Analysing the characteristics of texts written by Irish and Norwegian primary school students

Mette van Hell

Bachelor thesis Psychology

5th of July 2023

Faculty of Behavioural, Management and Social Sciences (BMS)

University of Twente

First Supervisor: Dr. J. Steinrücke

Second Supervisor: Dr. M. Amir Haeri

Introduction

Recent studies into the nature of reading among primary and secondary school students show a decline of reading motivation over the past ten years (OECD, 2019) PISA examines literacy and competence of 15-year-old students in three domains, which are reading, mathematics, and science. Its worldwide assessment of student learning and contributing factors is considered to be a highly impactful study for policymakers across countries, because of the insight it provides into the variations across countries for different school subjects. A first finding of the study is that digitalisation has caused text formats and reading mediums to change and develop (OECD, 2019). Reading habits have changed as well. Compared to the 2009 results, it is notable that considerably more students viewed reading as ‘a waste of time’ in PISA’s 2018 survey. Additionally, the study shows that students spend less time reading for leisure. Whereas they still read for practical reasons, the custom of reading books, magazines, and newspapers because students want to do so out of enjoyment, has dropped.

The decline in reading motivation and competence is alarming. The complexity of the digitalised world requires students to go beyond simple reading comprehension, which is illustrated by the ability to merely memorise and collect information. Rather, Binkley et al. (2011) highlight the need for students to be able to communicate, share, and use information to solve complex problems. This is reflected in the change in text which students read, according to PISA’s 2018 results. Now more than ever, reading requires complex information-processing strategies such as integration, analysis, synthesis, and evaluation of multiple sources of information (OECD, 2019). Elliot (2017) concludes that computers are already capable of solving the reading challenges that students with a low reading proficiency can take on. Because of this, it becomes evident that there is a need for high reading competence among students across the globe. To respond to changing demands in society, reading literacy is crucial for creating knowledge and expanding productivity. Reading literacy is not only related to academic achievement in other school subjects such as mathematics and science, but it is also a prerequisite for successful participation in personal and professional domains of adult life (Cunningham and Stanovitch, 1997). A strong foundation of literacy is needed from the early years of schools, to ensure the student's success in later education and professional life afterwards.

Reading motivation and engagement

Since the demands for reading and writing competence are high, researchers are examining ways to improve literacy among students. Experimental research has examined motivation as a predictor for literacy achievements (Graham et al., 2007; De Smedt, 2016; Teng & Zhang, 2018; Rogiers et al, 2020). Their findings highlight a crucial role for motivation in enhancing reading and writing skills among students. According to Fredricks et al. (2004), the construct of motivation is characterised by a willingness to learn and master a subject, as well as self-regulating behaviour and the active use of learning strategies. A motivated reader is therefore more likely to also be an engaged reader.

Apart from the motivation of students for a school subject or activity, differences in the type of motivation should also be considered. A leading theoretic framework which can be used for describing motivation is the Self-Determination Theory by Ryan and Deci (2000). SDT is a macro-theory of motivation and development that has proven to be especially interesting for application to educational practice (Niemec & Ryan, 2009). Ryan and Deci (2000) present a continuum of types of motivation. The two ends of the continuum are comprised of amotivation (e.g., lack of motivation to read) and intrinsic motivation (e.g., reading because of inherent enjoyment of this activity). In between, the SDT model differentiates between four types of extrinsic motivation. The first two types are classified as controlled motivation. These are external motivation (e.g., reading because of perceiving external pressure) and introjected regulation (e.g., reading because of perceiving internal pressure). The latter two types categorised as autonomous motivation, namely, identified regulation (e.g., reading because of personal importance of these activities), and integrated regulation (e.g., the activity of reading is congruent with the needs of the individual) (Ryan & Deci, 2000). Few studies have considered literacy from a SDT perspective so far, but De Naegel and colleagues (2012) have found that students in possession of autonomous types of motivation show better reading comprehension. Additionally, students who enjoy reading often, become more skilled compared to students who do not enjoy reading (Bates et al., 2016). Similar outcomes are found for writing (De Smedt et al. 2016). In conclusion, the importance of fostering autonomous motivation to enhance literacy becomes evident.

