94 studies were screened on full text. Based on the eligibility criteria for this phase, 44 studies were excluded, which resulted in a total of 50 studies to be included in the corpus for this review. To ensure the reliability of this screening step, a random selection of 27% of the articles (i.e., 26 out of 94 articles) was double coded by the second author for eligibility. Interrater agreement was 92% (24 out of 26 articles), and 100% after discussion.

Finally, three different hand-search procedures were conducted. First, we conducted *backward snowballing*, where we hand-searched the reference lists of all 50 studies included in the review and identified two additional studies. Of these, only one could be retrieved, which resulted in 51 studies. We tried to retrieve the other study by contacting the authors who conducted the study, the authors who referenced the study, and the journal who published the study, but to no avail. Second, in trying to identify newer studies, we conducted *forward snowballing*. For this search step, we used both Scopus and Google Scholar to identify all the papers who had cited the 51 included studies in our review, and screened these references as described by Wohlin (2014). This search step led to the inclusion of two additional studies, totaling 53 studies. Third, we hand-searched the reference lists of six relevant reviews/meta-analyses (Camacho et al., 2021; Ekholm et al., 2018; Graham & Perin, 2007; Graham et al., 2022, 2012a; 2012b; Troia et al., 2012) and identified three additional studies that met our criteria, which resulted in a total of 56 articles included in the corpus for the present review. Backward and forward snowballing was again conducted for these additional studies, but new studies meeting our inclusion criteria were not identified. Figure S1 in the supplementary material provides numerical totals for each phase of the screening process.

Assessing the Quality of the Studies

Before synthesizing results in a systematic review, it is common to assess the quality of these studies and exclude from review those that do not meet pre-defined quality criteria. However, according to Pawson et al., (2005, p. 30), “to synthesize *is* to make sense of the different contributions”, thus, the quality of studies should be “*established in synthesis* and not as a preliminary pre-qualification exercise” (originally italicized). In the present review, the quality of the studies is therefore assessed

Table 2 Criteria for assessing the studies' methodological quality

Methodological characteristic	Scoring options (Maximum total score = 14 points)	Distribution of characteristics among 56 reviewed studies	
		Frequency (n)	Percent (%)
Explicates theory and/or previous research in a way that builds the formulation of the posed question(s)/purpose(s)/objective(s) that can be investigated empirically	Yes = 2 points	37	66
	Partially = 1 point	19	34
	No = 0 point	0	0
	Quantitative or qualitative methods = 1 point	32	57
Research method	Mixed methods = 2 points	24	43
	Narrow sample = 1 point	7	13
Sample size	Small sample (< 100) = 2 points	27	48
	Medium sample (> 100 and < 300) = 3 points	17	30
	Large sample (> 300) = 4 points	5	9
	Explicitly = 2 points	19	34
Refers to/uses relevant theory to justify choice/design of motivation measure	Implicitly = 1 point	27	48
	Does not refer to relevant theory = 0 point	10	18
Characteristics/evidence of validity, reliability, credibility, and/or trustworthiness are addressed and reported	Reported = 2 points	51	91
	Not reported = 0 point	5	9
Findings and conclusions are legitimate or consistent with data collected	Yes = 2 points	51	91
	No = 0 point	5	9

as part of the synthesis, rather than used as a screening step for excluding studies from the review.

To assess the quality of the studies, a methodological quality score (MQS), adapted from Goodson et al., (2006, p. 313) and Miller et al., (2018, p. 90), was used to evaluate each study. Table 2 shows the six criteria used in the assessment, and their full score leads to a maximum of 14 points per study. Since ternary categories that are divided among *yes*, *partially*, and *no* award 2 points for *yes* and 0 point for *no*, the same scores of 2 for *yes* and 0 for *no* is also used for binary categories to maintain the balance in the scores.

Coding and Analysis of Eligible Studies

Following the lead of Reed et al. (2014), the coding process was carried out through three stages. In the first stage, a coding sheet was developed and refined by the four authors. In the second stage, the first author coded all studies included in the corpus ($n = 56$), and in the last stage, the second author double coded all 56 studies and discrepancies were resolved.

In stage 1 of the coding process, based on our research questions, all four authors met in a conference call to develop a spreadsheet divided into predefined categories (e.g., number of participants, type of self-report, motivation construct investigated, type of writing task) where the 56 included studies could be coded deductively. After initial categories were agreed upon, the first author coded five of the studies and met again with the other three authors to refine the coding scheme. The final version of the spreadsheet was organized into four main categories: (1) characteristics of the studies, (2) quality of the studies, (3) measures of writing motivation, and (4) factors affecting writing motivation. In the current review, we present results regarding the first three categories (for a detailed discussion of factors affecting writing motivation, see Alves-Wold et al.–in review).

During the second stage of the coding process, the first author coded all studies included in the review ($n = 56$), and all authors were available to discuss peculiarities of the studies and particularities of the coding. Finally, in the last stage, the second author double coded all 56 studies, and any discrepancies between the two researchers' scores were resolved through a second review, discussion of discrepancies, and a finalized consensus. Throughout the process, in addition to informal meetings among the authors, all four authors met at least monthly on a conference call to discuss particularities of the study and make consensual decisions regarding each stage of the review process.

Characteristics of the Studies

Studies were coded for the following eight characteristics: name of scientific journal, year of publication, country where the study was carried out, number of participants, participants' grade-level(s), research method, whether the study was an intervention, and a summary of the main findings (see Table S1 in the Supplementary Material for an overview).

Quality of the Studies

Notes were taken regarding the quality of the studies and scores were assigned for each study according to the six categories described in Table 2.

Measures

We coded the studies for type of writing task, type of student self-report, and details about the measures used in each study, such as if the measure was administered in a group or individually, or if images were used for support. In addition, we coded for type of triangulation data for studies that combined student self-report with other types of data.

Results and Discussion

In this section, we synthesize results and findings related to the research questions guiding this review: (1) What types of writing tasks are addressed in self-reports measuring students' motivation to write? (2) What motivation constructs are measured and how are they operationalized? (3) What emphasis is given to students' voices in the studies? Before presenting results related to these, we describe general characteristics of the studies, followed by an overview of their quality.

Characteristics of the Studies

Period

There is a significant growth in the number of studies published on writing motivation in early elementary education, as almost 70% ($n=38$) of the 56 reviewed studies were published in the last decade, rather than between 1996 and 2009 ($n=18$).

Place

Almost 60% of the studies were conducted in North America, with the USA contributing the most: USA ($n=28$), Canada ($n=4$), and Mexico ($n=1$). Asia was the second most represented continent with 28% of the studies: Turkey ($n=9$), China ($n=2$), Singapore ($n=2$), Indonesia, Jordan, and Taiwan (all $n=1$). Europe contributed with 11% of the studies: Finland ($n=2$), Cyprus, Italy, Portugal, and Sweden (all $n=1$). Finally, Oceania contributed only one study originating from Australia. Consistent with the review conducted by Camacho et al. (2021), the present review did not identify any studies from Africa nor South America. Given that one of the inclusionary criteria for this review requires that studies need to be written in English, it is not surprising that more than 50% of the studies are from English-speaking countries; however, none of the studies originated from the UK.

Publication

The studies were published in 44 different peer-reviewed journals, with the most represented journals being *Early Childhood Education Journal* ($n=4$), *Reading Psychology*, *Reading & Writing Quarterly* (both $n=3$), *Education, Elementary School Journal*, *Reading Improvement*, *Reading Horizons*, and *International Electronic Journal of Elementary Education* (all $n=2$).

Participants

Grade levels were investigated individually in 70% of the studies ($n=39$). Most of these ($n=28$) focused on the upper grade levels: 5th ($n=13$), 4th ($n=9$), and 3rd ($n=6$), whereas only less than a third ($n=11$) of these studies focused on the lower levels: Kindergarten ($n=6$), 1st ($n=2$), and 2nd ($n=3$). More than one grade level was investigated in 30% of the studies ($n=17$). Most of these studies ($n=12$) focused on grade pairs, with the most common combinations being 2nd and 3rd ($n=3$) and 4th and 5th ($n=3$). Only one study included participants from all six grades.

Quality of the Studies

MQS values were awarded to each study, as described in Table 2. Scores ranged from 5 to 13 points (maximum possible = 14), and the mean, median, and mode values were very similar to each other (mean = 10.23, median = 11, and mode = 11). Almost 75% of the studies received a score higher than 70% of the MQS, and only two studies received a score that was lower than 50%. Table 2 shows the frequency distributions for each category of the MQS, and additional comments regarding these categories are presented below.

Research Method

Approximately 45% of the studies ($n=25$) used a mixed-methods design, instead of exclusively quantitative ($n=18$) or qualitative methods ($n=13$). It is logical that researchers triangulate multiple types of data sources when dealing with complex affective constructs or measuring both writing skills and motivation constructs.

Sample Size

The number of participants included in each study varied significantly. Sample sizes spanned from qualitative investigations of one (Leroy, 2000) or two students (Abbott, 2000; Andrzejczak et al., 2005; Perry et al., 2003) to a quantitative investigation of the writing disposition of 2315 fourth and fifth graders (Unal, 2010). One of the studies did not specify the exact number of participants (Lee & Enciso, 2017), but according to our categories (narrow, small, medium, and large), we coded this study as investigating a large sample, as it included a sample of 29 classrooms.

As some of the studies investigating multiple grade levels provided only the total number of participants, we could not differentiate the total of participants per grade level. However, a total of approximately 8000 participants are investigated in the reviewed studies.

Theoretical Foundation of Studies and Rationale for Design of Motivation Measure

Most studies ($n=37$) presented relevant theory and previous research addressing both writing and motivation. Theories and models cited by authors include self-determination theory (e.g., Ryan & Deci, 2000), social cognitive theory (e.g., Bandura, 1986, 1997), social constructivist theory (e.g., Vygotsky, 1978), model of social self-interaction (e.g., Schunk, 1999), attribution theory (e.g., Weiner, 1986), and self-theories (e.g., Dweck & Master, 2009). Authors also cited relevant literature to address writing skills (e.g., Hayes, 1996) and to define motivation constructs (e.g., self-efficacy: Bandura, 1977). Nevertheless, more than 30% of the reviewed studies ($n=19$) tend to present theory and research that focus mainly on writing but lack relevant references to motivation theory and/or research. That is, even though these studies include research questions that investigate motivational constructs, relevant motivation theory is not used to explain what these constructs entail. In addition, in studies where both writing and reading were investigated, authors tend to refer mostly to reading research.

Only a third of the studies ($n=19$) explicitly referred to motivation theory in their methods section to explain the rationale for the design of the chosen motivation measure. Some referred to motivation theory to explain the choice of previously used measures, as Nicolaidou (2012) who argues that the Writer Self-Perception Scale was chosen because “it was grounded on Bandura’s (1997) theory of perceived self-efficacy.” Whereas others referred to motivation theory to justify the content of the measure developed for the study, as Liao et al. (2018) who explained that “the content of the WIQ [Writing Interest Questionnaire] was developed based on the four-phase model of interest development (Hidi & Renninger, 2006; Schraw & Lehman, 2001).”

Nevertheless, approximately half of the studies ($n=27$) only referred to motivation theory in the theory section of their articles, without explicitly linking the significance of the presented theory for the design of their motivational measures. In addition, ten of the studies did not refer to motivation theory to justify their choice of motivation measure, neither in the theory nor the methods sections of their articles.

Given that motivation is a multi-dimensional and dynamic construct, not clearly defining which constructs that are being investigated and how they are being operationalized in the chosen instruments might have consequences for the validity of the measure, as it becomes unclear which motivational components that are being measured and how. In other words, theoretical clarity can provide a needed link between a complex construct (i.e., motivation) and the measurement of that construct. To promote standardization in the definition of motivation constructs, Camacho et al., (2021, p. 224) provide a list of definitions that can be helpful for researchers to bring clarity to their investigations of motivation constructs.

Validity, Reliability, Credibility, and/or Trustworthiness

Authors addressed issues related to the trustworthiness of the studies in 91% of the cases ($n = 51$); however, the quality of the evidence reported by the authors varied among the studies. For instance, whereas some authors gave detailed accounts of the methods used to ensure the integrity of their investigations, others briefly addressed these issues under the Limitations section of their articles.

Out of 36 studies that included quantitative measures of motivation, such as scales, 22 reported evidence of reliability, and 14 of validity. In most of the studies, Cronbach's alpha internal consistency coefficient was reported as evidence of reliability, and confirmatory factor analysis was used for reporting evidence of validity. In four of the studies, authors referred instead to previous validation studies of the measures used, which might be insufficient, as although an instrument is reliable/valid for one sample, it may not be for another.

Out of 38 studies that included qualitative measures of motivation, such as interviews and open-ended questions, 35 provided evidence of credibility and trustworthiness. Most of the studies provided detailed information about the context, the participants, and the procedures used for collecting, coding, and analyzing the data.

It is important to notice that the number of studies employing quantitative and qualitative measures discussed here should not be totaled, as these methods are combined in mixed-methods studies.

Legitimacy of Findings and Conclusions

In 91% of the studies ($n = 51$), findings and conclusions showed consistency with the data collected. However, in 9% of the studies ($n = 5$), part of the conclusions was not clearly linked to the results. This discrepancy was mainly due to a lack of clear distinctions regarding the sources of the results, which impacts the validity of the findings. For instance, in some of the studies including responses from both teachers and students, it was not always clear which group had uttered the responses discussed. When conclusions are presented, it is therefore difficult to ascertain how much they represent the students' voices.

What Types of Writing Tasks Are Addressed in Self-Reports Measuring Students' Motivation to Write?

Given that the focus of our work is on motivation to *write*, researchers cannot consider this without operationalizing writing, so we start with how writing is addressed in the studies.

Some Studies Directly Measure Writing Skill and Others Do Not

Whereas all studies are measuring aspects of motivation, only a subset of studies also directly investigate writing skill (33 of the 56). In the studies that directly consider

writing skills, measurements of writing quality were used, and these included teacher-reported evaluations (e.g., Perry, 1998; Zumbrunn et al., 2019), researcher judgements (e.g., Hall et al., 2017; Teague et al., 2010), and students' self-assessments (e.g., Bradford et al., 2016; Nicolaidou, 2012). Quality levels were most commonly judged by the use of rubrics that set standards for *skill-related aspects of writing*, such as text organization (e.g., Boyacı & Güner, 2018), ideation (e.g., Schrodt et al., 2019), spelling accuracy (e.g., Jones et al., 2016), length of composition (e.g., Liao et al., 2018), and audience-awareness (e.g., Gallini & Zhang, 1997). Typically, multiple aspects of writing are investigated in synchrony (e.g., a writing sample that is analyzed from multiple lenses), and even seemingly simple writing tasks like spelling require the judgement of multiple aspects, like handwriting legibility, directionality (words are written from left to right), and spelling accuracy.

Although writing skill measures are not the primary focus of the present review, we note the value of these quality assessments in the studies' investigations of students' writing motivation. For instance, by measuring students' writing quality, researchers are able to compare motivation levels between high- and low-achieving students (e.g., Perry et al., 2003), check how calibrated students' self-efficacy beliefs are in comparison to their actual performance (e.g., Kim & Lorschbach, 2005), or investigate if changes in performance affect motivation levels (e.g., Li & Chu, 2018).

Types of Writing Tasks Addressed in the Studies

Various motivation constructs, to be discussed in the next section, were measured in the reviewed studies either in relation to writing as a *general task* (e.g., "what makes you a good writer?" from Kim & Lorschbach, 2005) or as *specific tasks* (e.g., writing a story, revising a text). Of the various types of writing tasks used in the studies, narrative/story writing ($n=17$) was clearly the most common genre investigated. Studies also included genres such as expository ($n=5$), informational/explanatory ($n=2$), and descriptive writing ($n=1$). However, it is difficult to give a precise number of the specific types of tasks investigated in the studies, as researchers used very different nomenclatures and categorizations of writing tasks. For instance, some authors used specific terms such as "letter writing" (e.g., Chohan, 2011) in their description of the investigated tasks, whereas others used more general terms such as "authentic writing tasks" (e.g., Boyacı & Güner, 2018), or "various writing activities" (e.g., Paquette, 2008) to denote them. Other types of writing tasks included, for example, spelling activities (Jones et al., 2016), linguistic games (Boscolo et al., 2012), poetry writing (Andrzejczak et al., 2005), collaborative writing (Li & Chu, 2018), and high- and low-challenge writing tasks (Miller & Meece, 1999).

In 24 of the studies, researchers indicated whether writing tasks or surveys were performed on paper ($n=12$), digitally ($n=6$), or both ($n=6$); however, in 32 of the studies, the technology used was not specified.

Writing tasks were also investigated in intervention studies ($n=33$) measuring the students' levels of motivation in relation to specific teaching practices ($n=25$), such as using artwork as a pre-writing activity (Andrzejczak et al., 2005), participating in writing workshops (Hertz & Heydenberk, 1997, Pollington et al., 2001), or through a

drama-based program (Lee & Enciso, 2017). In eight of these studies, motivation was also investigated in relation to the use of different digital tools such as mobile apps (Kanala et al., 2013; Sessions et al., 2016) and online blogging (Nair et al., 2013). In one study, researchers measured motivation to write before and after a common teaching practice, that is, instruction that prepares students for a state-mandated writing exam (Tunks, 2010).

Items and Responses

Most often, students were asked to answer questions about their writing motivation before and after executing writing tasks (e.g., Babayigit, 2019; Beck & Fetherston, 2003; Boscolo et al., 2012; Hier & Mahony, 2018; Ihmeideh, 2015; Liao et al., 2018), or only after working with them (e.g., Jones et al., 2016; Kim & Lorsbach, 2005; Miller & Meece, 1997, 1999; Wilson & Trainin, 2007). However, students also answered questions about writing motivation without being asked to perform specific writing activities (e.g., Hall & Axelrod, 2014; Mata, 2011; Merisuo-Storm, 2006; Unal, 2010; Zumbrunn et al., 2019). In addition, students were asked to answer questions about writing tasks that differed from those they were asked to execute. For instance, in Akyol and Aktas (2018), students were asked to write a *story*, but their survey included questions about *narrative*, *expository*, and *general* writing. Table 3 provides some examples of how writing was operationalized as either a general or specific task in both open- and close-ended questions.

Does the Type of Writing Addressed Matter?

As shown, the results indicate a split distribution regarding general or more specific writing tasks. Interestingly, when specific writing tasks are targeted in measures of motivation to write, storytelling/narrative writing is by far the most common genre investigated. However, although story writing is traditional in school, this choice is somewhat curious for motivation research when students are given the opportunity to rate their motivation for other genres. For instance, when given the opportunity to explore and produce their own informational texts, students' confidence and interest in this type of text increased (Hall et al., 2017). This difference in the occurrence of writing genres addressed in the studies, combined with findings showing that choice of task will have an influence on motivation levels (Alves-Wold et al.—in review), suggest a present bias concerning wider aspects of writing. From an assessment perspective, this bias may represent construct under-representation (Cook et al., 1979), a challenge that should be addressed when taking the first steps towards a coherent framework for measuring motivation to write.

A question that needs further research is whether this overweight of storytelling tasks in measures of motivation to write is mirroring either (a) researchers' oversimplifications of writing, where storytelling represents writing in general, or (b) storytelling is a task over-represented in classroom practices. Findings from the present study indicate that when motivation to write is measured, often a too narrow approach to writing is taken. This means that in order to develop valid self-report instruments for writing motivation, we need measures to target writing tasks that reflect a wider variety of classroom practices.

Table 3 Examples of how writing was operationalized as either a general or specific task in items and responses

Writing operationalization	Open-ended		Close-ended			
	Sample item	Sample response	Sample item	Sample response		
General task	Erdogan & Erdogan, (2013)	<p>“Writing is like... because...”</p> <p>Upon completion of the second posttest, students wrote a final reflection essay responding to the prompt “In your opinion, is it best to use a rubric when writing or not? Why?”</p>	<p>“Writing is like a book, because like books my writings give other people information.”</p> <p>- Assists in remembering to do all the things and check work</p> <p>- Helps you get 4 stars (achieve all the requirements)</p>	<p>Akyol and Aktas (2018)– Adapted from Codling and Gambrel (1997)</p> <p>Nair et al. (2013)</p>	<p>Knowing how to write well is ...</p> <p>I think learning how to blog is important</p>	<p>- Not important</p> <p>- Kind of important</p> <p>- Important</p> <p>- Very important</p> <p>SA = strongly agree; A = agree; N = neutral; D = disagree; SD = strongly disagree</p>
Specific task	Bradford et al. (2016)					

What Motivation Constructs Are Measured and How Are they Operationalized?

Motivation Constructs

Most of the studies ($n=36$) investigated one motivational construct, but in 20 studies, two or more motivational constructs were investigated simultaneously. Following the authors' terminology in each study, a total of 32 motivational constructs were identified. In Table 4, to systematize the variety of constructs identified in the studies, we categorized these constructs according to *theorized components* of motivation and *associated constructs*, as described in the work of Troia et al. (2012). Column 1 provides frequency numbers for each identified construct, and column 1 provides total numbers for the identified constructs in each category. Frequency numbers for associated constructs show that *attitude* was clearly investigated most often ($n=20$), followed by *self-efficacy* ($n=11$), and *motivation* ($n=9$). These frequency numbers also correspond with total numbers for each category, showing that the theorized component of *interest and value* was investigated most often ($n=39$), followed by *self-beliefs* ($n=28$), and *motivation* ($n=11$).

Types of Self-Reports

Self-reports were used 83 times in the 56 reviewed studies to measure the above-mentioned motivational constructs. Although authors have used different nomenclature, the types of self-reports fall into three main categories: (1) interviews, (2) surveys and questionnaires, and (3) alternative written responses. *Student interviews* were used 32 times and included discussion sessions, student-teacher conferences, portfolio-based conferences, and individual interviews where students answered orally a questionnaire that included ratings from 1 to 10, and also open-ended

Table 4 Sorting of motivation constructs

Theorized components	Associated constructs
Motivation ($n=11$)	Motivation ($n=9$), intrinsic motivation ($n=1$), and achievement motivation ($n=1$)
Self-beliefs ^a ($n=28$)	Self-efficacy ($n=11$), self-concept ($n=3$), perceptions of themselves as writers ($n=4$), self-perception ($n=2$), self-perception of competence ($n=1$), confidence ($n=1$), anxiety ($n=1$), performance expectancies ($n=1$), beliefs ($n=1$), writer identity ($n=1$), perceived competence ($n=1$), outcome expectation ($n=1$)
Goal-orientations ($n=2$)	Goal orientations ($n=1$), cognitive engagement ($n=1$)
Interest and value ($n=39$)	Attitude ($n=20$), interest ($n=5$), value ($n=3$), enjoyment ($n=3$), task value ($n=2$), perception/belief about writing ($n=2$), liking ($n=1$), writing disposition ($n=1$), aversion ($n=1$), perception of teacher writing enjoyment ($n=1$)
Outcome attributions ($n=5$)	Effort ($n=1$), control ($n=1$), support ($n=1$), growth mindset ($n=1$), attributions ($n=1$)

^aGiven that *self-efficacy beliefs* are task-oriented, we use the broader cover term *self-beliefs* to represent this broad component of motivation and its associated constructs

questions where they could give reasons for their ratings (Miller & Meece, 1999). *Surveys and questionnaires* were used 46 times, and it is noteworthy that often the terms *surveys and questionnaires* were used interchangeably. In the reviewed studies, both surveys and questionnaires were often described as a set of written questions/statements (items) where students could indicate their responses by choosing one of the options in a Likert scale. In ten of these studies, surveys also included open-ended questions. Finally, *alternative written responses* were investigated five times, through qualitative analyses of the narrative portion of students' feedback on exit slips (Truax, 2018), the students' final reflection essay on the use of rubrics (Bradford et al., 2016), the students' drawings and written description of a recent writing experience (Zumbrunn et al., 2017), and the students' completion of the metaphorical sentence "Writing is like... because..." (Erdoğan & Erdoğan, 2013).

Self-reports were administered both individually ($n=23$) and in a group ($n=37$), but in 23 of the identified self-reports, this distinction was not specified in the studies' methodological descriptions. Surveys and questionnaires were mainly administered in whole-class groups ($n=31$), whereas interviews were often administered individually ($n=20$) or in focus-groups ($n=2$). Although most of the student interviews were conducted by the researchers at school, some were also conducted in the students' homes (e.g., Abbott, 2000).

Adapting Previously Used Self-Reports

Out of the 83 identified self-reports, authors chose to employ previously used motivational measures 41 times, either without modifying them ($n=10$), or by adapting them to the needs of the studies ($n=31$), whereas authors developed new measures for their investigations 42 times. In less than 15% of the times when interviews were used ($n=7$), authors indicated that interview guides were developed based on previously used questions. In contrast, authors often used surveys and questionnaires that had previously been used ($n=34$), either without modifying them ($n=10$), or with modifications ($n=24$). With regards to alternative written responses, authors did not indicate whether they were adaptations of previously used motivational measures.

Self-report modifications included linguistic adaptations, such as translations (e.g., Babayigit, 2019; Nicolaidou, 2012) and changes in the wording of items to account for age-adequate content (e.g., Hier & Mahony, 2018). In addition, items were also modified from different domains to be applicable to writing, for instance, by adapting items from reading to writing (e.g., Kanala et al., 2013). However, in some cases, it is not always clear what these modifications entail, for example, when authors state that part of the items from a specific scale is used, but do not specify which items (e.g., Göçen, 2019).

Items and Responses

A variety of items were used in the studies to measure motivation constructs. Items included both statements and questions, and students could provide their responses by answering open-ended questions or through marking options in close-ended questions.

Open-ended questions were most commonly used in student interviews and often asked students about their self-beliefs (e.g., “What do you do really well as a writer?” from Hillyer & Ley, 1996) and preferences (e.g., “Are there any things in particular you like to write about?” from Nolen, 2007). In surveys, open-ended questions were often used to ask students to provide reasons to their answers in close-ended questions (e.g., Silver & Lee, 2007).

Close-ended scales varied from two-tiered frequency responses divided between “always” and “usually” (Wilson & Trainin, 2007) to 10-point Likert scales ranging from “1 = not very sure” to “10 = very sure” (Miller & Meece, 1999). Some surveys consistently used the same close-ended responses for all of the items in the survey (e.g., Bayat, 2016). However, some surveys combined different response scales, depending on the items. For instance, in Akyol and Aktas (2018), although most of the items are judged by the students through a 4-point Likert scale, the responses include both frequency scales from “almost never” to “always,” and evaluation scales from “a poor author” to “a very good author,” as well as its inverted sequencing from “very good” to “poor.”

Given the young age of the students being investigated, in 19 of the studies researchers also chose to use stuffed animals or images to support the students in their responses. Stuffed animals were used three times for different purposes. In Mata (2011), two different stuffed animals represented contradictory statements, and kindergarten children could choose if they were “a little” or “a lot like” the chosen stuffed animal for each item, which resulted in a 4-point Likert scale. In Nolen (2007), a monkey hand puppet was used to ask children in first grade to describe reading and writing in school, and in Schrodt et al. (2019), it is not specified how the stuffed animals were used in the study. When it comes to images, although Garfield, the cat was used most often ($n=6$) (e.g., Paquette, 2008), a variety of other symbolic images was also used in all grades to represent different responses in Likert scales. These included, for instance, dogs (e.g., Paquette et al., 2013), teddy bear faces (Merisuo-Storm, 2006), boxes (e.g., Hier & Mahony, 2018), and happy and sad faces (Jones et al., 2016). In Table 5, we present some examples of the items used in the studies to measure the five theorized components of motivation, and samples of both open- and close-ended responses.

Why is Validation of Self-Reports of Motivation to Write Important?

The results above show that new unvalidated self-report measures are used extensively in the research targeting writing motivation in early elementary education. Interestingly, out of the 83 identified self-report measures identified in the present review, previously used motivational measures were used 41 times, either with or without modifications. This means that 42 times, authors developed new measures for their studies. This preponderance of newly developed instruments documented in the present review is symptomatic of the claimed limbo state of writing motivation research, and therefore worthy of attention and reflections. According to Haladyna et al. (2002), new measures represent the most common validity threats in the field of assessment. Although validity in testing refers to the accurate and meaningful interpretation of test scores and the reasonableness of the inferences drawn from test

Table 5 Examples of motivation constructs addressed in items and responses

Theorized component	Open-ended		Close-ended	
	Sample item	Sample response	Sample item	Sample response
Motivation	Abbott (2000) Why do you choose to write?	"I don't know. I just like to write. I know if you write enough, then you'll get better and better. [...]"	Li and Chu (2018) "I want to spend more time in writing because of using "Joyous Writing Club."	5-point Likert scale with 1 = strongly disagree and 5 = strongly agree
Self-beliefs	Nicolaidou (2012) Do you think your essays in your portfolio show progress?	Yes, I believe that they do	Grenner et al. (2020) - I can quickly write a text on the computer - I can read through my text and correct spelling mistakes - I can read through my text and make changes to improve it	Beneath each statement was a 100-mm, visual-analogue scale. The VAS scale was marked with the phrases "not at all" and "yes, completely" below the left and right endpoints. One of the researchers explained how students should mark the scales according to their beliefs, with illustrations on the whiteboard

Table 5 (continued)

Theorized component	Open-ended		Close-ended	
	Sample item	Sample response	Sample item	Sample response
Goal-orientations	Hillyer and Ley (1996) What would you like to do better as a writer?	"I need to improve on using quotation marks correctly in my stories. I'm not sure how to use them when I let my characters have conversation."	Miller and Meece (1997) <i>Task-mastery</i> item example: "I wanted to learn as much as possible on this assignment." <i>Ego-social</i> item example: "I wanted the teacher to think I am doing a good job on this assignment."	Students rated on a 4-point scale according to how well the statement described how they completed an assignment (1 = not at all to 4 = very much)
Interest and value	Chen and Liu (2019) Think and describe your attitude towards. Does your writing attitude change before and after the use of the 4S approach? Why or why not?	"Before using the 4S approach, I initially thought that writing and composition were both difficult for me. Those two tasks are interesting and now I am more willing than before to share stories with my classmates. In addition, the story structures exhibited in the tasks are clear and they are useful when I am writing the story."	Babayigit (2019)–Adapted from Graham, Berninger and Fan (2007) "How do you feel" ... "about writing for fun at home", "when you write in school during free time", "about writing during summer vacation", "about writing instead of playing", "about writing in school", "about spending free time writing", and "when it's time for writing at school."	Students indicated their attitude by marking one of four images of Garfield the Cat, ranging from very happy (score of 4) to very unhappy (score of 1)

Table 5 (continued)

Theorized component	Open-ended		Close-ended	
	Sample item	Sample response	Sample item	Sample response
Outcome attributions	Truax (2018)	<p>“When I work hard at my writing, my writing gets better”</p> <p>Strongly disagree = 1, Disagree = 2, Agree = 3, Strongly agree = 4</p> <p>Explain why you think that</p>	<p>Hier and Mahony (2018)</p> <p>“Did you work as hard as possible?”</p> <p>“Did you try to do a good job writing your story?”</p>	<p>Students responded to items on a five-point, Likert-type scale. Response options were presented as boxes with the words not at all, a little bit, some, a lot, and very, very much under each one. The box above each response option became successively larger from not at all through very, very much to denote the relative level of agreement each option represented</p>

scores (AERA et al., 2014; Messick, 1998), most threats for valid interpretations relate directly to the quality of the constructed measures (Cook et al., 1979). That is, with a preponderance of poorly crafted tests, misinterpretations and inappropriate uses of test score data are likely. The use of unvalidated measures also seemed evident in studies investigating motivation to write using student interviews and alternative written responses, as they often did not present a rationale for the construction of these measures.

However, although we see the value of creativity and innovation, new measures should be accompanied by pilot testing and discussions of validity, and in failing to do so, instead they threaten to hamper scientific progress. In total, we recognize that creativity is truly needed in assessments of writing motivation but should be combined with sound measurement development to form a solid basis for this evolving field of research. To do this, it should be mandatory that researchers define the motivation constructs explored in their studies, as remarked by Camacho et al. (2021), and make explicit their rationale for choice/design of measure of motivation. Only then, can measures be scrutinized by the research community and end users, for scientific purposes or classroom uses.

Does It Matter What Kind of Pictorial Support We Use in Self-Report Measures?

The format of scales designed for children is an important area to consider, to confirm we are on the right track when measuring something. Specifically, we must acknowledge the role of pictorial support—e.g., typical scales including faces with different moods—as this format is most frequently applied to elicit student thinking in the studies included in the present review. Pictorial support is a recommended practice when measuring young children’s motivation (Bandura, 2006), as an aid to capture their perspectives, especially if they lack the necessary vocabulary to express themselves. This broad guideline is a relatively uncontroversial recommendation in the field, but exactly what type of pictorial support to use is far less unanimous.

More precisely, a further finding in the current study is that a myriad of pictorial supports was used, including stuffed animals and picture-scales of different cats and dogs, and in most cases objectively constructed out of convenience. This diversity is somewhat surprising in the field of motivation, as proponents of this field are likely to have a particular sensitivity to how nuances in external factors may trigger humans’ inner drives differently. Returning to the focus on validity, the spectrum of different varieties of pictorial support revealed in this study can be seen as uncontrolled variables which may erroneously inflate or deflate scores for some or all examinees. Such uncontrolled variables can lead to construct-irrelevant variance—reducing the accuracy of test score interpretations and thereby the validity of the test in question (Cook et al., 1979; Messick, 1998).

The array of unvalidated pictorial supports is particularly problematic, given the early advice of Bandura (2006, p. 313) to use *circles* with progressively larger sizes for representing students’ increasing confidence that they can perform the tasks addressed in the measures. In this, Bandura clearly stated that *happy or sad faces* in self-efficacy scales should be avoided, because children may misread such scales as measuring their emotional states rather than how confident they are that they can

perform given tasks. Bandura's caution about the use of pictorial scales is based on a clear rationale, and ought to be validated as a starting point for investigating the convenience of other ways to provide pictorial support for young students. None of the studies included in the present review refers to the guidelines of Bandura (2006) or other relevant guidelines for their use of pictorial support in scale construction.

What Emphasis Is Given to Students' Voices in the Studies?

Students' Self-Reports Versus Reports from Others

Although every study in this review included at least one measure of motivation based on *students' self-reports*, the degree of emphasis placed on the students' voices when reporting results and findings varied considerably among the studies. Mainly, this difference was a consequence of the type of data used in the studies. That is, students' voices were given a predominant role in some studies as they only included students' self-reports as their primary data (e.g., Mata, 2011; Seban & Tavsanli, 2015). However, other studies combined the students' self-reports with reports from teachers (e.g., Li & Chu, 2018), researchers (e.g., Hall et al., 2017), and parents (e.g., Teague et al., 2010), which would juxtapose the students' voices with the viewpoints of others.

In studies where data from students' self-reports were triangulated with teachers' and/or researchers' observations/evaluations, it was sometimes unclear what contribution each source had for the findings, as they were not clearly presented in the results. For instance, although Hertz and Heydenberk (1997, p. 205) state that "informal and formal assessments, observations of students' writing process behaviors, and parent, teacher, and student interviews" were used for data collection, these are not presented separately in the results. When findings are discussed, it is therefore difficult to pinpoint how much the students' own viewpoints contributed to the studies' findings.

In addition, whereas some studies provide examples of items and students' responses (e.g., Erdoğan & Erdoğan, 2013; Snyders, 2014), others provide only examples of the questions asked to students, without explicitly providing examples of what the students actually answered, which again makes it difficult to gain a better insight of the students' perspectives. For instance, in Lee and Enciso (2017), the open questions "What is good writing?" and "What does a good writer do?" were included in their survey, but examples of student answers were not provided.

In other cases, even if results are presented separately, the students' responses may sometimes play a smaller role in the findings than the teachers' evaluations. For example, after analyzing data from both teachers' and students' responses, Chohan (2011, p. 39) recommends the implementation of a schoolwide mailing program, arguing that "data analysis indicated that children enjoyed the responsive letter writing process." However, although qualitative results from teachers' evaluations suggested that this intervention had contributed positively to students' writing motivation, quantitative data from student surveys indicated otherwise.

Measuring More Than One Domain

Determining the students' perspectives is also important in studies measuring other domains, in addition to writing. For instance, in seven of the included studies, reading motivation was also investigated, and in one of the studies (Hall et al., 2017), it was difficult to ascertain how interested students were in writing, as most of the information on interest derived from students' self-reports were about reading, while results regarding motivation to write seemed to be mostly derived from observations and reports from teachers and parents. In addition, looking at the results derived from all 56 studies, students seem to have high levels of motivation to write. However, in five of the seven studies that measured both reading and writing motivation, students demonstrated higher levels of motivation towards reading, rather than writing, which could indicate issues related to conformity in the self-reports. That is, when only asked questions about writing, students might provide positive answers as they think this is what is expected of them, whereas when asked about different domains, they might feel "freer" to be honest about each domain separately. At the same time, when asked about their motivations towards different domains, students might also feel that choosing their favorite is necessary, which might influence their choices. Whether measuring only one domain or more than one thus needs deliberate consideration.

Combining Multiple Sources of Student Self-Reports

Even in studies that only included students' self-reports, these were sometimes derived from different data sources, such as multiple surveys (e.g., Göçen, 2019; Seban, 2012) or surveys and interviews (e.g., Kholisiyah et al., 2018; Nicolaidou, 2012; Perry, 1998), and the role that each data source played in the studies also varied. For instance, Truax (2018) combined quantitative data from students' surveys with qualitative data from students' interviews, and although quantitative results indicated that teacher's growth mindset feedback had no statistically significant effects, qualitative findings suggested that objective feedback positively impacted writing motivation.

Combining quantitative and qualitative methods was commonly used for complementary purposes, where quantitative methods, such as scales, were used to measure *frequencies* and *levels*, that is, *how often* students worked with writing activities and *how much* they enjoyed or valued these activities. Qualitative methods alone, such as open-ended questions in interviews and surveys, were used to investigate *reasons* for their choices and *factors* that influenced those choices. However, although some studies only investigated frequencies and levels, such as how much students were motivated to write, often in the studies' discussion section, authors also argued about reasons and factors for the students' answers. For instance, after analyzing quantitative data from three different scales, Göçen (2019) concludes that creative writing activities had a positive effect on students' creative writing achievement, writing attitude, and motivation. The author then suggests in the implications section of the article that "it is necessary to adopt a process-based writing approach and conduct creative writing activities in teaching mother tongue in order to enable students to

advance their writing skills, develop positive attitudes towards writing and enjoy writing” (Göçen, 2019, p. 1038). Nevertheless, the scales used in the study only measured *whether* motivation levels increased, rather than *why* they would increase. In such cases, questions about reasons and factors should be explicitly asked.

Finally, studies also combined different sample sizes, depending on the type of data collected. Commonly, larger groups responded to surveys, and smaller (mostly purposeful) samples participated in interviews (e.g., Bayraktar, 2013). For example, in Nair et al. (2013), a larger sample of 197 students responded to a survey, whereas a smaller purposeful sample of 12 students who did not submit their online assignments, and 6 who did it, participated in interviews to give reasons that affected their motivation to finish and submit (or not) their assignments.

Presenting Results from Each Data Source and Reporting Findings

In sum, the present analysis shows in what ways students’ self-reports were weighted differentially across studies. That is, students’ voices were given more prominence when students’ self-reports represented the primary data source (e.g., Mata, 2011). However, students’ perspectives were often given less prominence when combined with other data sources, such as reports from teachers (e.g., Li & Chu, 2018), researchers (e.g., Hall et al., 2017), and parents (e.g., Teague et al., 2010). Therefore, future studies should be aware when presenting results on students’ motivation from different data sources, i.e., publications should clearly state *who* has uttered *what* and further on, from *whose voices* the conclusions are drawn. Frameworks for guiding the integration of mixed methods data provide guidance in this area (Fetters et al., 2013). For clarity purposes, although analyses combine and synthesize different data sources, data for each source should be presented separately, and examples of items/questions *and* responses should be provided alongside the publication.

Limitations

The present review has four main limitations. First, even though comprehensive database and hand searches were conducted to identify relevant studies, we acknowledge that not all available studies were captured in this review as a result of stringent inclusion criteria. For instance, only including studies published in English give more prominence to research derived from specific countries and may overlook relevant advancements in the assessment of writing motivation originating from other parts of the world. In addition, to lessen limitations related to the quality of included studies, like Camacho et al. (2021) and Miller et al. (2018), the present review only includes peer-reviewed studies, as these studies have undergone the rigorous demands of the peer-review process, which is generally accepted “as a method to ensure a level of academic credibility” (Miller et al., 2015, p. 467). However, even though relevant as a quality threshold, this criterion can be a source of bias. For instance, as documented by Polanin et al. (2016), the presence of publication bias in education and psychology can lead to skewed effect sizes in meta-analyses. Nevertheless, publication bias does not represent the same type of threat in the present

review, given that the way the RQs of our review are formulated—partly resembling the ones of a scoping review instead of a meta-analysis—points rather to sampling in order to achieve saturation. As such, by purposively sampling studies from peer-reviewed journals that purportedly reflect research trends and standards in writing motivation, we argue that the studies included in our review provide the variability necessary for investigating the RQs addressed in our study.

Second, we acknowledge that ending the search in April 2020 excludes studies published after the occurrence of the COVID-19 pandemics, which may have influenced students' motivation and perhaps how writing motivation was measured during lockdown periods. We acknowledge this limitation and encourage researchers to explore whether such a change has occurred, which would be an interesting starting point for future reviews; however, this question is beyond the scope of the present study.

Third, inconsistencies of construct definitions in writing motivation coupled with the use of a diverse range of self-report practices limited our ability to synthesize findings towards specific standards for designing writing motivation self-reports. However, we provide recommendations that can serve as initial suggestions towards more standardized practices.

Finally, we focus specifically on self-reports used with K-5 students, but it is possible that a systematic review investigating older students could reveal more uniform practices and guidelines, which in turn could potentially inform the design of self-reports for younger students. However, to the best of our knowledge, no such reviews have been conducted, hence we could not compare our findings with those for older students.

Conclusion and Future Directions

Particularly in the study of writing motivation, researchers must be sensitized to the target of the motivation—the writing task itself. The greater construct of “writing” represents an array of genres and formats, and each type is endowed with potential challenges and joys. Therefore, how writing is presented can greatly impact how a young writer responds. We must not assume that because a young writer is unmotivated to compose a creative story that they are unmotivated towards the larger construct of “writing.” This is akin to assuming that, for example, if a student did not want to play a game of dodgeball they did not like athletics. In the reviewed studies, an over-simplification or narrow operationalization of writing represented a significant validity risk to the study of writing motivation. However, this also provides an important opportunity to course-correct and strengthen the rigor of research in this area.

Findings from the present study show extensive—some would even say reckless—use of unvalidated self-report measures of motivation to write in the early grades. New measures should be solidly anchored in theory of motivation and writing, and systematically piloted before implementation in classrooms or for research purposes. In particular, pictorial support in measures of writing motivation should follow existing guidelines for scale construction (e.g., Bandura, 2006).

Furthermore, in our quest for better understanding students' motivation to write, listening to their own perspectives on this matter is an important step towards building this understanding. However, the role that this rich source of data plays when analyzing findings needs to be clearly delineated. That is, especially in studies combining the students' views with the perspectives of others, such as teachers and parents, researchers need to report clearly whose perspectives are being portrayed and what role they play in the studies' findings. In addition, to strengthen transparency of how the students' reports are evaluated, researchers should provide examples, not only of the *items* asked, but also of *responses* uttered by the students.

The conditions highlighted above all pose threats to validity in different ways, as validity is defined as "The degree to which evidence and theory support the interpretation of test scores for proposed uses of tests" (AERA et al., 2014, p. 1). The emphasis on interpretation and use underscores that validity exceeds the boundaries of the test itself (i.e., psychometric qualities), rendering a situation where validity in the groundbreaking sense is dependent on intended and appropriate use of test scores. The lack of a consistent framework of constructs of motivation to write, as shown in this study, will continue to pose validity threats until solved.

Over a long period of time, assessment researchers have emphasized challenges of reliability and psychometrics over challenges posed by construct-underrepresentation. The lack of attention towards the different writing genres used when investigating motivation to write therefore represents a clear construct-underrepresentation in current research base on motivation to write. Argument-based validity (Kane, 1992, 2006) emphasizes the importance of starting validation from aspects of interpretation and use. An obvious starting point would be to listen to the students' own voices: "If we want to know what is actually in the interest of the child, it is logical to listen to him or her" (UNCRC, 1989; article 12:1). According to established definitions, this is also the most likely place from which constructs should be "constructed," as constructs are theoretical and not directly observable (Thorndike & Thorndike-Christ, 2014, p. 135). This is also why Chappelle (2020) claims that argument-based validation is an intended means for bridging the gap between theory and practice.

Our findings may provide some initial suggestions for approaching the limbo state of the field of motivation to write, as addressed above. Future research and development should (1) work to counteract existing biases in writing tasks, (2) provide a rationale for their choice/design of measure of motivation, and explicitly state what the chosen instrument is measuring, and (3) report clearly whose voices are being heard (e.g., students', teachers', or researchers') and the appropriateness of this choice regarding study purpose, design, and findings.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-023-09732-6>.

Open access funding provided by University of Stavanger & Stavanger University Hospital

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

*Studies marked with an asterisk were included in the present review.

- *Abbott, J. A. (2000). "Blinking out" and "having the touch": Two fifth-grade boys talk about flow experiences in writing. *Written Communication*, 17(1), 53–92. <https://doi.org/10.1177/0741088300017001003>
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- *Akyol, H., & Aktas, N. (2018). The relationship between fourth-grade primary school students' story-writing skills and their motivation to write. *Universal Journal of Educational Research*, 6(12), 2772–2779. <https://doi.org/10.13189/ujer.2018.061211>
- Alexander, P. A., & Fox, E. (2004). A historical perspective on reading research and practice. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical Models and Processes of Reading* (5th ed., Vol. 1, pp. 33–68). DE: International Reading Association. <https://doi.org/10.1598/0872075028.2>
- *Andrzejczak, N., Trainin, G., & Poldberg, M. (2005). From image to text: Using images in the writing process. *International Journal of Education & the Arts*, 6(12), 1–17.
- Arnesen, A., Braeken, J., Ogden, T., & Melby-Lervåg, M. (2019). Assessing children's social functioning and reading proficiency: A systematic review of the quality of educational assessment instruments used in Norwegian elementary schools. *Scandinavian Journal of Educational Research*, 63(3), 465–490.
- *Babayigit, O. (2019). Examining the effect of creative writing activities on reading, writing and language lesson attitudes of elementary school fourth grade students. *European Journal of Educational Research*, 8(1). <https://doi.org/10.12973/eu-jer.8.1.213>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. *Self-Efficacy in Changing Societies*, 15, 334.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, 5(1), 307–337.
- Bandura, A. (1994). Self-efficacy. In V.S. Ramachaudran (Ed.) *Encyclopedia of Human Behavior* (Vol. 4, pp. 71–81). New York: Academic Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- *Bayat, S. (2016). The effectiveness of the creative writing instruction program based on speaking activities (CWIPSA). *International Electronic Journal of Elementary Education*, 8(4), 617–628.
- *Bayraktar, A. (2013). Nature of interactions during teacher-student writing conferences, revisiting the potential effects of self-efficacy beliefs. *Eurasian Journal of Educational Research*, 13(50), 63–86.
- *Beck, N., & Fetherston, T. (2003). The effects of incorporating a word processor into a year three writing program. *Information Technology in Childhood Education Annual*, 2003(1), 139–161.
- Boscolo, P., & Gelati, C. (2007). Best practices in promoting motivation for writing. *Best practices in writing instruction*, 202–221.

- Boscolo, P., & Gelati, C. (2018). Motivating writers. In S. Graham, C. A. MacArthur, & M. Hebert (Eds.), *Best Practices in Writing Instruction, Third Edition*. New York, UNITED STATES: Guilford Publications.
- *Boscolo, P., Gelati, C., & Galvan, N. (2012). Teaching elementary school students to play with meanings and genre. *Reading & Writing Quarterly*, 28(1), 29–50. <https://doi.org/10.1080/10573569.2012.632730>
- *Boyacı, S. D. B., & Güner, M. (2018). The impact of authentic material use on development of the reading comprehension, writing skills and motivation in language course. *International Journal of Instruction*, 11(2), 351–368. <https://doi.org/10.12973/iji.2018.11224a>
- *Bradford, K. L., Newland, A. C., Rule, A. C., & Montgomery, S. E. (2016). Rubrics as a tool in writing instruction: Effects on the opinion essays of first and second graders. *Early Childhood Education Journal*, 44(5), 463–472. <https://doi.org/10.1007/s10643-015-0727-0>
- Brindle, M., Graham, S., Harris, K. R., & Hebert, M. (2016). Third and fourth grade teacher's classroom practices in writing: A national survey. *Reading and Writing*, 29(5), 929–954. <https://doi.org/10.1007/s11145-015-9604-x>
- Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychologist*, 35(1), 25–37. https://doi.org/10.1207/S15326985EP3501_4
- Camacho, A., Alves, R. 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>
- Chapelle, C. A. (2020). *Argument-based validation in testing and assessment*. Sage Publications.
- *Chen, Z.-H., & Liu, W.-Y. (2019). A six-stage story structure approach for elementary students' story production: Quality, interest, and attitude. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2019.1688354>
- *Chohan, S. K. (2011). Any letter for me? Relationships between an elementary school letter writing program and student attitudes, literacy achievement, and friendship culture. *Early Childhood Education Journal*, 39(1), 39–50. <https://doi.org/10.1007/s10643-010-0438-5>
- Codling, R. M., & Gambrell, L. B. (1997). The motivation to write profile: an assessment tool for elementary teachers. National Reading Research Center, Universities of Georgia and Maryland Instructional Resource, 38
- Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings* (Vol. 351). Houghton Mifflin.
- Dockrell, J. E., Marshall, C. R., & Wyse, D. (2016). Teachers' reported practices for teaching writing in England. *Reading and Writing*, 29(3), 409–434. <https://doi.org/10.1007/s11145-015-9605-9>
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256. <https://doi.org/10.1037/0033-295X.95.2.256>
- Dweck, C. S., & Master, A. (2009). Self-theories and motivation: Students' beliefs about intelligence. In K. R. Wenzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 123–140). Routledge/Taylor & Francis Group.
- Eccles, J. S., Adler, T., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco: W. H. Freeman.
- Ekholm, E., Zumbunn, 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>
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54(1), 5–12. <https://doi.org/10.1037/0022-3514.54.1.5>
- *Erdoğan, T., & Erdoğan, Ö. (2013). A metaphor analysis of the fifth grade students' perceptions about writing. *The Asia-Pacific Education Researcher*, 22(4), 347–355. <https://doi.org/10.1007/s40299-012-0014-4>
- Evers, A., Muñiz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25(3), 283–291.
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. *Health Services Research*, 48, 2134–2156.
- *Gallini, J. K., & Zhang, Y.-L. (1997). Socio-cognitive constructs and characteristics of classroom communities: An exploration of relationships. *Journal of Educational Computing Research*, 17(4), 321–339. <https://doi.org/10.2190/Y71J-T3EJ-VFTJ-P6WG>

- *Göçen, G. (2019). The effect of creative writing activities on elementary school students' creative writing achievement, writing attitude and motivation. *Journal of Language and Linguistic Studies*, 15(3), 1032–1044.
- Goodson, P., Bui, E. R., & Dunsmore, S. C. (2006). Self-esteem and adolescent sexual behaviors, attitudes, and intentions: A systematic review. *Journal of Adolescent Health*, 38(3), 310–319. <https://doi.org/10.1016/j.jadohealth.2005.05.026>
- Gough, D., Oliver, S., & Thomas, J. (2012). Introducing systematic reviews. In D. Gough, S. Oliver, and J. Thomas (Eds.), *An introduction to systematic reviews* (pp. 1–16). Thousand Oaks, CA: Sage.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- 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(Emphasis>(3), 516–536. <https://doi.org/10.1016/j.cedpsych.2007.01.002>
- Graham, S., McKeown, D., Kihara, S., & Harris, K. R. (2012b). 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., Tavsanlı, O. F., & Kaldırım, A. (2022). Improving writing skills of students in Turkey: A meta-analysis of writing interventions. *Educational Psychology Review*, 34(2), 889–934. <https://doi.org/10.1007/s10648-021-09639-0>
- *Graham, S., Berninger, V., & Abbott, R. (2012a). Are attitudes toward writing and reading separable constructs? A study with primary grade children. *Reading & Writing Quarterly*, 28(1), 51–69. <https://doi.org/10.1080/10573569.2012.632732>
- *Grenner, E., Johansson, V., van de Weijer, J., & Sahlén, B. (2020). Effects of intervention on self-efficacy and text quality in elementary school students' narrative writing. *Logopedics Phoniatrics Vocology*, 1–10. <https://doi.org/10.1080/14015439.2019.1709539>
- Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The ethics of using digital trace data in education: A thematic review of the research landscape. *Review of Educational Research*, 91(5), 671–717. <https://doi.org/10.3102/00346543211020116>
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333.
- *Hall, A. H., & Axelrod, Y. (2014). “I am kind of a good writer and kind of not”: Examining students' writing attitudes. *Journal of Research in Education*, 24(2), 34–50.
- *Hall, A. H., Boyer, D. M., & Beschoner, E. A. (2017). Examining kindergarten students' use of and interest in informational text. *Early Childhood Education Journal*, 45(5), 703–711.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, 94(3), 638–645. <https://doi.org/10.1037/0022-0663.94.3.638>
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Erlbaum.
- *Hertz, M., & Heydenberk, W. (1997). A kindergarten writing workshop: How kindergarten students grow as writers. *Reading Horizons*, 37(3), 203–214.
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 60(4), 549–571.
- Hidi, S., & Boscolo, P. (2008). Motivation and writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144–157). Guilford Press.
- 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
- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: An intervention study. *Learning and Instruction*, 12(4), 429–446. [https://doi.org/10.1016/S0959-4752\(01\)00009-3](https://doi.org/10.1016/S0959-4752(01)00009-3)
- *Hier, B. O., & Mahony, K. E. (2018). The contribution of mastery experiences, performance feedback, and task effort to elementary-aged students' self-efficacy in writing. *School Psychology Quarterly*, 33(3), 408–418. <https://doi.org/10.1037/spq0000226>
- *Hillyer, J., & Ley, T. C. (1996). Portfolios and second graders' self-assessments of their development as writers. *Reading Improvement*, 33(3), 148–159.

- *Ihmeideh, F. (2015). The impact of dramatic play centre on promoting the development of children's early writing skills. *European Early Childhood Education Research Journal*, 23(2), 250–263. <https://doi.org/10.1080/1350293X.2014.970848>
- *Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2016). Beyond the rainbow: Retrieval practice leads to better spelling than does rainbow writing. *Educational Psychology Review*, 28(2), 385–400. <https://doi.org/10.1007/s10648-015-9330-6>
- *Kanala, S., Nousiainen, T., & Kankaanranta, M. (2013). Using a mobile application to support children's writing motivation. *Interactive Technology and Smart Education*, 10(1), 4–14. <https://doi.org/10.1108/17415651311326419>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527.
- Kane, M. T. (2006). Content-related validity evidence in test development. *Handbook of Test Development*, 1, 131–153.
- Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.
- Kane, M. T. (2017). Using empirical results to validate performance standards. In *Standard Setting in Education* (pp. 11–29). Springer, Cham.
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3), 118–121.
- *Kholisiyah, R. N., Rukayah, & Indriayu, M. (2018). Achievement motivation analysis of outstanding students in learning writing at primary schools. *International Journal of Educational Methodology*, 4(3), 133–139. <https://doi.org/10.12973/ijem.4.3.133>
- Kierkegaard, S. (1998). The point of view: Kierkegaard's writings (Vol. 22) (V. Howard & Edna H. Hong, Trans.) Princeton University Press. (Original work published 1859)
- *Kim, J.-A., & Lorsbach, A. W. (2005). Writing self-efficacy in young children: Issues for the early grades environment. *Learning Environments Research*, 8(2), 157–175. <https://doi.org/10.1007/s10984-005-7248-5>
- Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational Psychology Review*, 31.
- Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). Preface. In S. Lane, M. R. Raymond, and T. M. Haladyna. *Handbook of test development* (Vol. 2) (pp. xv-xvi). New York, NY: Routledge.
- *Lee, B. K., & Enciso, P. (2017). The big glamorous monster (or Lady Gaga's adventures at sea): Improving student writing through dramatic approaches in schools. *Journal of Literacy Research*, 49(2), 157–180. <https://doi.org/10.1177/1086296X17699856>
- *Leroy, C. (2000). "Why we do this is important": An inner-city girl's attitude toward reading and writing in the classroom. *Reading Horizons*, 41(2), 69–92.
- *Li, X., & Chu, S. K. W. (2018). Using design-based research methodology to develop a pedagogy for teaching and learning of Chinese writing with wiki among Chinese upper primary school students. *Computers & Education*, 126, 359–375. <https://doi.org/10.1016/j.compedu.2018.06.009>
- *Liao, C. C. Y., Chang, W.-C., & Chan, T.-W. (2018). The effects of participation, performance, and interest in a game-based writing environment. *Journal of Computer Assisted Learning*, 34(3), 211–222. <https://doi.org/10.1111/jcal.12233>
- *Mata, L. (2011). Motivation for reading and writing in kindergarten children. *Reading Psychology*, 32(3), 272–299. <https://doi.org/10.1080/02702711.2010.545268>
- McTigue, E. M., Solheim, O. J., Walgermo, B., Frijters, J., & Foldnes, N. (2019). How can we determine students' motivation for reading before formal instruction? Results from a self-beliefs and interest scale validation. *Early Childhood Research Quarterly*, 48, 122–133. <https://doi.org/10.1016/j.ecresq.2018.12.013>
- *Merisuo-Storm, T. (2006). Girls and boys like to read and write different texts. *Scandinavian Journal of Educational Research*, 50(2), 111–125. <https://doi.org/10.1080/00313830600576039>
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45(1), 35–44.
- Miller, D. M., McTigue, E. M., & Scott, C. E. (2015). The quality of recent studies in content-area writing in secondary classrooms. *Literacy Research: Theory, Method, and Practice*, 64(1), 461–477. <https://doi.org/10.1177/2381336915617602>
- Miller, D. M., Scott, C. E., & McTigue, E. M. (2018). Writing in the secondary-level disciplines: A systematic review of context, cognition, and content. *Educational Psychology Review*, 30(1), 83–120. <https://doi.org/10.1007/s10648-016-9393-z>
- *Miller, S. D., & Meece, J. L. (1997). Enhancing elementary students' motivation to read and write: A classroom intervention study. *The Journal of Educational Research*, 90(5), 286–299. <https://doi.org/10.1080/00220671.1997.10544585>

- *Miller, S. D., & Meece, J. L. (1999). Third graders' motivational preferences for reading and writing tasks. *Elementary School Journal*, 100(1), 19.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Møller, H. L., Mortensen, J. O., & Elbro, C. (2022). Effects of integrated spelling in phonics instruction for at-risk children in kindergarten. *Reading & Writing Quarterly*, 38(1), 67–82.
- *Nair, S. S., Tay, L. Y., & Koh, J. H. L. (2013). Students' motivation and teachers' teaching practices towards the use of blogs for writing of online journals. *Educational Media International*, 50(2), 108–119. <https://doi.org/10.1080/09523987.2013.795351>
- National Early Literacy Panel. (2008). Developing early literacy: Report of the national early literacy panel. Washington, DC: National Institute for Literacy. Available at: <https://www.nifl.gov/earlychildhood/NELP/NELPreport.htmlGoogleScholar>.
- *Nicolaidou, I. (2012). Can process portfolios affect students' writing self-efficacy? *International Journal of Educational Research*, 56, 10–22. <https://doi.org/10.1016/j.ijer.2012.08.002>
- *Nolen, S. B. (2007). Young children's motivation to read and write: Development in social contexts. *Cognition and Instruction*, 25(2–3), 219–270. <https://doi.org/10.1080/07370000701301174>
- *Paquette, K. R., Byrd, C. M., & Dragich, D. (2013). An investigation of implementing drop everything and write (DEW) in early childhood elementary classrooms. *Education*, 133(3), 356–369.
- *Paquette, K. R. (2008). Study of elementary students' attitudes about writing after a cross-age tutoring experience. *Reading Improvement*, 45(4), 181–191.
- Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review—A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10(1_suppl), 21–34. <https://doi.org/10.1258/1355819054308530>
- *Perry, N. E., Nordby, C. J., & VandeKamp, K. O. (2003). Promoting self-regulated reading and writing at home and school. *Elementary School Journal*, 103(4), 317–338.
- *Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology*, 90(4), 715–729.
- Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, 86(1), 207–236. <https://doi.org/10.3102/0034654315582067>
- *Pollington, M. F., Wilcox, B., & Morrison, T. G. (2001). Self-perception in writing: The effects of writing workshop and traditional instruction on intermediate grade students. *Reading Psychology*, 22(4), 249–265. <https://doi.org/10.1080/02702710127640>
- Reed, D. K., Cummings, K. D., Schaper, A., & Biancarosa, G. (2014). Assessment fidelity in reading intervention research: A synthesis of the literature. *Review of Educational Research*, 84(2), 275–321. <https://doi.org/10.3102/003465431452131>
- Rietdijk, S., van Weijen, D., Janssen, T., van den Bergh, H., & Rijlaarsdam, G. (2018). Teaching writing in primary education: Classroom practice, time, teachers' beliefs and skills. *Journal of Educational Psychology*, 110(5), 640–663. <https://doi.org/10.1037/edu0000237>
- 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. <https://doi.org/10.1037/0003-066X.55.1.68>
- Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review*, 13(1), 23–52. <https://doi.org/10.1023/A:1009004801455>
- *Schrodt, K. E., Elleman, A. M., FitzPatrick, E. R., Hasty, M. M., Kim, J. K., Tharp, T. J., & Rector, H. (2019). An examination of mindset instruction, self-regulation, and writer's workshop on kindergarteners' writing performance and motivation: A mixed-methods study. *Reading & Writing Quarterly*, 35(5), 427–444. <https://doi.org/10.1080/10573569.2019.1577778>
- Schunk, D. H. (1999). Social-self interaction and achievement behavior. *Educational Psychologist*, 34(4), 219–227. https://doi.org/10.1207/s15326985Sep3404_3
- Schunk, D. H. (1994). Self-regulation of self-efficacy and attributions in academic settings. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 75–99). Routledge.
- *Seban, D., & Tavsanlı, Ö. F. (2015). Children's sense of being a writer: Identity construction in second grade writers workshop. *International Electronic Journal of Elementary Education*, 7(2), 217–234.

- *Seban, D. (2012). Researching the effect of authoring cycle on third grade students' attitudes towards writing, self-perception and writing ability. *Egitim ve Bilim*, 37(164), 147–158.
- *Sessions, L., Kang, M. O., & Womack, S. (2016). The neglected “R”: Improving writing instruction through iPad apps. *TechTrends*, 60(3), 218–225. <https://doi.org/10.1007/s11528-016-0041-8>
- *Silver, R., & Lee, S. (2007). What does it take to make a change? Teacher feedback and student revisions. *English Teaching: Practice and Critique*, 6(1), 25–49.
- *Snyders, C. S. B. (2014). ‘I wish we could make books all day!’ An observational study of kindergarten children during writing workshop. *Early Childhood Education Journal*, 42(6), 405–414. <https://doi.org/10.1007/s10643-013-0625-2>
- *Teague, B. L., Smith, P. H., & Jiménez, R. T. (2010). Learning to write in a Mexican school. *Journal of Language and Literacy Education*, 6(1), 1–19.
- Thorndike, R. M., & Thorndike-Christ, T. (2014). *Measurement and evaluation in psychology and education* (8th ed). Pearson.
- Tooste, J. R., Didion, L., Peng, P., Filderman, M. J., & McClelland, A. M. (2020). A meta-analytic review of the relations between motivation and reading achievement for K–12 students. *Review of Educational Research*, 90(3), 420–456. <https://doi.org/10.3102/0034654320919352>
- 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., 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>
- *Truax, M. L. (2018). The impact of teacher language and growth mindset feedback on writing motivation. *Literacy Research and Instruction*, 57(2), 135–157. <https://doi.org/10.1080/19388071.2017.1340529>
- *Tunks, K. W. (2010). Attitudes toward writing in fifth grade students preparing for state-mandated writing exam. *International Journal of Applied Educational Studies*, 8(1), 1–9.
- *Unal, E. (2010). An analysis of the writing disposition of fourth and fifth grade elementary school pupils. *Education*, 131(2), 319–330.
- United Nations Convention on the Rights of the Child (UNCRC), November 20, 1989, 1577 UNTS 3. <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>
- Unrau, N. J., Rueda, R., Son, E., Polanin, J. R., Lundeen, R. J., & Muraszewski, A. K. (2018). Can reading self-efficacy be modified? A meta-analysis of the impact of interventions on reading self-efficacy. *Review of Educational Research*, 88(2), 167–204. <https://doi.org/10.3102/0034654317743199>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Walgermo, B. R., & Uppstad, P. H. Assessment as an element for enhancing students identities as readers and writers. In C. Wagner, K. Frankel, and C. Leighton, *Becoming readers and writers: Literate identities across childhood and adolescence*. Routledge. (in press).
- Walgermo, B. R., Foldnes, N., Uppstad, P. H., & Solheim, O. J. (2018). Developmental dynamics of early reading skill, literacy interest and readers' self-concept within the first year of formal schooling. *Reading and Writing*, 31(6), 1379–1399. <https://doi-org.ezproxy.uis.no/10.1007>
- Weiner, B. (1986). An attributional theory of achievement motivation and emotion. In B. Weiner, *An attributional theory of motivation and emotion* (pp. 159–190). New York, NY: Springer US. https://doi.org/10.1007/978-1-4612-4948-1_6
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-Value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*, 89(3), 420.
- *Wilson, K. M., & Trainin, G. (2007). First-grade students' motivation and achievement for reading, writing, and spelling. *Reading Psychology*, 28(3), 257–282. <https://doi.org/10.1080/02702710601186464>
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, 1–10. <https://doi.org/10.1145/2601248.2601268>

- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91. <https://doi.org/10.1006/ceps.1999.1016>
- *Zumbrunn, S., Ekholm, E., Stringer, J. K., McKnight, K., & DeBusk-Lane, M. (2017). Student experiences with writing: Taking the temperature of the classroom. *The Reading Teacher*, 70(6), 667–677. <https://doi.org/10.1002/trtr.1574>
- *Zumbrunn, S., Marrs, S., Broda, M., Ekholm, E., DeBusk-Lane, M., & Jackson, L. (2019). Toward a more complete understanding of writing enjoyment: A mixed methods study of elementary students. *AERA Open*, 5(2), 233285841985079. <https://doi.org/10.1177/2332858419850792>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Aline Alves-Wold¹  · Bente Rigmor Walgermo¹  · Erin McTigue¹  ·
Per Henning Uppstad¹ 

Bente Rigmor Walgermo
bente.r.walgermo@uis.no

Erin McTigue
erin.m.mctigue@uis.no

Per Henning Uppstad
per.h.uppstad@uis.no

¹ Norwegian Centre for Reading Education and Research, University of Stavanger, Stavanger, Norway

Appendix 3 – Article 3

Alves-Wold, A., Walgermo, B. R., & Foldnes, N. (2024). Assessing writing and spelling interest and self-beliefs: Does the type of pictorial support affect first and third graders' responses? *Assessing Writing*, 60, 100833. <https://doi.org/10.1016/j.asw.2024.100833>

ASSESSING WRITING AND SPELLING INTEREST AND SELF-BELIEFS

1

Assessing writing and spelling interest and self-beliefs: Does the type of pictorial support affect first and third graders' responses?

Abstract

An array of pictorial supports (e.g., emojis, geometrical figures, animals) is often used in studies assessing young students' writing motivation with Likert scales. However, although these images may influence the students' responses, sufficient rationales for these choices are often absent from the studies. To the best of our knowledge, the present study is the first to investigate two different types of pictorial support (circles vs. faces) in Likert scales assessing first and third graders' writing interest, self-concept, and spelling interest and self-efficacy. The samples consist of 2197 first graders (mean age 6.8 years) and 1740 third graders (mean age 8.4 years). Results show statistically significant differences among the scales indicating that when face-scales are used, first-graders skip motivation items more often, and students in both grades avoid the minimum values of the scale more often. Gender differences are also found indicating that when face-scales are used, boys in third grade avoid maximum values more often, and girls in both grades avoid the minimum values more often. These findings suggest that the use of circle-scales compared to face-scales seem more appropriate in scales measuring young students' writing and spelling interest and self-beliefs.

Keywords: self-report, writing, spelling, motivation, interest, self-concept, self-efficacy, scale validation, circles, emojis/faces, students' voices

Assessing writing and spelling interest and self-beliefs: Does the type of pictorial support affect first and third graders' responses?

1 Introduction

To date, the field of writing motivation research lacks consensus not only on how motivation constructs are defined (Camacho et al., 2021), but also on how they are measured and operationalized (Authors, xxxx). Even less attention has been given to how valid scales for measuring these constructs with young children should be designed. Although the literature provides some theoretical and general directions for constructing e.g., self-efficacy scales (Bandura, 2006), guidelines are often not accompanied by sufficient empirical support. In addition, pragmatics and convenience tend to guide researchers' design of motivation scales. For instance, one common practice in the field is that when measuring motivation with young children, pictorial supports like emojis, geometrical figures, and animals are used, but a recent review of assessments of writing motivation with K-5 students (Authors, xxxx) shows that despite an array of pictorial supports being used in these assessments, valid rationales for these choices are often missing. This is in contrast to recommendations made by Kane (2006, p. 24) who cautions that test tasks should not be "overly influenced by extraneous factors (e.g., test format)." To address this gap, the present study is, to the best of our knowledge, the first empirical investigation that compares two different types of pictorial supports – specifically *circles* vs. *faces* – and explores the potential influence of such extraneous factors on students' responses regarding their writing interest, self-concept, and spelling interest and self-efficacy.

2 Literature review

2.1 Why do we need to assess students' writing motivation?

Although the extent and frequency to which young children should be tested in school remain an unresolved issue (Green et al., 2022), the use of relatively short screening tests has been promoted as a valuable and less invasive tool for identifying underachieving students at risk of developing learning

difficulties (Green et al., 2022; Solheim et al., 2021)(Authors, xxxx). However, although measuring the students' skill levels is necessary for identifying and addressing specific skill limitations, we know from motivation research that only targeting the students' skills is not sufficient. Since the 1980s, motivational factors, such as the students' interests and beliefs, have received increased emphasis as necessary ingredients for the development of adequate literacy skills (Hidi & Boscolo, 2008). That is, literacy success is dependent on the development of both skills and motivation (Graham et al., 2007; McTigue et al., 2019). Consequently, to capture a broader picture of the students' needs in their literacy development, it is also necessary to measure their motivation levels.

Still, despite an increased focus on motivation in educational contexts, motivation is seldom assessed in national screening tests. This is surprising, given that underachieving students – the target group of screening tests – often also report lower levels of interest (Graham et al., 1993). We thus argue that to get a better picture of these students' complex needs, motivational levels should also be assessed alongside skill levels in literacy screening tests. Moreover, as suggested by Authors (xxxx), measuring motivation constructs like interest and self-beliefs may have beneficial effects for learning, as it contributes to an increased focus on these constructs in learning situations, in addition to providing teachers with useful information on the students' actual motivational states.

With regards to writing motivation, in particular, studies show that students often report lower levels of motivation to write compared to reading (Babayigit, 2019; Graham et al., 2012; Mata, 2011; Merisuo-Storm, 2006; Perry et al., 2003). This means that it is of particular importance to have valid tools to provide teachers and researchers with sound information about students' levels of writing motivation.

2.2 Writing interest and self-beliefs

The motivation constructs investigated in the present study are intentional, as these two broad components of writing motivation, namely *interest* and *self-beliefs*, represent two of the most research-

supported aspects of literacy motivation in the early grades (Authors, xxxx) (Troia et al., 2012). In this section, these two constructs will be defined and their relevance for the development of writing skills and motivation will be briefly presented.

Interest is commonly divided between *situational* and *individual* (or *personal*) *interest*, where the former is more fluctuating and triggered by specific characteristics of tasks, and the latter tends to be more stable as it originates from individual preferences towards tasks or domains (Hidi et al., 2002; Troia et al., 2012). For instance, a student may have low interest in writing argumentative texts (individual interest) but writing a text to his or her parents arguing as to why the student should be allowed to play a specific game may awake the students' interest in writing this type of text (situational interest). According to Hidi et al. (2002, p. 430) *situational interest* "is generated by certain conditions and/or stimuli in the environment and it represents a relatively immediate affective reaction that focuses attention and that may not be long term," whereas *individual interest* concerns "a relatively enduring individual disposition to re-engage with certain objects and events."

Different studies have shown that interest (both situational and individual) positively affects academic learning and facilitates writing performance (Z.-H. Chen & Liu, 2019; A. H. Hall et al., 2017; Hidi et al., 2002; Liao et al., 2018). Another important aspect of interest is that it is not static, that is, it is malleable and can be developed, but "without self-generated or environmental support for continued engagement, it is also possible, for a person's interest in something to decrease or drop off altogether" (Renninger & Hidi, 2017, p. 12). In educational settings, this is of particular importance, because even though students may show little interest towards specific tasks (e.g., structuring essays) or broader domains (e.g., writing), with support from their teachers, their interest can be developed, which can positively affect their motivation levels and consequently their writing skills development.

Although interest is a strong motivational factor for writing (Bruning & Horn, 2000), research shows that the *beliefs individuals have about themselves*, i.e., *self-beliefs*, are also an important

ingredient in the development of individuals' writing motivation (Graham, 2018; Troia et al., 2012). The term *self-beliefs* is commonly used as a general cover term referring to different types of beliefs individuals have about themselves (McTigue et al., 2019; Pajares et al., 1999; Wright et al., 2019). These can refer to more general beliefs towards a domain (e.g., writing self-concept), as well as more focused beliefs towards specific tasks (e.g., spelling self-efficacy). In addition, Pajares and Valiante (2001, p. 367) remark the importance of self-worth beliefs for *writing self-concept* and define it as one's "judgements of self-worth associated with one's self-perception as a writer", while *self-efficacy* concerns individuals' convictions about their capabilities in succeeding with more specific activities (e.g., ideation self-efficacy) or accomplishing tasks (e.g., self-efficacy for poetry writing).

As remarked by Troia et al. (2012, p. 8), self-concept and self-efficacy beliefs may vary from each other. For example, a student can have positive self-beliefs about his or her competence towards the broad domain of writing ("I believe I am a good writer" – i.e., positive writing *self-concept*), but have more varied beliefs regarding specific tasks or skills ("I think that I was not able to spell properly the words in my text" – i.e., negative spelling *self-efficacy*). Like interest, self-concept and self-efficacy beliefs are also malleable and can be developed (Bandura, 1997). However, whereas self-efficacy beliefs tend to be more malleable and future-oriented, self-concept appears to be more stable and oriented towards past experiences (Bong & Skaalvik, 2003; Troia et al., 2012; Unrau et al., 2018). In fact, as suggested by Unrau et al. (2018, p. 169), self-efficacy "acts as a precursor to the development of self-concept." That is, as an individual experiences mastery with task-specific activities, like successfully structuring paragraphs, spelling words, or writing an essay, the individual will develop beliefs about his or her efficacy in successfully accomplishing these activities. These collective beliefs may in turn strengthen the individual's general self-perception within the broader domain of writing, i.e., the individual's writer self-concept.

This understanding of self-efficacy as a forerunner for writer self-concept is important in educational settings, as each small writing task students encounter may add a brick to their broader writer self-concept puzzle. As proposed by Authors (xxxx), in classrooms we should therefore “work hands-on in every learning situation so as to strengthen students’ feeling of mastery when they are engaged in their daily reading and writing tasks—and, importantly, this includes assessments.” For the teacher, mindful assessment of students’ writing interest and self-beliefs is of potential great value, as these motivational factors influence students’ subsequent desire or decision to re-engage with writing tasks (Renninger & Hidi, 2017) and “are positively related to the amount of effort expended to perform a task, persistence with a difficult task, the recruitment of strategies to accomplish a task, and actual task performance, regardless of one’s age, gender, or ethnicity” (Troia et al., 2012, p. 6). However, for teachers and researchers to obtain accurate information regarding the students’ interest and beliefs, valid assessment instruments are needed.

2.3 Assessing young students’ writing interest and self-beliefs

The use of self-reports with young children has been debated, as there is discussion regarding reliability issues like children’s overestimation of their abilities, and their limited capacity and lack of opportunities for comparing themselves with other children, as well as with their own previous performances (Mata, 2011, p. 289). However, many researchers have advocated for the importance and richness of children’s self-reported data (A. H. Hall & Axelrod, 2014; Zumbunn et al., 2017) (Authors, xxxx), and from the literature reviewed, Sturgess et al. (2002, p. 115) argue that “there is convincing evidence that the use of self-report with young children is valid, desirable and helpful,” and that self-reports have been “demonstrated to be reliable in many instances and contexts.” In addition, various studies have shown that children in primary grades are indeed able to give an indication of their motivations to write (Graham et al., 2012; Mata, 2011), even though some may lack a more robust vocabulary to express themselves (Paquette, 2008).

Depending on the type of information that researchers wish to investigate regarding students' writing motivation, different types of self-report have been used to generate data, including surveys (e.g., Bayat, 2016; Liao et al., 2018), questionnaires (e.g., Jones et al., 2016; Kanala et al., 2013), interviews (e.g., Abbott, 2000; Beck & Fetherston, 2003), and drawings (e.g., Zumbrunn et al., 2017). In their review of writing motivational measures for K-5 students, Authors (xxxx) point out that quantitative assessment methods, like scales, were often "used to measure frequencies and levels, that is, how often students worked with writing activities and how much they enjoyed or valued these activities," while qualitative assessments like "open-ended questions in interviews and surveys, were used to investigate reasons for their [the students] choices and factors that influenced those choices."

Although qualitative methods are also used for investigating students' writing interest and self-beliefs (A. H. Hall & Axelrod, 2014; Kim & Lorschach, 2005; Zumbrunn et al., 2017), when it comes to large samples, these motivational constructs are often assessed in terms of levels, in which quantitative methods are then commonly used. When measuring students' levels of interest, questions in which students indicate how much they like (or enjoy) writing in general or specific writing activities are commonly used (e.g., Boscolo et al., 2012). Levels of writing self-beliefs are, in turn, sometimes measured through questions in which students indicate how good they think they are to write in general (writing self-concept) or when executing specific writing tasks (task-specific self-efficacy) (e.g., S. D. Miller & Meece, 1997). Self-efficacy is also measured through statements, in which students indicate their level of confidence at executing the tasks described in the statements, like "I can read through my text and correct spelling mistakes" (Grenner et al., 2020, p. 3). Often, to capture these levels, Likert-scales ranging for instance from "not at all" to "very much" (e.g., Hier & Mahony, 2018) are used, and with young students, these scales are accompanied by pictorial supports (e.g., geometrical figures ranging from small to big, or faces ranging from sad to happy) to help the students visualize different response levels (e.g., Babayigit, 2019; Grenner et al., 2020; Jones et al., 2016).

These pictorial supports can be seen as an aid for students to rate their responses, but they may also indicate a source of bias, as these images themselves may influence the students' responses. For instance, in a study on pain ratings, Chambers and Craig (1998) found that the type of face (i.e., smiling or neutral face to indicate 'no pain') used in the scales influenced the children's self-reported ratings of pain, and that the biasing impact was more pronounced among younger children (5-8 years old) compared to their older peers (9-12 years old). Similar biases related to the type of face scale used were reported by L. Hall et al. (2016), in a study investigating children's (9-11 years old) evaluations of an interactive narrative ('serious game'). According to the authors, results showed that children avoided selecting the negative options of the scale, even when a neutral face was used as the end point.

Although there are limited studies on the use of face-scales, a growing number of studies (Gallo et al., 2017; Schouteten et al., 2018; Setty et al., 2019) have investigated the use of emojis, which are "small icons used in digital communication to express emotions/feelings" (Sick et al., 2020, p. 1). These studies have also found gender differences in emoji use. For example, in a study investigating the use of emojis to indicate food preferences among 9-13-year-old children, Sick et al. (2022) found gender differences in the participants' responses, in which girls (74 %) showed greater preference towards positive emojis than boys (64 %)($p = 0.023$). In fact, in 2018, Z. Chen et al. (2018) argued that gender differences were not only statistically significant, machine learning algorithms were already at that time able to deduce the gender of a person solely on the basis of the emojis inserted in the text. Sick et al. (2020) also note that gender differences in emoji preferences should be seen in relation to gender differences in non-verbal communication and emotion expression, since meta-analytic reviews indicate that girls tend to express more positive emotions compared to boys (Kring & Gordon, 1998), and more internalized emotions (e.g., fear and sadness), compared to more externalized emotions (e.g., anger) expressed by boys (Chaplin & Aldao, 2013).

With regards to motivation measures in particular, Bandura (2006, p. 313) has also brought attention to a potential bias effect related to face-scales, and the author recommends avoiding face-scales ranging from happy to sad for ratings of self-efficacy beliefs, as children “may misread such a scale as measuring their happiness or sadness rather than how confident they are that they can perform given tasks.” For this matter, the author recommends for example a circle-scale instead, ranging from smaller to bigger circles. However, although studies like the ones mentioned above have investigated potential influences of the design of face-scales on children’s responses, to the best of our knowledge, the present investigation is the first empirical study to compare circle- and face-scales and their potential influence on measures of writing motivational constructs.

Finally, although examining students’ response patterns is necessary in investigations of scale validity, scholars have brought attention to the importance of also investigating students’ *lack of responses*, i.e., their *skipping behavior* in low-stakes assessments. For instance, in an investigation of non-serious behavior in the Programme for International Student Assessment (PISA), Akyol et al. (2021) uses the students’ skipping behavior (i.e., not answering questions) as one of their markers for non-seriousness, and found that girls were less likely to express non-serious behavior. Similarly, (Liu & Hau, 2020, p. 1121) argue that if students intentionally skip items (i.e., user missingness), such missingness information may be seen as an indication of “(lack of) test-taking motivation.” The authors also remark that it is important to explore potential reasons of missingness, like slow answering pace, which commonly leads to non-reached items at the end of the test, or issues related to the design of the test. For example, in an investigation of a Turkish university entrance exam, Akyol et al. (2022) found that skipping behavior was associated with an important feature of the test design, which was the inclusion of penalties for choosing wrong answers. The authors found gender differences regarding skipping behavior and risk taking, and noted that although females were significantly more averse to risk, the effect of gender was small, favoring the inclusion of higher guessing penalties for increasing

the precision of the exam. However, even though a few studies have examined the significance of students' skipping behavior in low-stakes assessments, none have examined this behavior in relation to different types of scales. Thus, in the present study, we also investigate whether the students' skipping behavior in the writing interest and self-beliefs items may be related to the type of pictorial support used in the test, that is, if scale types may lead to more or less skipping.

2.4 The present study

The present study is undertaken as part of the development of a new national literacy screening test in Norway for first and third graders. This test has been developed by the Norwegian Centre for Reading Education and Research as requested by the Norwegian Directorate for Education and Training (NDET) with the purpose of helping teachers identify students who need extra support to develop adequate reading and writing skills.

Although NDET recommends the completion of this test to all schools, participation in first grade is voluntary, both at school and individual levels. That is, school administrators are allowed to refrain from implementing this test in their schools, and in cases where teachers do not see the completion of this test as beneficial for individual students (e.g., students with disabilities, or students who do not comprehend the Norwegian language), these students are offered alternative activities and do not have to participate in this screening test. In addition, in order not to force students to answer tasks that they are not comfortable with answering or do not want to answer, NDET requires that the design of the test must allow students to skip tasks.

Following the common practice of using pictorial support (e.g., geometrical figures, emojis, animals) in Likert scales when assessing young children's writing motivation through self-reports, the present study uses two types of pictorial support separately: *circles* and *faces* (as shown in Figure 1), where we investigate whether these types of pictorial support influence the students' responses.

The research question guiding this investigation is as follows: *Do students report, or skip reporting, levels of writing interest and self-beliefs differently depending on the type of pictorial support used in the Likert scales? If so, to what extent are the differences explained by grade, gender, and type of motivational construct measured?*

To investigate this research question, two hypotheses have been formulated, and here we present the reasoning for them. First, given that in the present study, students are allowed to skip items, we hypothesize that students may choose to skip items that they do not feel comfortable with answering, or that they are not sufficiently motivated to answer. According to Bandura (2006), face-scales tap into the respondents' emotional register more than circle-scales. Consequently, providing ratings about one's interest and self-beliefs may be more costly when face-scales are used. Therefore, we expect higher skipping rates among students when face-scales are used.

Second, following the proposition of Bandura (2006), we argue that the extreme ends of the face-scale are more emotionally loaded than the middle, more neutral options. Based on this assumption of the emotional load of face-scales, we thus expect more avoidance of the extreme ends when these are used in comparison with circle-scales.

Specifically, in the present study, we thus evaluate whether data supports the following two hypotheses:

*H*₁ We hypothesize that the students' likelihood of skipping motivation items will be significantly higher for the face-scale compared to the circle-scale.

*H*₂ We hypothesize that, when not skipping, students will tend to use the scale extremes to a lesser degree with the face-scale compared to the circle-scale.

If confirmed, these hypotheses can be an indication that – due to their emotional load – face-scales are less optimal for measuring students' writing interest and self-beliefs, as they may lead to

higher levels of skipping behavior (i.e., missing data) and avoidance of extreme ends, when compared to circle-scales.

3 Method

3.1 Study context

As part of the development of a new national literacy screening test, students from different Norwegian schools were asked to complete a digital booklet comprising four short sections: word reading, spelling, reading comprehension, and language comprehension. Each student carried out two booklets within a period of 2 weeks, and had 30 minutes to complete each booklet in their respective classrooms. The structure of the booklets, containing the four short sections mentioned above, was the same in both weeks, but the tasks included in each section contained of course different texts to be read and interpreted, and different words to be spelled. Given the focus of the present study, data for our investigation was derived from the spelling section of the test.

Although a myriad of tasks (e.g., composing narrative or informative texts, writing shopping lists, or birthday cards) can be used to operationalize writing in writing assessments, *spelling* is chosen for this purpose in the present screening test for two reasons. First, given the aim of identifying students at risk of developing literacy difficulties, tasks with strong prediction of later literacy difficulties were needed, and spelling tasks have proven to be powerful predictors. For instance, according to Lundetræ and Thomson (2018), although spelling has received less attention in research, it demands a deeper and more actively generated understanding of written word forms, which frequently serves as a stronger indicator of literacy difficulties in alphabetic languages. Moreover, as noted by Troia et al. (2012, p. 13) “lower level transcription skills such as spelling and handwriting, exert a powerful influence on how well students accomplish composing tasks when these skills are underdeveloped,” and serve therefore as significant indicators of literacy difficulties. Second, given the ambitious aim of considerably reducing the length of earlier versions of this national literacy screening

test from 60 to 30 minutes, shorter tasks like spelling tasks were preferred. However, although writing is operationalized as a spelling task in the present test, this task is designed to mirror high-quality and motivating literacy practices, in which the spelling task is presented in an authentic and hopefully motivating scenario, as it will be described in more detail in section 3.3.2.

Prior to the students' participation, teachers led all students through an instruction including sample items representing each of the sections in the test. Students then completed the test's tasks by using their individual school digital devices (tablet computers or Chromebooks) during class. Each participating student was asked to log into the system and complete the literacy test according to on-screen instructions that were read out loud digitally to them. To hear these instructions, all students used headphones during the test.

3.2 Participants

The sample consists of 2197 first grade students (50.5 % girls, mean age 6.8 years) and 1740 third grade students (51.3 % girls, mean age 8.4 years). Data for third graders was collected in September 2021, meaning that they were at the onset of their third year of formal literacy instruction, while data for first graders were collected towards the end of their first year of formal literacy instruction, between April and June 2022.

3.3 Measures

3.3.1 Writing interest and self-concept

The first task the students encountered in each test booklet was concerned with their general interest and self-concept for reading and writing. However, as mentioned above, given the focus of the present study, data related to writing, rather than reading, is presented.

The students' general interest and self-concept were each measured using one single item, adapted from Nuutila et al. (2020), as described in items 1 and 2 in Table 1.

A picture of students writing, as shown in Figure 2, accompanied each of these two items. This picture was designed for the students to give their responses according to an elaborated view of writing, which includes both analog and digital writing, as well as authentic writing activities, like writing wishing or shopping lists.

Both items were measured with a 5-point Likert scale ranging from 5 (Very much) to 1 (Not at all), and as pictorial support, these answers were accompanied by either a face scale (as shown in Figure 1a) or a circle scale (as shown in Figure 1b).

3.3.2 Authentic word spelling task

After answering the questions regarding their writing (and reading) interest and self-concept, every participant carried out computerized literacy tasks belonging to each of the four short sections of the literacy test, as mentioned above. The sub-section measuring the students' spelling skill consists of 7 words with orthographic transparent and opaque spelling patterns, which reflect the semi-transparent Norwegian orthography. For these tasks, a picture of a boy sitting in his room appeared on the screen (see Figure 3a). The students were introduced to this boy as 'Josef', whose birthday is soon coming up, and to which he was really looking forward. Then, the students were asked whether they could help Josef with writing a wishing list for his birthday. For each word in the wishing list, a thought bubble with an illustration of that word appeared to the students, as shown in Figure 3b, which is an illustration of the word 'hammock' ("hengekøye" in Norwegian). Each word was then read out loud digitally to the students, and they could type in their answers.

3.3.3 Interest and self-efficacy for spelling task

Before and after completing these spelling tasks, the students were asked questions regarding their interest and self-efficacy for these tasks, totaling 7 items related to their interest and self-beliefs, as shown in Table 1.

Items 3 to 6 were also measured with a scale ranging from 5 (Very much) to 1 (Not at all), and the same type of pictorial support used for items 1 and 2. However, for item 7, following the lead from Nuutila et al. (2020), we used a 3-point Likert scale with the response options “Yes”, “*Maybe*”, and “No”.

3.4 Study Design

All participants responded to these 7 items twice within a period of two weeks (once in each test booklet), and the order of the scales used in each booklet was reversed, meaning that a flipped design was employed, as shown in Figure 4. That is, for all 7 items, approximately half of the students were presented with the face-scale in booklet 1, and with the circle-scale in booklet 2 (1146 first-graders, and 832 third-graders), whereas the other half was presented with the opposite sequence (1051 first-graders, and 908 third-graders). Given that the nature of the present study implies a flipped design, all participating students completed their assigned test booklet 1 (wave 1) before their assigned test booklet 2 (wave 2).

3.5 Analysis

Each hypothesis was analyzed first by considering descriptive statistics and then proceeding to statistical modeling and inferential statistics. Descriptive statistics are mostly presented visually, relying on plot panels fully crossing grade, wave, gender, and format.

For formal significance testing, we modeled item responses using mixed linear logistic regression. Given that there are significant developmental differences between grade 1 and 3, we chose to model the fixed effects of item and gender in separate grades. This is because there are expected differences in what constitutes interest and self-beliefs in grade 1 when students have just had their first formal year of literacy instruction compared to grade 3, when these skills are expected to be more established. Running a model with both grades included would mean that we free up fixed effects across grades, leading to a much more complex model, with possibly more unstable estimates. For this

reason, we ran separate models in grade 1 and in grade 3. To further support this decision, we calculated the correlation matrices among the 7 items in grade 1 and in grade 3 and tested whether these were equal. This hypothesis was rejected (chi-square=220.9 with 21 dfs, p=0) by Steiger’s test (Steiger, 1980).

Further nesting the students in classes was not possible, as NDET did not allow the collection of class information during this test.

Let us consider the outcome “skipping” on an item response. The skipping behavior of student s in relation to on item I , in wave W , and under format F , is modeled by the binary variable $Y_{I,W,F,s}$ being equal to 1 if a response was skipped, 0 otherwise. The overall likelihood of skipping is allowed to vary between students, and this is modelled by a random intercept. Also, the change in likelihood of skipping between waves is allowed to randomly vary between students, and this is modelled by a random slope. The seven motivation items were modeled as fixed effects, with dummy variable coding, using the first item, general interest (INT_GEN), as reference.

In summary, the mixed logistic model of $Y_{I,W,F}$ contains a random effect of wave W ($W=0$ and $W=1$ correspond to wave 1 and 2, respectively), a fixed format effect F ($F=0$ and $F=1$ correspond to circle and face, respectively), and a fixed student level gender effect G_s ($G_s=0$ and $G_s=1$ correspond to girl and boy, respectively). Of primary concern is whether gender moderates the effect of pictorial support format on responses. We therefore also included a cross-level interaction between F_i and G_w in our mixed model:

$$\text{logit } t(Y_{I,W,F,s}) = \alpha_s + \sum_{i=1}^6 \gamma_i D_i + (\beta_W + \beta_s) W + \beta_F F + \beta_G G_s + \beta_{FG} F \cdot G_s + \epsilon_{I,W,F,s} \quad (1)$$

where α_s is a random intercept, and β_W and β_s are the fixed and random effect of wave, respectively. Also, β_F is the fixed format effect for girls, $\beta_F + \beta_{FG}$ is the fixed format effect for boys, and β_G is the fixed gender effect for the circle format. The fixed effect of each item, relative to the reference item

INT_GEN, is represented by coefficients $\gamma_1, \dots, \gamma_6$. Given that any effect of wave over and beyond the influence of format is likely small and therefore of little practical interest in test design, we do not add the complexity of an interaction W*F in equation (1) for reasons of parsimony.

Modeling and plotting were conducted in the R software environment (R Core Team, 2022), using packages *lme4* (Bates et al., 2015) and *ggplot2* (Wickham, 2016), respectively. The scripts are available upon request.

4 Results

Univariate descriptive statistics for the items in Table 1 are presented in Table 2. Correlations among items within each wave are given in Table 3a, while the repeated measure correlations across waves are presented in Table 3b.

4.1 Hypothesis 1

4.1.1 *Skipping behavior*

The number of participants and the percentages of skipping responses (across all seven motivation items) over the two administration waves are given in Table 4, for each combination of grade, format, wave, and gender.

The observed skipping pattern is further refined in Figure 5, where each of the seven motivation items are separated out. A consistent finding for first graders is that item skipping occurs more frequently in Wave 2 compared to Wave 1, and skipping consistently occurs more often with the face format compared to the circle format. Also, the sensitivity to item format is generally more pronounced among boys compared to girls. In grade 3, the differences between skipping in Wave 1 and skipping in Wave 2 are generally negligible, as are the differences in skipping between the face and circle formats.

To further inquire into whether the findings in grade 1 reflect significant effects, we ran a mixed logistic regression model with skipping as the outcome. The estimated model is given in Table 5. The model yielded a significant ($p < .1$) effect of format, with the face format associated with an increase in

the likelihood of skipping, while significant gender differences were not found. For Grade 3, the effects of Wave are attenuated and opposite to the results for Grade 1 (see Table 6). Moreover, in Grade 3 we found no significant effect of gender nor item format on skipping behavior, and no interaction between gender and item format.

However, although both first and third graders skipped general interest and self-concept items (i.e., INT_GE and SC_GE) more often than task-specific items (i.e., INT_PRE, SE_PRE, INT_POST, SE_POST, and INT_CONT), only third graders consistently demonstrated significantly less skipping behavior for all 5 task-specific items.

4.2 Hypothesis 2

4.2.1 Endorsing maximum values

Consider the percentage of students (who did not skip) endorsing the highest level of the motivation items, depicted in Figure 6. Apart from grade 1 boys in wave 2, it is hard to visually find clear support for H_2 in this Figure.

To further investigate this hypothesis, estimation of the mixed linear logistic model with endorsement of the max value as outcome variable was conducted (see Table 7 for grade 1). Table 7 shows that first graders tend to endorse max values more often in Wave 2. Boys endorse max values significantly less often than girls, but this difference is regardless of the scale format. With regards to scale format, both genders tend to endorse max values to a less extent when face scales are used, and this difference is more pronounced for boys, but these tendencies are not statistically significant.

For grade 3, however, as shown in Table 8, there is a marginally significant moderation of the effect of the face format for boys, that is, the likelihood of endorsing max values in the face scale is significantly less for boys compared to girls.

Finally, estimate values for endorsing max values in both grades are higher for task-specific self-efficacy items (i.e., SE_PRE and SE_POST), than for general self-concept and items assessing the

students' interest. That is, first and third graders endorse max values more often for task-specific items assessing their self-beliefs, rather than their interest and general self-concept.

4.2.2 Endorsing minimum values

The observed percentages of endorsing the minimum scale value are presented in Figure 7. Clearly, boys consistently endorse the minimum value more often than do girls. Also, in wave 1 it is seen consistently that with circles there is a higher percentage of minimum values compared to the face format. Surprisingly, in wave 2, the situation is reversed for boys in grade 1, where the face format is associated with a clearly higher percentage of minimal values.

The model estimates for grade 1 are given in Table 9, and for grade 3 in 10. Similar patterns are found for both grades, in which there is less likelihood of choosing the minimum values in the face scale compared to the circle format. However, compared to girls, boys in both grades choose the minimum value in face scales more often compared to girls.

Finally, an interesting consistent zigzag pattern can be seen in Figure 7, in which students in both grades seem to endorse minimum values less often when the items measure their writing self-beliefs (i.e., SC_GE, SE_PRE, and SE_POST) compared to items measuring their writing interest (i.e., INT_GE, INT_PRE, INT_POST, and INT_CONT). Estimated coefficients for these items reflecting this zigzag pattern are reported in Tables 9 and 10.

5 Discussion and implications of findings

The present study set out to investigate whether first and third grade students' responses on measures of writing interest, self-concept, and spelling self-efficacy are influenced by the type of pictorial support used in the Likert-scales measuring these constructs. Specifically, we compared results from two scales - faces and circles - and investigated if potential differences could be explained in relation to the students' grade, gender, and type of pictorial support. In this investigation, we hypothesized that students' likelihood of skipping motivation items would be significantly higher for

the face-scale compared to the circle-scale (H_1), and that, when not skipping, students would tend to use the scale extremes to a lesser degree with the face-scale compared to the circle-scale (H_2). In this section, we will discuss our findings with relation to these two hypotheses, and the implications of these findings for the development of motivation scales for young children.

5.1 Skipping behavior

Results from the present study show that hypothesis 1 is partially supported. That is, boys and girls in first grade skip motivational items significantly more often when face-scales are used compared to circle-scales. However, although students in third grade show similar patterns, these differences are not statistically significant, which may indicate that the skipping behavior of third graders is less influenced by the face-scale format. This finding is in line with Chambers and Craig (1998), who found that the biasing effect of the type of face used in their scales was more pronounced among younger children (5-8 years old) compared to their older peers (9-12 years old).

For third graders, skipping behavior seems to be more related to the type of motivational construct measured. That is, even though students in both grades skip motivational items less often when these are measuring their task-specific writing interest and self-efficacy compared to general writing interest and self-concept, only third graders consistently show statistically significant differences. This tendency of responding more frequently to task-specific rather than general motivational items may be related to students' difficulties in interpreting what general items entail. For instance, Wilson and Trainin (2007, p. 275) remarks that researchers have previously "identified the propensity to rely on a generalized sense of self-concept as a confounding problem when working with young children." The authors then advocate for the importance of investigating motivation at the task level and of providing young students with concrete tasks when assessing their motivations to write as these make the items more concrete and accessible to young children. When providing students with tasks in such assessments, Authors (xxxx) also remark that an overuse of specific writing tasks (e.g.,

story-writing) in these assessments represent “a clear construct-underrepresentation in current research base on motivation to write.” The authors then recommend the inclusion of more varied writing tasks in future assessments of writing motivation, which more realistically reflect the myriad of tasks, activities, tools, and genres that writing entails.

However, although first graders also tend to skip task-specific items more often, their skipping behavior is not as consistently statistically significant as for third graders. This additional difference between first and third graders may be related to their views of what writing - as a general construct - entails. That is, studies have shown that younger students often have narrower views of writing, in which writing is often equated with specific tasks like having neat handwriting or spelling words accurately (Beck & Fetherston, 2003; Kim & Lorsbach, 2005; Paquette et al., 2013). Thus, when answering general items, first graders may picture more concrete and limited tasks than their older peers, which make these items more accessible and easier to interpret.

In addition to interpretation difficulties regarding general items, students’ skipping behavior may also be linked to characteristics of the constructs measured. For instance, Authors (xxxx) argue that general reading and writing interest and self-concept are much closer related to the students’ reader and writer *identities*. That is, judging one’s interest and self-efficacy *towards specific tasks* is emotionally less costly than rating general constructs that are more directed *towards one’s identities*. Moreover, providing such ratings about oneself may be more costly when face-scales are used, as according to Bandura (2006), these types of scales tap into the respondents’ emotional register more than circle-scales. In fact, in his renowned book *Thinking, Fast and Slow* on behavioral economics and cognitive biases, Kahneman (2011) refers to the research of Ekman (2006) and argues that humans assign stronger value to negative face expressions than positive ones. Further, the author remarks that face expressions are also related to our negativity bias and self-esteem maintenance. This claim is based on research conducted by Baumeister et al. (2001) who found that humans tend to be more

influenced by negative events and experiences than positive ones, and that negative emotions and face expressions, such as fear, anger, or sadness, have a stronger impact on our cognition and behavior compared to positive emotions and expressions. On the basis of such studies, Kahneman (2011) argues that to protect ourselves from danger and negative experiences, we assign stronger value to negative face expressions than positive ones and are more motivated to avoid bad self-definitions than to pursue good ones. This asymmetry of negative bias leads researchers to conclude that our brain in evolutionary terms is designed to assign priority to bad news. Thus, to protect oneself from negative representations of one's identities, students in the present study may prefer to skip motivation items to avoid choosing negative face expressions to represent themselves.

Finally, as suggested by Liu and Hau (2020, p. 1121), students' skipping behavior may also be "used as a direct measure of (lack of) test-taking motivation." However, in the present study, we speculate whether the students' high frequency of skipping can also be seen as a direct measure of *lack of writing motivation*. That is, in measures of motivation, skipping test items may be seen not only as an indication of lack of motivation towards test-taking, but also as an indication of low levels of motivation towards the tasks or domains being assessed in the skipped items.

5.2 Endorsing extreme values

Results from the present study show that hypothesis 2 is partially supported with regards to maximum values, but confirmed with regards to minimum values. That is, although the likelihood of endorsing *maximum* values decreased significantly only for boys in third grade when face-scales were used, the likelihood of endorsing *minimum* values decreased significantly in both grades when face-scales were used. The significant avoidance of maximum values in face-scales for third-grader boys compared to girls aligns with findings from a meta-analysis conducted by Kring and Gordon (1998), in which it was found that girls tend to express more positive emotions compared to boys. Thus, it may be easier for girls to choose positive face expressions from the extreme end of the scales, compared to

boys. However, a meta-analysis conducted by Chaplin and Aldao (2013) indicate that boys, in turn, are more likely to express externalizing emotions, like anger, compared to girls, which may explain why it seems to be easier for boys than girls in both grades to choose negative face expressions at the extreme end of the face-scale.

Interestingly, results from the present study also show that students in both grades are significantly more likely to endorse maximum values for items measuring their task-specific self-beliefs compared to their interest, and at the same time less likely to endorse minimum values for the same items (see zigzag pattern in Figures 6 and 7). This finding indicates that students seem to be more comfortable with choosing the negative end of scales to rate their *interest towards tasks*, but may find it more difficult to use the negative end of scales for rating *beliefs about their own competence*. Given that assessment situations may also contribute to the students' formation of their writer identities (Authors, xxxx), it is important to avoid exposing them to such uncomfortable assessment situations, as they can potentially be harmful for their identities.

5.3 Avoiding the biasing effect of face-scales

Results from the present study support the recommendations of Bandura (2006) to avoid face-scales in measures of self-efficacy, and extends these recommendations also to measures of interest and self-concept. The findings discussed here are also in line with prior research on the use of emojis, in which facial expressions were demonstrated to be often open to reinterpretation and misinterpretation (D. M. Miller et al., 2015). Indeed, from everyday life, we know that a smile can signal a myriad of different underlying emotions. This everyday assumption is underpinned by studies that have found that even within the same cultures, people often do not add the same meaning to emojis - regardless of whether emojis are presented in context or isolation (H. Miller et al., 2016). These findings may indicate that introducing facial expressions in assessments can lead to confusion regarding what the faces actually mean. Additionally, this confusion may to a larger degree be evident among boys, and

lead to higher skipping rates of items containing facial expressions, compared to girls. In fact, research suggests that females more accurately than males are able to identify emotions from facial expressions, even under conditions of minimal stimulus information (J. A. Hall & Matsumoto, 2004).

Goldshmidt and Weller (2000) suggest that females in general use significantly more emotional words than males and that such gender differences in language and emotion stems from the socialization of the sexes. For instance, Kuebli and Fivush (1992) argue that mothers and fathers both tend to use a greater number and variety of emotion words with daughters compared with their sons. Moreover, in the literature, females have been reported to value, experience, and express emotions more than men (Berghout Austin et al., 1987; Block, 1983; Sprecher & Sedikides, 1993).

Despite these gender differences, results from the present study show that the choice of circle-scales compared to face-scales seem more appropriate for both genders, especially for first graders, who seem to be more sensitive to face expressions in scales measuring their writing interest and self-beliefs.

5.4 Limitations and future directions

The face-scales used in the present study are quite neutral concerning gender and color. This means that emojis with stronger emotional value (e.g., scales rendering from happy green to furious red faces) could potentially have resulted in more extreme values and larger group differences. However, our choice of faces for measuring motivation for writing were considerate towards the fact that we wanted to limit possibilities for measuring adjacent emotional states to students' motivation for writing. Therefore, in accordance with recommendations of both Bandura (2006) and Kane (2006, p. 24), we aimed to within the two scales as purely as possible measure students' interest and self-beliefs related to writing. In addition, given that these scales are used in national screening tests requested by NDET, these scales need to comply with national regulations regarding universal design, which adds external constraints regarding the shaping of these scales. To comply with these regulations, the shaping of these

scales was therefore developed in cooperation with professional graphic designers, where we aimed to create scales that were as neutral as possible. For instance, by avoiding the addition of different colors to the scales. This choice was deliberate to avoid introducing too much “noise” (Kahneman et al., 2021) to the scales, as we aimed to ensure that the distinctions derived from the face-scale remained primarily influenced by the facial expressions themselves, rather than other extraneous factors like color variations.

In order to even more specifically investigate students’ perceptions of the two different scales, we could actually also within the test have asked the students: ‘Do you prefer answering these tasks with circles or faces?’ (e.g., giving the students the choice between: circles – faces – eventually other types of scale). Adding an open-ended question where the test-takers could justify their choice of scale could also have provided valuable information. Importantly, however, inserting such additional questions would increase the test length and could have possibly challenged the students’ test stamina. Given that the triggers for the students’ choices are likely to be partly sub-conscious, it remains an empirical question whether qualitative investigations into students’ thoughts after finishing these tasks might be preferable for gaining more valuable in-depth information about the students’ perception of the scales. Additional qualitative investigations were originally a part of the present study design, but due to Covid-19 restrictions, these investigations could not be carried out.

Furthermore, in the present study, due to the students’ young age, we chose 5-point Likert scales. This was in spite of the fact that prior research have suggested that more categories on a scale are associated with higher reliability, and that 7-point scales may be superior in terms of reliability (Preston & Colman, 2000) (Authors, xxxx).

An open question is also whether the students’ reporting of motivational levels was influenced by the sequence within the scales. In the present study all tasks were presented from the left ranging from 5 to 1: positive towards more negative faces and bigger towards smaller circles. To avoid this

potential anchoring effect, L. Hall et al. (2016) suggest to flip the sequence of the scale for 50 percent of the given sample (e.g. from negative towards positive, ranging from 1 to 5). Future investigations should consider including this flipped design when comparing the quality of different scales.

Given that our current flipped design only includes two constellations (i.e., face-circle and circle-face), another limitation of the present study regards the impossibility of separately modeling intra-individual changes that are due to factors other than the change of format between the two waves, as this type of variance is confounded with the format change in our current design. To extract intra-individual variance from the format-based change, we would need to include two additional constellations in our test design, namely face-face, and circle-circle. However, any systematic pattern in intra-individual changes will at least be cancelled in our current model, given the random assignment of face-circle vs circle-face in our flipped design. In addition, to account for random pattern in intra-individual change, we included a random slope associated with wave.

Finally, given the widespread use of different types of pictorial support in writing research and other fields, we argue that more attention and further testing of different types of pictorial supports is needed in the development of valid scales for young students.

5.5 Concluding remarks

The take home message from the present investigation is that it is not indifferent what kind of pictorial support is used in measures of elementary students' motivation for writing. As remarked by Kahneman et al. (2021, p. 199), "the choice of a scale can make a large difference in the amount of noise in judgments, because ambiguous scales are noisy," thus choosing appropriate scales require deliberate considerations. With regard to measures of writing and spelling interest and self-beliefs, findings from the present study suggest that in comparison to face-scales, circle-scales seem to be a more appropriate choice for measuring these constructs among young students. More specifically the circle-scales seemed to be the least noisy for these students. In sum, the results from the current study

underpin the recommendations of Bandura (2006) to avoid face-scales in measures of self-efficacy, and further extends these recommendations to measures of writing interest and self-concept.

References

- Abbott, J. A. (2000). "Blinking Out" and "Having the Touch": Two Fifth-Grade Boys Talk about Flow Experiences in Writing. *Written Communication, 17*(1), 53–92.
<https://doi.org/10.1177/0741088300017001003>
- Akyol, P., Key, J., & Krishna, K. (2022). Hit or miss? Test taking behavior in multiple choice exams. *Annals of Economics and Statistics, (147)*, 3–50.
- Akyol, P., Krishna, K., & Wang, J. (2021). Taking pisa seriously: How accurate are low-stakes exams? *Journal of Labor Research, 42*, 184–243.
- Babayigit, O. (2019). Examining the Effect of Creative Writing Activities on Reading, Writing and Language Lesson Attitudes of Elementary School Fourth Grade Students. *European Journal of Educational Research, 8*(1). <https://doi.org/10.12973/eu-jer.8.1.213>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5*(1), 307–337.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology, 5*(4), 323–370.
- Bayat, S. (2016). The Effectiveness of the Creative Writing Instruction Program Based on Speaking Activities (CWIPSA). *International Electronic Journal of Elementary Education, 8*(4), 617–628.
- Beck, N., & Fetherston, T. (2003). The Effects of Incorporating a Word Processor into a Year Three

- Writing Program. *Information Technology in Childhood Education Annual*, 2003(1), 139–161.
<https://www.learntechlib.org/p/17765/>
- Berghout Austin, A. M., Salehi, M., & Leffler, A. (1987). Gender and developmental differences in children’s conversations. *Sex Roles*, 16, 497–510.
- Block, J. H. (1983). Differential premises arising from differential socialization of the sexes: Some conjectures. *Child development*, 1335–1354.
- Bong, M., & Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review*, 15(1), 1–40.
<https://doi.org/1040-726X/03/0300-0001/0>
- Boscolo, P., Gelati, C., & Galvan, N. (2012). Teaching Elementary School Students to Play With Meanings and Genre. *Reading & Writing Quarterly*, 28(1), 29–50.
<https://doi.org/10.1080/10573569.2012.632730>
- Bruning, R., & Horn, C. (2000). Developing Motivation to Write. *Educational Psychologist*, 35(1), 25–37. https://doi.org/10.1207/S15326985EP3501_4
- Camacho, A., Alves, R. 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>
- Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children’s faces pain scales. *Pain*, 78(1), 27–37. [https://doi.org/10.1016/S0304-3959\(98\)00112-2](https://doi.org/10.1016/S0304-3959(98)00112-2)
- Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological bulletin*, 139(4), 735.
- Chen, Z., Lu, X., Ai, W., Li, H., Mei, Q., & Liu, X. (2018). Through a gender lens: Learning usage patterns of emojis from large-scale android users. *Proceedings of the 2018 world wide web conference*, 763–772.

- Chen, Z.-H., & Liu, W.-Y. (2019). A six-stage story structure approach for elementary students' story production: Quality, interest, and attitude. *Computer Assisted Language Learning*.
<https://doi.org/10.1080/09588221.2019.1688354>
- Ekman, P. (2006). *Darwin and facial expression: A century of research in review*.
- Gallo, K. E., Swaney-Stueve, M., & Chambers, D. H. (2017). Comparing visual food images versus actual food when measuring emotional response of children. *Journal of Sensory Studies*, 32(3), e12267.
- Goldshmidt, O. T., & Weller, L. (2000). "Talking emotions": Gender differences in a variety of conversational contexts. *Symbolic Interaction*, 23(2), 117–134.
- 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., Berninger, V., & Abbott, R. (2012). Are Attitudes Toward Writing and Reading Separable Constructs? A Study With Primary Grade Children. *Reading & Writing Quarterly*, 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., Schwartz, S. S., & MacArthur, C. A. (1993). Knowledge of writing and the composing process, attitude toward writing, and self-efficacy for students with and without learning disabilities. *Journal of learning disabilities*, 26(4), 237–249.
- Green, C. P., Nyhus, O., & Salvanes, K. (2022). How Does Testing Young Children Influence Educational Attainment and Well-Being? *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4275327>

- Grenner, E., Johansson, V., van de Weijer, J., & Sahlén, B. (2020). Effects of intervention on self-efficacy and text quality in elementary school students' narrative writing. *Logopedics Phoniatrics Vocology*, 1–10. <https://doi.org/10.1080/14015439.2019.1709539>
- Hall, A. H., & Axelrod, Y. (2014). "I Am Kind of a Good Writer and Kind of Not": Examining Students' Writing Attitudes. *Journal of Research in Education*, 24(2), 34–50.
- Hall, A. H., Boyer, D. M., & Beschoner, E. A. (2017). Examining Kindergarten Students' Use of and Interest in Informational Text. *Early Childhood Education Journal*, 45(5), 703–711.
- Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4(2), 201.
- Hall, L., Hume, C., & Tazzyman, S. (2016). Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. *Proceedings of the The 15th International Conference on Interaction Design and Children*, 311–321. <https://doi.org/10.1145/2930674.2930719>
- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: An intervention study. *Learning and Instruction*, 12(4), 429–446. [https://doi.org/10.1016/S0959-4752\(01\)00009-3](https://doi.org/10.1016/S0959-4752(01)00009-3)
- Hidi, S., & Boscolo, P. (2008). Motivation and Writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144–157). Guilford Press.
- Hier, B. O., & Mahony, K. E. (2018). The contribution of mastery experiences, performance feedback, and task effort to elementary-aged students' self-efficacy in writing. *School Psychology Quarterly*, 33(3), 408–418. <https://doi.org/10.1037/spq0000226>
- Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2016). Beyond the Rainbow: Retrieval Practice Leads to Better Spelling than does Rainbow Writing. *Educational Psychology Review*, 28(2), 385–400. <https://doi.org/10.1007/s10648-015-9330-6>

- Kahneman, D. (2011). *Thinking, fast and slow*. macmillan.
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Hachette UK.
- Kanala, S., Nousiainen, T., & Kankaanranta, M. (2013). Using a mobile application to support children's writing motivation. *Interactive Technology and Smart Education, 10*(1), 4–14. <https://doi.org/10.1108/17415651311326419>
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). American Council on Education; Praeger.
- Kim, J.-A., & Lorsbach, A. W. (2005). Writing Self-Efficacy in Young Children: Issues for the Early Grades Environment. *Learning Environments Research, 8*(2), 157–175. <https://doi.org/10.1007/s10984-005-7248-5>
- Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of personality and social psychology, 74*(3), 686.
- Kuebli, J., & Fivush, R. (1992). Gender differences in parent-child conversations about past emotions. *Sex Roles, 27*(11-12), 683–698.
- Liao, C. C. Y., Chang, W.-C., & Chan, T.-W. (2018). The effects of participation, performance, and interest in a game-based writing environment. *Journal of Computer Assisted Learning, 34*(3), 211–222. <https://doi.org/10.1111/jcal.12233>
- Liu, Y., & Hau, K.-T. (2020). Measuring motivation to take low-stakes large-scale test: New model based on analyses of “participant-own-defined” missingness. *Educational and Psychological Measurement, 80*(6), 1115–1144.
- Lundetræ, K., & Thomson, J. M. (2018). Rhythm production at school entry as a predictor of poor reading and spelling at the end of first grade. *Reading and writing, 31*(1), 215–237.
- Mata, L. (2011). Motivation for Reading and Writing in Kindergarten Children. *Reading Psychology, 32*(3), 272–299. <https://doi.org/10.1080/02702711.2010.545268>

- McTigue, E. M., Solheim, O. J., Walgermo, B., Frijters, J., & Foldnes, N. (2019). How can we determine students' motivation for reading before formal instruction? Results from a self-beliefs and interest scale validation. *Early Childhood Research Quarterly, 48*, 122–133.
<https://doi.org/10.1016/j.ecresq.2018.12.013>
- Merisuo-Storm, T. (2006). Girls and Boys Like to Read and Write Different Texts. *Scandinavian Journal of Educational Research, 50*(2), 111–125. <https://doi.org/10.1080/00313830600576039>
- Miller, D. M., McTigue, E. M., & Scott, C. E. (2015). The Quality of Recent Studies in Content-Area Writing in Secondary Classrooms. *Literacy Research: Theory, Method, and Practice, 64*(1), 461–477. <https://doi.org/10.1177/2381336915617602>
- Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). “blissfully happy” or “ready to fight”: Varying interpretations of emoji. *Proceedings of the international AAAI conference on web and social media, 10*(1), 259–268.
- Miller, S. D., & Meece, J. L. (1997). Enhancing Elementary Students' Motivation to Read and Write: A Classroom Intervention Study. *The Journal of Educational Research, 90*(5), 286–299.
<https://doi.org/10.1080/00220671.1997.10544585>
- Nuutila, K., Tapola, A., Tuominen, H., Kupiainen, S., Pásztor, A., & Niemivirta, M. (2020). Reciprocal Predictions Between Interest, Self-Efficacy, and Performance During a Task. *Frontiers in Education, 5*, 36. <https://doi.org/10.3389/educ.2020.00036>
- Pajares, F., Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. *Journal of Educational Psychology, 91*(1), 50–61.
<https://doi.org/10.1037/0022-0663.91.1.50>
- 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>

- Paquette, K. R. (2008). STUDY OF ELEMENTARY STUDENTS' ATTITUDES ABOUT WRITING AFTER A CROSS-AGE TUTORING EXPERIENCE. *Reading Improvement, 45*(4), 181–191.
- Paquette, K. R., Byrd, C. M., & Dragich, D. (2013). AN INVESTIGATION OF IMPLEMENTING DROP EVERYTHING AND WRITE (DEW) IN EARLY CHILDHOOD/ELEMENTARY CLASSROOMS. *Education, 133*(3), 356–369.
- Perry, N. E., Nordby, C. J., & VandeKamp, K. O. (2003). Promoting Self-Regulated Reading and Writing at Home and School. *Elementary School Journal, 103*(4), 317–338.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta psychologica, 104*(1), 1–15.
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Renninger, K. A., & Hidi, S. E. (2017). *The Power of Interest for Motivation and Engagement* (1st ed.). Routledge. <https://doi.org/10.4324/9781315771045>
- Schouteten, J. J., Verwaeren, J., Lagast, S., Gellynck, X., & De Steur, H. (2018). Emoji as a tool for measuring children's emotions when tasting food. *Food quality and preference, 68*, 322–331.
- Setty, J. V., Srinivasan, I., Radhakrishna, S., Melwani, A. M., & DR, M. K. (2019). Use of an animated emoji scale as a novel tool for anxiety assessment in children. *Journal of dental anesthesia and pain medicine, 19*(4), 227–233.
- Sick, J., Monteleone, E., Dinnella, C., Pierguidi, L., & Spinelli, S. (2022). Development of an emoji-based self-report measurement tool to measure emotions elicited by foods in preadolescents. *Food Quality and Preference, 104585*. <https://doi.org/10.1016/j.foodqual.2022.104585>

Sick, J., Spinelli, S., Dinnella, C., & Monteleone, E. (2020). Children's selection of emojis to express food-elicited emotions in varied eating contexts. *Food Quality and Preference*, *85*, 103953.

<https://doi.org/10.1016/j.foodqual.2020.103953>

Solheim, O. J., Torppa, M., Henning Uppstad, P., & Lerkkanen, M.-K. (2021). Screening for Slow Reading Acquisition in Norway and Finland – a Quest for Context Specific Predictors.

Scandinavian Journal of Educational Research, *65*(4), 584–600.

<https://doi.org/10.1080/00313831.2020.1739130>

Sprecher, S., & Sedikides, C. (1993). Gender differences in perceptions of emotionality: The case of close heterosexual relationships. *Sex roles*, *28*(9-10), 511–530.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, *15*(3), 335-352.

Sturgess, J., Rodger, S., & Ozanne, A. (2002). A Review of the Use of Self-Report Assessment with Young Children. *British Journal of Occupational Therapy*, *65*(3), 108–116.

<https://doi.org/10.1177/030802260206500302>

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>

Unrau, N. J., Rueda, R., Son, E., Polanin, J. R., Lundeen, R. J., & Murszewski, A. K. (2018). Can Reading Self-Efficacy Be Modified? A Meta-Analysis of the Impact of Interventions on Reading Self-Efficacy. *Review of Educational Research*, *88*(2), 167–204.

<https://doi.org/10.3102/0034654317743199>

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

<https://ggplot2.tidyverse.org>

ASSESSING WRITING AND SPELLING INTEREST AND SELF-BELIEFS

36

Wilson, K. M., & Trainin, G. (2007). First-Grade Students' Motivation and Achievement for Reading,

Writing, and Spelling. *Reading Psychology, 28*(3), 257–282.

<https://doi.org/10.1080/02702710601186464>

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>

Zumbrunn, S., Ekholm, E., Stringer, J., McKnight, K., & DeBusk-Lane, M. (2017). Student

Experiences With Writing: Taking the Temperature of the Classroom. *The Reading Teacher, 70*(6),

667–677. <https://doi.org/10.1002/trtr.1574>

Table 1

Seven items measuring interest and self-beliefs

Item	Description
Pre-test items	
1	General writing interest (label: INT_GE): “How much do you like to write?”
2	General writer self-concept (label: SC_GE): “How good are you at writing?”
Pre-task items	
3	Interest for spelling task (label: INT_PRE): “How much do you think you will like these tasks?”
4	Self-efficacy for spelling task (label: SE_PRE): “How good do you think you will be at completing these tasks?”
Post-task items	
5	Interest for spelling task (label: INT_POST): “How much did you like these tasks?”
6	Self-efficacy for spelling task (label: SE_POST): “How good do you think you were at completing these tasks?”
7	Continued interest for spelling task (label: INT_CONT): “Would you like to work with similar tasks some other time?”

Table 2

Univariate descriptives. Mean (SD) for item scores by format.

Grade	Item	Circle	Face	p-value
		N=1377	N=1314	
1	INT_GE	3.87 (1.37)	3.80 (1.37)	0.195
	SC_GE	4.05 (1.27)	4.05 (1.23)	0.984
	INT_PRE	4.00 (1.44)	3.99 (1.41)	0.783
	SE_PRE	4.22 (1.27)	4.20 (1.25)	0.695
	INT_POST	3.89 (1.51)	3.86 (1.47)	0.590
	SE_POST	4.23 (1.27)	4.17 (1.29)	0.198
	INT_CONT	3.93 (1.59)	4.01 (1.52)	0.240
		Circle	Face	p-value
		N=1456	N=1443	
3	INT_GE	3.45 (1.29)	3.41 (1.29)	0.460
	SC_GE	3.74 (1.11)	3.77 (1.11)	0.504
	INT_PRE	3.75 (1.26)	3.71 (1.24)	0.326
	SE_PRE	4.03 (1.10)	3.95 (1.09)	0.081
	INT_POST	3.65 (1.38)	3.64 (1.36)	0.825
	SE_POST	4.00 (1.16)	4.00 (1.14)	0.936
	INT_CONT	3.77 (1.55)	3.93 (1.43)	0.003

Table 3

Pearson correlations within and across waves.

Grade 1	INT_GE	SC_GE	INT_PRE	SE_PRE	INT_POST	SE_POST	INT_CONT
INT_GE		.56	.43	.35	.39	.29	.37
SC_GE	.56		.33	.35	.30	.35	.23
INT_PRE	.41	.30		.60	.66	.49	.59
SE_PRE	.31	.37	.63		.50	.61	.44
INT_POST	.41	.28	.71	.51		.60	.63
SE_POST	.32	.41	.49	.67	.56		.52
INT_CONT	.37	.25	.62	.45	.69	.49	

Grade 3	INTGE	SCGE	INTPRE	SEPRE	INTPOST	SEPOST	INTCONT
INTGE		.47	.44	.24	.45	.25	.39
SCGE	.48		.31	.38	.33	.36	.27
INTPRE	.41	.33		.47	.63	.33	.52
SEPRE	.24	.36	.45		.37	.50	.32
INTPOST	.40	.30	.65	.36		.48	.64
SEPOST	.20	.34	.35	.52	.48		.38
INTCONT	.34	.26	.55	.27	.65	.36	

(a) *Correlations within waves. Values for face and circle formats are above and below the diagonal, respectively.*

Grade	INT_GE	SC_GE	INT_PRE	SE_PRE	INT_POST	SE_POST	INT_CONT
1	.52	.46	.45	.39	.46	.35	.48
3	.65	.52	.52	.42	.52	.42	.51

(b) *Intra-item correlations across waves.*

Table 4

Number of respondents (n) and the percentage missing aggregated over seven motivation items

Grade	Format	Gender	Wave 1	Wave 2	Wave 1	Wave 2
			<i>n</i>		Missing %	
1	Circle	Boy	541	546	22	54
		Girl	510	600	23	55
	Face	Boy	546	541	27	63
		Girl	600	510	26	61
3	Circle	Boy	468	380	19	22
		Girl	440	452	19	21
	Face	Boy	380	468	20	23
		Girl	452	440	18	22

Note. Wave 1=First week when students took their first test booklet; Wave 2=Second week when students took their second test booklet.

Table 5

Grade 1 mixed logistic regression model for skipping with interaction between gender and format

<i>Dependent variable: Skip</i>		
Predictor	Estimate	Standard error
SC_GE (γ_1)	0.695***	(0.082)
INT_PRE (γ_2)	-0.472***	(0.087)
SE_PRE (γ_3)	-0.218**	(0.085)
INT_POST (γ_4)	-0.320***	(0.086)
SE_POST (γ_5)	-0.133	(0.085)
INT_CONT (γ_6)	-0.200**	(0.085)
Wave (β_W)	10.858***	(0.287)
Boy (β_G)	-0.141	(0.212)
Face (β_F)	0.414*	(0.218)
Boy*Face (β_{FG})	0.291	(0.310)
Constant	-13.517***	(0.375)
Observations	30,758	
Log Likelihood	-10,467.880	
Akaike Inf. Crit.	20,963.760	
Bayesian Inf. Crit.	21,080.430	

Note. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Reference for item effects is INT_GE=General interest;

SC_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy;

INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Table 6

Grade 3 mixed logistic regression model for skipping with interaction between gender and format

<i>Dependent variable: Skip</i>		
Predictor	Estimate	Standard error
SC_GE (γ_1)	-0.075	(0.089)
INT_PRE (γ_2)	-1.880***	(0.112)
SE_PRE (γ_3)	-1.644***	(0.107)
INT_POST (γ_4)	-1.718***	(0.109)
SE_POST (γ_5)	-1.617***	(0.107)
INT_CONT (γ_6)	-1.711***	(0.109)
Wave (β_W)	-2.502***	(0.548)
Boy (β_G)	0.101	(0.283)
Face (β_F)	-0.154	(0.269)
Boy*Face (β_{FG})	0.275	(0.392)
Constant	-1.530	(0.998)
Observations	24,360	
Log Likelihood	-7,150.937	
Akaike Inf. Crit.	14,329.880	
Bayesian Inf. Crit.	14,443.280	

Note. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Reference for item effects is INT_GE=General interest;

SC_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy;

INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Table 7

Grade 1 mixed logistic regression model for endorsing max, with interaction between gender and format

<i>Dependent variable: Endorse max</i>		
Predictor	Estimate	Standard error
SC_GE (γ_1)	0.345***	(0.074)
INT_PRE (γ_2)	0.788***	(0.073)
SE_PRE (γ_3)	1.200***	(0.075)
INT_POST (γ_4)	0.575***	(0.073)
SE_POST (γ_5)	1.249***	(0.075)
INT_CONT (γ_6)	0.790***	(0.073)
Wave (β_W)	0.142*	(0.078)
Boy (β_G)	-0.291**	(0.127)
Face (β_F)	-0.152	(0.098)
Boy*Face (β_{FG})	-0.050	(0.140)
Constant	-0.150	(0.140)
Observations		
	18,109	
Log Likelihood		
	-9,598.283	
Akaike Inf. Crit.		
	19,224.570	
Bayesian Inf. Crit.		
	19,333.820	

Note. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Reference for item effects is INT_GE=General interest; SC_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy; INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Table 8

Grade 3 mixed logistic regression model for endorsing max, with interaction between gender and format

<i>Dependent variable: Endorse max</i>		
Predictor	Estimate	Standard error
SC_GE (γ_1)	0.359***	(0.074)
INT_PRE (γ_2)	0.674***	(0.071)
SE_PRE (γ_3)	1.055***	(0.071)
INT_POST (γ_4)	0.803***	(0.071)
SE_POST (γ_5)	1.209***	(0.071)
INT_CONT (γ_6)	1.225***	(0.071)
Wave (β_W)	-0.119***	(0.045)
Boy (β_G)	-0.159	(0.098)
Face (β_F)	0.036	(0.058)
Boy*Face (β_{FG})	-0.146*	(0.084)
Constant	-1.290***	(0.103)
Observations	19,423	
Log Likelihood	-10,715.230	
Akaike Inf. Crit.	21,458.470	
Bayesian Inf. Crit.	21,568.710	

Note. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Reference for item effects is INT_GE=General interest;

SC_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy;

INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Table 9

Grade 1 mixed logistic regression model for endorsing minimum, with interaction between gender and format

<i>Dependent variable: Endorse min</i>		
Predictor	Estimate	Standard error
SC_GE (γ_1)	-0.680***	(0.136)
INT_PRE (γ_2)	0.413***	(0.119)
SE_PRE (γ_3)	-0.515***	(0.130)
INT_POST (γ_4)	0.843***	(0.117)
SE_POST (γ_5)	-0.318**	(0.128)
INT_CONT (γ_6)	1.481***	(0.116)
Wave (β_W)	-0.670**	(0.336)
Boy (β_G)	1.247***	(0.210)
Face (β_F)	-0.560***	(0.187)
Boy*Face (β_{FG})	0.592**	(0.253)
Constant	-4.664***	(0.481)
Observations	18,109	
Log Likelihood	-4,866.150	
Akaike Inf. Crit.	9,760.300	
Bayesian Inf. Crit.	9,869.558	

Note. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Reference for item effects is INT_GE=General interest;

SC_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy;

INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Table 10

Grade 3 mixed logistic regression model for endorsing minimum, with interaction between gender and format

<i>Dependent variable: Endorse min</i>		
Predictor	Estimate	Standard error
SC_GE (γ_1)	-1.404***	(0.140)
INT_PRE (γ_2)	-0.515***	(0.119)
SE_PRE (γ_3)	-1.673***	(0.143)
INT_POST (γ_4)	0.093	(0.113)
SE_POST (γ_5)	-1.320***	(0.134)
INT_CONT (γ_6)	1.200***	(0.108)
Wave (β_W)	0.377**	(0.182)
Boy (β_G)	1.703***	(0.188)
Face (β_F)	-0.692***	(0.149)
Boy*Face (β_{FG})	0.493***	(0.184)
Constant	-5.412***	(0.343)
Observations	19,423	
Log Likelihood	-4,341.395	
Akaike Inf. Crit.	8,710.790	
Bayesian Inf. Crit.	8,821.029	

Note. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Reference for item effects is INT_GE=General interest;

SC_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy;

INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Figure 1

Two types of pictorial support: Faces and circles



(a) Scale using face expressions ranging from happy to sad.



(b) Scale using circles ranging from large to small

Figure 2

Illustration depicting an elaborated view of writing

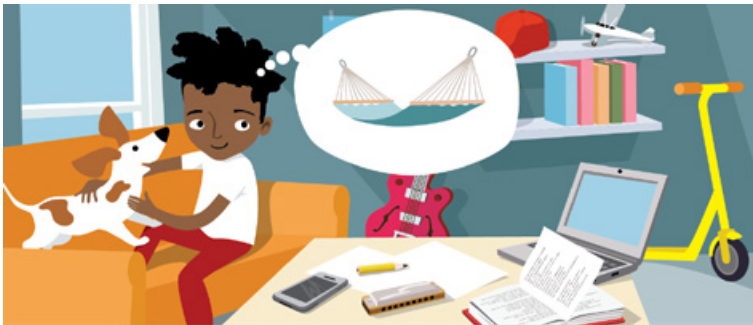


Figure 3

Authentic spelling task screens



(a) *Spelling task asking students to help a boy with writing his birthday wishing list.*



(b) *Spelling task asking students to spell the birthday wishing list item 'hammock'.*

Figure 4

Flipped design, grade 1

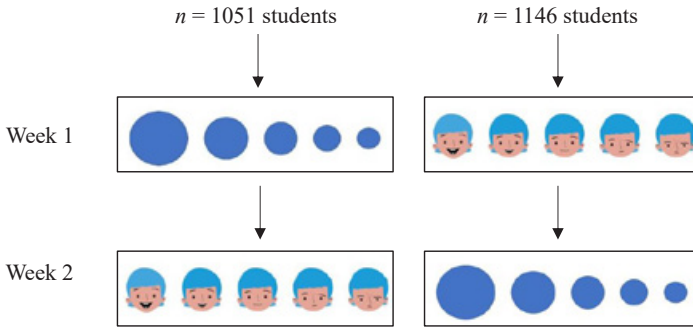
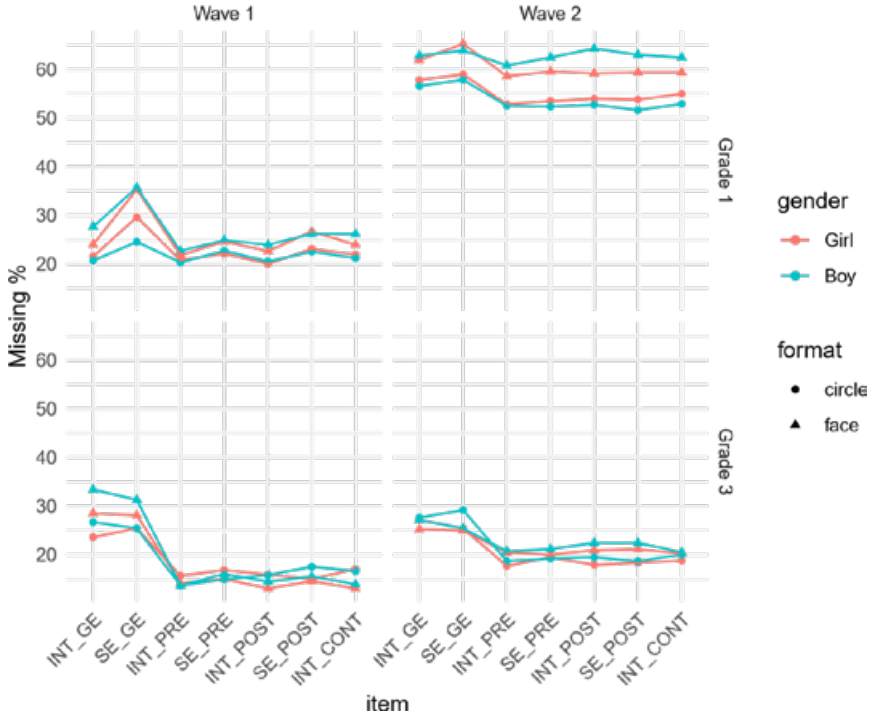


Figure 5

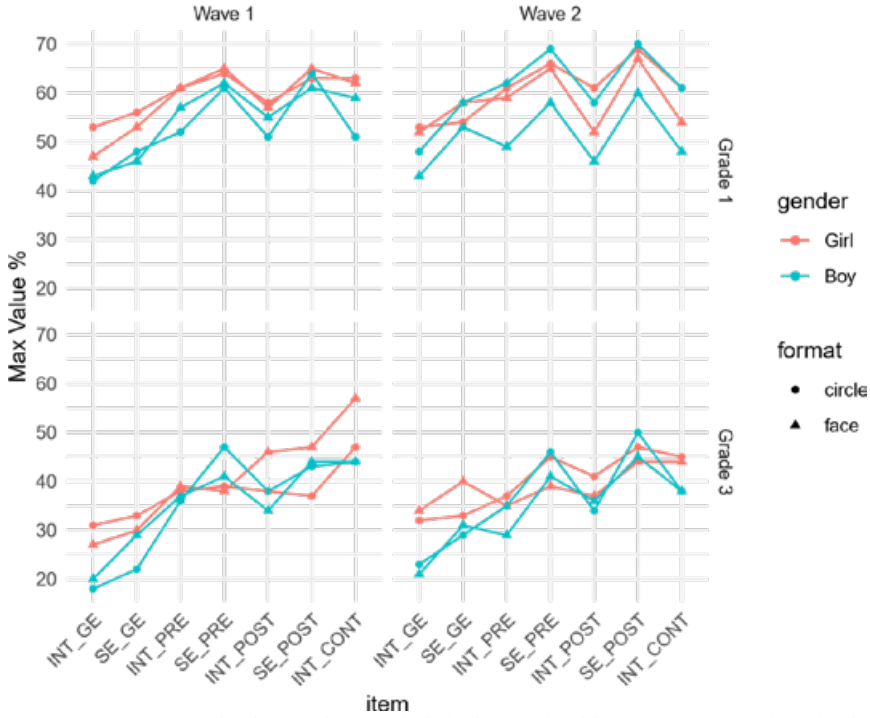
Percentage of skipping across grades, waves, gender, format, and items



Note. Wave 1=First week when students took their first test booklet; Wave 2=Second week when students took their second test booklet. INT_GE=General interest; SE_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy; INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Figure 6

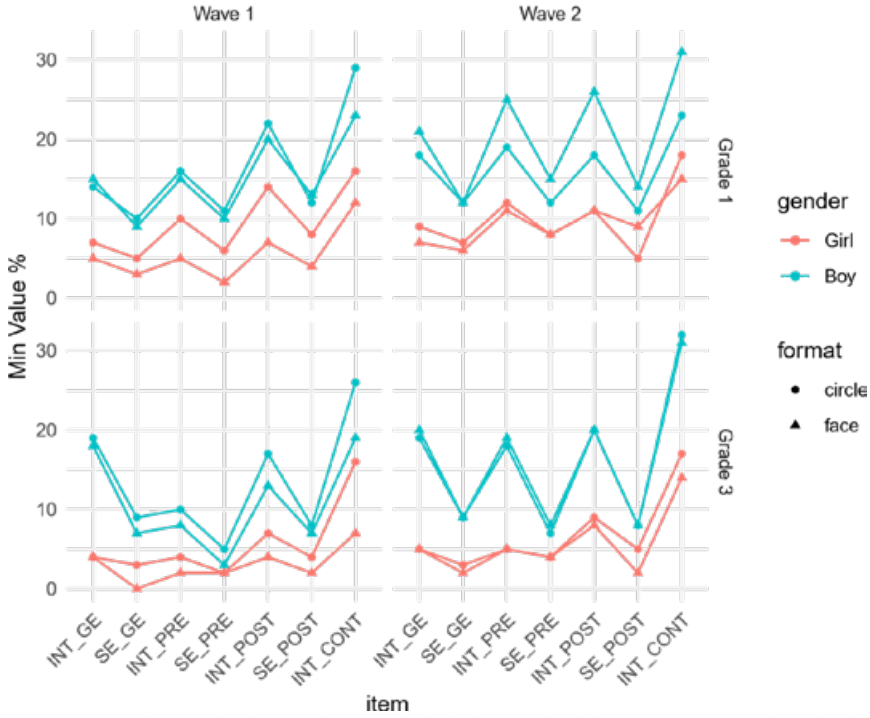
Percentage of students endorsing the maximum value across grades, waves, gender, format, and items



Note. Wave 1=First week when students took their first test booklet; Wave 2=Second week when students took their second test booklet. INT_GE=General interest; SE_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy; INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Figure 7

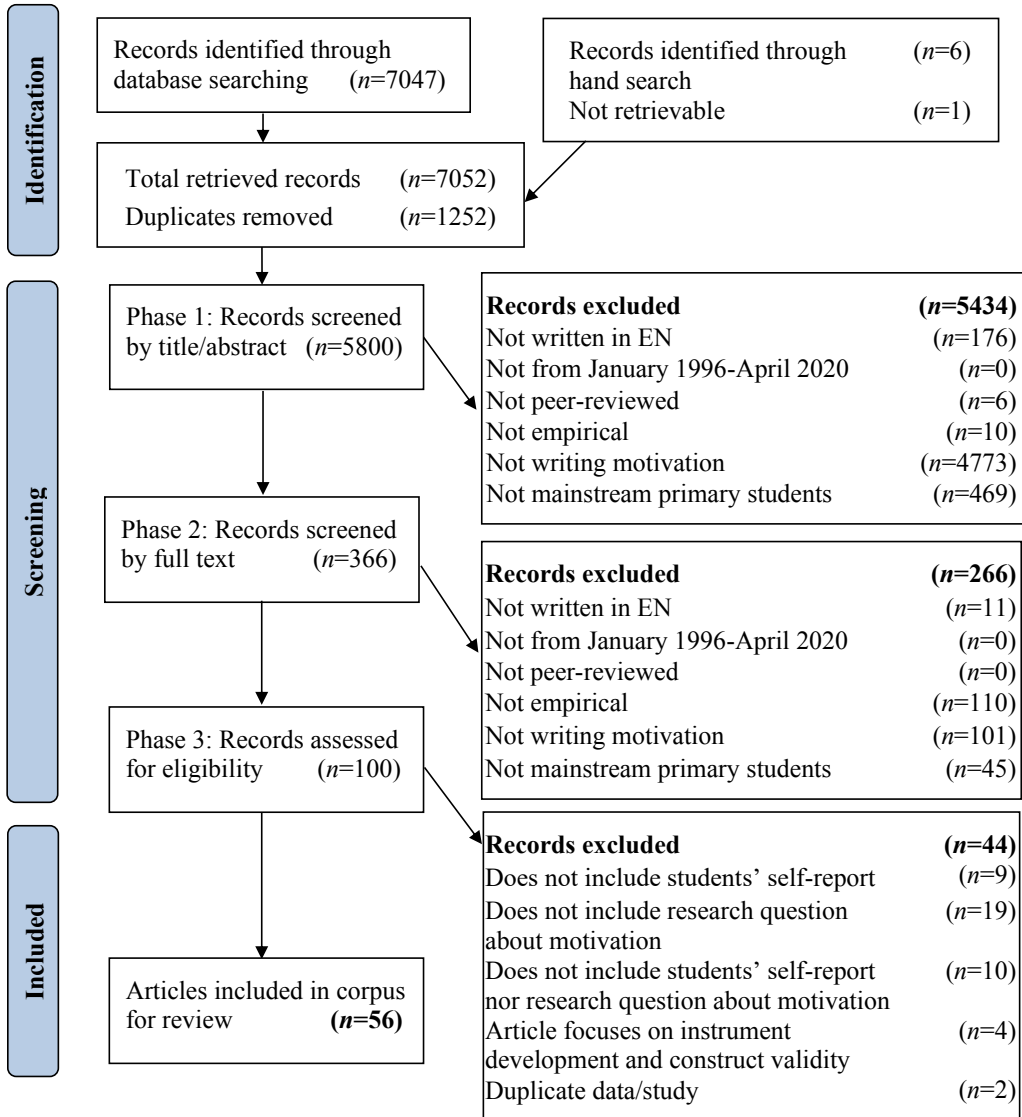
Percentage of students endorsing the minimum value across grades, waves, gender, format, and items



Note. Wave 1=First week when students took their first test booklet; Wave 2=Second week when students took their second test booklet. INT_GE=General interest; SE_GE=General self-concept; INT_PRE=Pre-task interest; SE_PRE=Pre-task self-efficacy; INT_POST=Post-task interest; SE_POST=Post-task self-efficacy; INT_CONT=Continued interest.

Appendix 4 – Supplementary material: Figure S1

Figure S1
Flow Diagram of Article Selection Process.



Adapted from: Page, M. J., McKenzie, J. E., Bossuyt, P. M, Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71

Appendix 5 – Supplementary material: Table S1

Table S1
Overview of Included Studies

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Abbott (2000)	USA	5th	2	Qualitative	Interviews	Intrinsic motivation	N/A	Interpretive study of two avid writers showed that they describe flow experiences similarly to older students and that the classroom social context influenced their intrinsic motivation to write.
Akyol & Aktas (2018)	Turkey	4th	185	Quantitative	Survey	Task value Self-efficacy (The survey uses the term <i>self-concept</i> , but the discussion uses <i>self-efficacy</i>)	Task: Story-writing Survey questions about narrative, expository and general writing	As the value placed on writing by students and their perception of themselves as competent writers increase, their story-writing scores also increase.
Andrzejczak et al. (2005)	USA	2nd and 3rd	2	Qualitative	Interviews	Motivation	Poetry, narrative, or expository writing	Visual art creation enhances the writing process and artwork as a pre-writing activity becomes a motivational entry point.
Babayigit (2019)	Turkey	4th	35	Quantitative	Survey	Attitude	Creative writing	Although creative writing activities led to an increase in reading, writing and language lesson attitude, only increase in language lesson attitude was significant.
Bayat (2016)	Turkey	4th	42	Quantitative	Survey	Attitude	Creative writing (letters, stories, tales, fables, poems)	Students receiving a creative writing program based on speaking activities increased their writing attitudes more than students in the control group.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Bayraktar (2013)	USA	5th	22	Mixed methods	Survey with entire sample for purposive sampling of 4 students to be interviewed	Self-efficacy Enjoyment	Student-teacher writing conferences about diverse genres (e.g., historical fiction, expository text, and persuasive essay)	Students' interactions differed during student-teacher writing conferences, depending on their levels of writing self-efficacy. Students with lower levels displayed weaker commitment to their tasks, and had difficulties in sharing their writings and seeking their teacher's help. Students with higher levels were more self-disciplined and committed to their writing goals, and were not hesitant in discussing their writings and seeking help when needed.
Beck & Fetherston (2003)	Australia	3rd	7	Qualitative	Interviews	Attitude	Story writing on paper and in a word processor (Students were given story starters)	Student self-reports indicated that the incorporation of a word processor influenced positively the students' enjoyment of the writing process as it facilitated the completion of writing tasks and the editing process.
Boscolo et al. (2012)	Italy	4th	114	Mixed methods	Survey and questionnaire with open-ended questions	Liking Self-perception of competence	Linguistic games and narrative writing	Students in an intervention teaching them to play with narrative texts reported that they liked writing more at posttest compared to students in a business-as-usual condition, however no effect on their perceived competence was found.

Study	Country	Grade level(s)	Student sample size	Methodology	Type of student self-report(s)	Motivation construct	Type of writing task	Main findings
Boyacı & Güner (2018)	Turkey	4th	46	Mixed methods	Survey and semi-structured interviews	Motivation	Authentic writing	The use of authentic task-based material improved the reading comprehension, writing skills and writing motivation of the students.
Bradford et al. (2016)	USA	1st and 2nd	32	Mixed methods	Survey including one open written response and a final reflection essay on the use of rubrics	Attitude	Opinion paragraphs in response to given writing prompts	Although an intervention focusing on the use of rubrics did not show a significant difference between conditions according to student attitude ratings, student-reported reasons indicated that the use of rubrics promoted positive attitudes toward writing later and higher self-efficacy.
Chen & Liu (2019)	China	5th	49	Mixed methods	Two questionnaires and semi-structured interviews	Attitude Interest	Story writing	Findings indicated that the 4S approach increased students' story quality, attitudes, and interest.
Chohan (2011)	Canada	1, 2, 3, 4, and 5	122	Mixed methods	Survey including open-ended questions	Attitude Perceptions of themselves as writers	Schoolwide letter writing	Although the number of students who indicated that they should write letters at school to each other almost doubled, the intervention did not show a significant impact on their attitudes toward writing.
Erdogan & Erdogan (2013)	Turkey	5th	594	Qualitative	Students' completion of the metaphor "writing is like..."	Perception about writing	Completion of metaphorical sentence: "Writing is like... because..."	Students' metaphors were sorted under 14 categories, like writing is 'joyful', 'informative', and 'developing', and results suggested that students mostly conceptualize writing as an activity that they like, rather than something boring.

Study	Country	Grade level(s)	Student sample size	Methodology	Type of student self-report(s)	Motivation construct	Type of writing task	Main findings
Gallini & Zhang (1997)	USA	4th and 5th	84	Mixed methods	Survey and discussion sessions	Confidence Self-efficacy	Writing about real-world science	Students who preferred group work enjoyed interaction and receiving input from peers. Higher audience awareness indicated greater orientation towards group work and the use metacognitive skills in writing, whereas students who preferred individual work showed higher self-efficacy rates and marginal significance in using metacognitive skills. Lower ratings of confidence in writing were more present among those with less audience awareness.
Graham et al. (2012)	USA	1st and 3rd	241	Quantitative	Survey	Attitude	Task: Personal narrative Survey questions about writing at school and at home	Students' ratings indicated that reading and writing are separable constructs, but some aspects like literacy interactions might not be. No statistically significant correlations between SES measures and attitude measures were found.
Grønner et al. (2010)	Sweden	5th	55	Quantitative	Survey	Self-efficacy	Personal narrative	Students' pre-test ratings indicated strong self-efficacy for narrative writing, which increased after intervention. Moderate correlations between self-efficacy and writing performance were found pre- and post-intervention.

Study	Country	Grade level(s)	Student sample size	Methodology	Type of student self-report(s)	Motivation construct	Type of writing task	Main findings
Göçen (2019)	Turkey	1, 2, 3, and 4	630	Quantitative	Three different surveys	Attitude Motivation	Creative writing	Findings indicated a positive effect of creative writing on students' creative writing achievement, writing attitude and motivation.
Hall & Axelrod (2014)	USA	K, 1, 2, 3, 4, and 5	81	Qualitative	Focus group interviews	Attitude Self-efficacy (including students' feelings and emotions for writing)	N/A	Students who perceived their teachers as enthusiastic in their practices and towards writing expressed positive writing attitudes. Students also expressed preferences regarding writing, including choice of topic, genre, writing for different purposes and audiences, and whether sharing their writings or not.
Hall et al. (2017)	USA	Kindergarten	41	Mixed methods	Interviews	Interest	Informational text	After using The Tools Approach strategy, children increased their abilities to identify and use informational text features. Their interest in reading and writing this type of text also increased after the intervention.
Hertz & Heydenberk (1997)	USA	Kindergarten	19	Mixed methods	Interviews	Attitude	Story writing	Findings indicate that process writing instruction increased students' writing skills, and that student motivation was positively influenced by the interactive format of the intervention.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Hier & Mahony (2018)	USA	2nd	117	Quantitative	Survey	Effort Self-efficacy	Narrative and expository writing	After participating in a performance feedback writing intervention, students reported higher levels of writing self-efficacy. Although effort was a significant predictor of writing self-efficacy, their experiences with task mastery and positive feedback were not.
Hillyer & Ley (1996)	USA	2nd	32	Qualitative	Interviews	Perceptions of themselves as writers	Different types of writing to be included in portfolios and self-assessments	A portfolio-based intervention showed positive effects on students' perceptions of themselves as writers; their ability to set goals for their literacy development, and to self-assess their writings and their progression toward their goals.
Imheideh (2015)	Jordan	Kindergarten	46	Mixed methods	Interviews	Attitude	Different types of functional writing (e.g., shopping lists, prescriptions, cards, and messages)	After the integration of writing activities into a dramatic play center with child-size furniture and real-life scenarios, such as grocery shops and doctor offices, children demonstrated significant improvement in their writing skills and attitudes.
Jones et al. (2016)	USA	1st and 2nd	42	Quantitative	Questionnaire	Interest	Spelling practice through rainbow writing and quizzing (retrieval practice)	Both retrieval practice and rainbow writing promoted student learning, but retrieval practice led to better results. Students indicated enjoying retrieval practice as much or more than rainbow writing.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Kanala et al. (2013)	Finland	5th	25	Mixed methods	Questionnaire with multiple choice and open-ended questions	Motivation	Writing in a mobile app	Results indicated that the use of a mobile application to write showed a moderate positive effect on the students' motivation to write.
Kholisyah et al. (2018)	Indonesia	5th	18	Mixed methods	Survey and interview	Achievement motivation	Explanatory text	Results indicated that outstanding students display high levels of writing achievement motivation and that their motivation is influenced by individual (e.g., competence and beliefs) and environmental factors (e.g., competitive environments, parent support and teachers).
Kim & Lorschach (2005)	USA	K to 1st	18	Qualitative	Interview	Self-efficacy	Story writing by completing a sentence like "To be a good friend, I can..."	Despite their young age, children were able to indicate their levels of writing self-efficacy. Their ratings mostly showed consistency with teachers' and researchers' ratings, and differences seemed to be related to the students' definitions of writing.
Lee & Enciso (2017)	USA	3rd	29 classrooms	Quantitative	Survey	Self-efficacy	Story writing following a story spine	An intervention with teaching artists from a theater increased students' self-efficacy for pre-writing and also the students' skill and confidence in idea generation and revision.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Leroy (2000)	Canada	5th	1	Qualitative	Interview	Attitude	Creative and journal writing	Although the girl in the study expressed enjoyment in writing poems and stories in class, and believed that the purpose of these was to entertain her readers, she complained that writing responses to her teacher about the books and texts she had read ruined the joy of the reading activity.
Li & Chu (2018)	China	4th	109	Mixed methods	Questionnaire, individual interviews, and focus group interviews	Motivation	Collaborative writing in a wiki	Students were mostly positive about the collaborative process writing through wikis, but reported on learning difficulties related collaborations and the wiki-platform.
Liao et al. (2018)	Taiwan	3rd	245	Quantitative	Survey	Individual and situational interest	Writing in a game-based environment vs. online-based environment	Students writing in a game-based environment showed greater situational interest and were more motivated to generate ideas and write longer articles than students writing in an online-based environment.
Mata (2011)	Portugal	Kindergarten	451	Quantitative	Survey	Enjoyment Value Self-concept	Survey questions about writing stories	Although students indicated high levels of motivation for both reading and writing, motivation levels for reading were higher. Reading and writing value scores were higher than enjoyment and self-concept.
Meristo-Strom (2006)	Finland	4th	145	Quantitative	Survey	Attitude	N/A	Students reported more negative attitudes towards writing than towards reading.

Study	Country	Grade level(s)	Student sample size	Methodology	Type of student self-report(s)	Motivation construct	Type of writing task	Main findings
Miller & Meece (1997)	USA	3rd	187	Quantitative	Survey	Goal orientations Cognitive engagement Anxiety Self-concept	Complex writing tasks	In classrooms where students had more opportunities to work with challenging writing tasks, write collaboratively, and self-monitor their progress over time, they demonstrated more positive motivational patterns.
Miller & Meece (1999)	USA	3rd	24	Mixed methods	Interview where students answered a survey by rating their answers from 1 to 10, and were also asked open-ended questions to explain their answers	Performance expectancies Task value	High-challenge tasks (e.g., essays on a topic of students' choice, research papers on an animal for an interdisciplinary science unit, a character analysis for a class novel, letters to politician describing election concerns, and letters to the next year's third-grade class explaining what students needed to know in their new grade) and low-challenge tasks (e.g., worksheets on unstressed vowels, pronouns, simple contractions, and vocabulary, spelling, and handwriting exercises)	In general, students preferred high-challenging writing tasks, rather than low-challenging, as these were seen as boring and not intellectually challenging. Those who did not have enough opportunities to work with challenging tasks, questioned whether they had the necessary skills.
Nair et al. (2013)	Singapore	5th	224	Mixed methods	Online survey with entire sample and interviews with 18 students	Motivation	Expository text and blogging	Although most of the students agreed that blogging was useful, important and interesting, both students and teachers tended to take traditional paper-based writing assignments more seriously than online-based.

Study	Country	Grade level(s)	Student sample size	Methodology	Type of student self-report(s)	Motivation construct	Type of writing task	Main findings
Nicolaidou (2012)	Cyprus	4th	146	Mixed methods	Two different surveys with entire sample and semi-structured interviews with 9 students	Self-efficacy	Process portfolios	Results indicated that portfolio affordances connected to sources of self-efficacy, like verbal persuasion, and mastery and vicarious experiences, influenced positively the students' writing self-efficacy.
Nolen (2007)	USA	1, 2, and 3	67	Mixed methods	Interviews	Interest Motivation	N/A	Results indicated that the classroom social context played an essential role for the children's motivation to read and write.
Paquette (2008)	USA	2nd and 4th	85	Mixed methods	Survey and interviews	Attitude	Various writing tasks	Survey results did not indicate significant differences between treatment and control groups. However, even though second-graders lacked language skills to express themselves accurately, interview answers from second and fourth graders indicated that students appreciated the cross-age tutoring program, despite collaboration issues.
Paquette et al. (2013)	USA	K, 1, 3, 4, and 5	133	Mixed methods	Survey including 3 open-ended questions	Attitude	Various writing tasks during Drop Everything and Write sessions, including personal written communication (note passing), free writing (creative, personal narrative, expository writing), and prompt (personal response to an event or literature).	Although teachers' observations of students' behaviors indicated that the implementation of Drop Everything and Write was successful, students' survey data were contradictory.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Perry (1998)	Canada	2nd and 3rd	94	Mixed methods	Questionnaire including open-ended questions and interviews	Control Support Beliefs Values Expectations	N/A	Students in classroom contexts characterized by high self-regulated learning demonstrated more positive motivational orientations.
Perry et al. (2003)	Canada	1st	2	Mixed methods	Interviews	Motivation	Task: Story writing Interview questions about story writing and spelling.	Although both students indicated constructive views on failure, the high achieving student was more negative towards errors. Both had strategies for coping with challenging writing tasks, but the lower achieving preferred less challenging tasks.
Pollington et al. (2001)	USA	4th and 5th	130	Quantitative	Survey	Self-perception	Writing during workshop vs. traditional writing instruction	There were no significant differences on the students' scores on writer self-perception between students receiving writing workshop instruction and those receiving traditional instruction. Findings suggest that teachers are more decisive in affecting the students' writer self-perception.
Schrodt et al. (2019)	USA	Kindergarten	27	Mixed methods	Interviews, survey, and recorded conversations during measurement tasks	Motivation (including questions about, attribution, competence and difficulty, self-efficacy, value and enjoyment, and mindset)	Enhanced writer's workshop	Quantitative and qualitative results indicated that adding instruction on mindset and self-regulation to the writer's workshop improved the students' writing motivation and independence.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Seban (2012)	Turkey	3rd	42	Quantitative	Two different surveys	Attitude Self-perception	Picture and poem books	The authoring cycle program had a positive effect on the experimental group's attitudes toward writing, but no significant effect on their self-perception as writers, compared to the control group. No significant differences were found between the groups' writing achievement, as measured in terms of appearance, organization, development, language, and length of composition.
Seban & Tavsani (2015)	Turkey	2nd	27	Qualitative	Interviews	Students' perceptions about themselves as writers	Writer's workshop	Students' participation in writing workshops seemed to influence their writer identities.
Sessions et al. (2016)	USA	5th	30	Qualitative	Interviews	Attitude	Story writing using picture books as model text	Both students receiving traditional instruction and students working with iPad apps increased their writing skills. However, students using iPad apps produced better texts, experienced the writing activity as more socially engaging, and showed higher levels of motivation to write.

Study	Country	Grade level(s)	Student sample size	Methodology	Type of student self-report(s)	Motivation construct	Type of writing task	Main findings
Silver & Lee (2007)	Singapore	4th	33	Mixed methods	Survey including two open-ended questions	Attitude	Two 120-word narrative compositions (with the topics 'A robbery in the park' and 'An outing to the kelong'. Each topic had three pictures to guide students in their writing, but they were required to come up with their own conclusions)	Teacher feedback characterized by advice encouraged the most revision. However, most of the revisions were judged to be minimal, and only 10% of the revisions were considered successful. Students preferred praise and advice, but although they disliked criticism, this feedback type also led to revisions.
Snyders (2014)	USA	Kindergarten	3	Qualitative	Interviews and videotaped student-teacher conferences	Self-efficacy Writer identity	Book writing during workshop	Prior writing experiences and participation in writing workshops where students could associate themselves with the authors explored during mini-lessons influenced the students' identities as writers.
Teague et al. (2010)	Mexico	1s and 4th	6	Qualitative	Interviews	Attitude	Various writing tasks (authors noticed many copious school writing including dictations, copying, and simple texts)	Results indicated that home and community texts were seldom included in classroom instruction, and that writing attitudes were often associated with expectations of correctness and neatness.
Truax (2018)	USA	2nd and 3rd	56	Mixed methods	Interviews, survey, students' feedback on exit slips, and recorded teacher-student conferences	Growth mindset Self-concept Value	Writing workshop	Although quantitative results did not show significant effects, qualitative findings indicated that growth mindset feedback combined with objective compliments increased students' writing motivation.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Tunks (2010)	USA	5th	215	Quantitative	Survey	Attitude	Descriptive writing Narrative writing Expository writing	Students' motivation to write decreased after preparing for and taking the state-mandated exam.
Unal (2010)	Turkey	4th and 5th	2315	Quantitative	Survey	Writing disposition (divided into three subcategories: confidence, persistence, and passion)	N/A	Results indicated that students' writing disposition levels are positive, but that the writing disposition levels among schools vary.
Wilson & Trainin (2007)	USA	1st	198	Quantitative	Survey	Attributions Perceived competence Self-efficacy	Survey questions about story writing	Results suggested that students can differentiate their self-efficacy for reading, writing, and spelling and that literacy attributions mediate between achievement and self-efficacy.
Zumbrunn et al. (2017)	USA	5th	114	Qualitative	Drawings of a recent writing experience and a written description of the drawings	Perception of themselves as writers Beliefs about writing	N/A	Results indicated that students expressed both positive (e.g., joy) and negative (e.g., apathy, anxiety) emotions in relation to writing. Often, drawings illustrating positive experiences depicted teachers near the students, whereas drawings illustrating negative experiences depicted the teacher further away. Students also expressed feeling more motivated to write when confident in their writing abilities and allowed choice opportunities.

<i>Study</i>	<i>Country</i>	<i>Grade level(s)</i>	<i>Student sample size</i>	<i>Methodology</i>	<i>Type of student self-report(s)</i>	<i>Motivation construct</i>	<i>Type of writing task</i>	<i>Main findings</i>
Zumbrunn et al. (2019)	USA	5th	263	Mixed methods	Survey including two open-ended questions	Students' writing enjoyment and aversion Students' perceptions of teacher writing enjoyment	Survey questions about writing in general	Quantitative findings indicated a positive relationship between students' perceptions of teacher writing enjoyment and their own writing enjoyment. Qualitative findings suggested that their writing enjoyment and aversion is influenced by their writing preferences, as well as their mood and motivation.