Looking at reading motivation from a SDT perspective offers the opportunity to explore the factors that influence motivation. Effectively, these factors can help teachers and other stakeholders in education to improve autonomous motivation in students. Ryan and Deci (2000) take an SDT perspective. They argue for three basic psychological needs which must be met for individuals to foster growth and personal well-being. These needs entail autonomy, competence, and relatedness. Autonomy is described by having a choice in one's own behaviour and self-endorsing this. Competence refers to the confidence and effectiveness that an individual portrays in their actions. Lastly, relatedness refers to the experience of feeling connected to others and being accepted by them (De Naegel et al., 2014).

Identifying factors that influence motivation

Various studies have found influencing factors of motivation to increase autonomous motivation, and they can be linked to the psychological needs proposed by Ryan and Deci (2000). With regards to teacher support, Guthrie and colleagues (2012) highlighted the importance of teachers supporting the autonomy of students. According to Renniger (2000), teachers function as an 'expert other' to generate interest in a subject and provide scaffolding, which leads to motivated behaviour. Additionally, peer collaboration leads to higher engagement (Fredricks et al., 2004; Miller & Meece, 1997; Lutz et al., 2006; Hidi & Renniger, 2006), emphasising the role of relatedness in student learning. This is supported by the success of educational practices based on Johnson and Johnson's social interdependence theory (2009). By providing reading texts that are relevant to the students, as well as communicating that the texts hold personal value to them, autonomous motivation increases (Guthrie et al., 2012). Contrary to what some may assume, students prefer complex tasks over easy tasks, especially when they are used to them. (Miller & Meece, 1999).

CORI is an example of a reading instruction programme which sought to include these factors to enhance motivation. Guthrie and Wigfield (2000) developed Concept-Oriented Reading Instruction for increasing reading engagement and comprehension in elementary school. CORI includes scientific topics, teaching different reading strategies and various instructional practices. It is based on the idea that that high engagement leads to high comprehension, and targets autonomous motivation through its instructional properties. Wigfield and colleagues (2008) studied the effectiveness of CORI compared to traditional reading instruction. A quasi-experiment with fourth-grade students (9- to 10-year-old) was conducted, providing a different type of reading instruction for each class. Significantly higher reading comprehension, engagement and strategy use were found for students who received concept-oriented reading instruction (Wigfield et al. 2008). Therefore, it can be concluded that learning outcomes will improve when educators ensure to employ autonomous motivation-enhancing strategies in their instructional practices.

Digital learning environments

The rapid digitalisation of society brings about a shift in competences that students need to succeed in their later life. A gradual shift of paper-based to digital-based learning methods takes place in education as well. Digital learning environments offer features that can be advantageous for improving engagement in reading education. For instance, digital education can offer meaningful and relevant contexts, interaction, and can be tailored to the individual learner (Vanbecelaere et al., 2020). To guide the development of (digital) learning environments, Downes (2010) proposes that learning environments should maximise learner autonomy, openness, interactivity, and diversity. These criteria can be linked back to the psychological needs from the before-mentioned SDT (Ryan & Deci, 2000). Thus, digital learning environments can improve student motivation and engagement. When effective design of learning environments is applied to reading literacy, it can be expected that reading comprehension will increase consequently.

Additionally, digital learning environments provide the unique feature of adapting to the individual student. For instance, computer-assisted education makes it possible to tailor instruction to the needs and characteristics of the student (Murphy, 2019), as well as give fitting feedback. The latter is especially attractive for educational practice (Martin, 2020) since it supports teachers in the instructional process. Research also suggests that adaptive feedback improves a student's motivation and self-regulation (Koenka & Anderman, 2019; Ouyiang & Jiao, 2021). Maier and Klotz (2022) argue for consideration of AI-based machine learning techniques for giving adaptive feedback. Some of the advantages include predicting student success and providing feedback on this basis, as well as detecting so-far-unknown behaviour patterns. An adaptive feedback approach ensures that every student receives tailored support so that they can progress their literacy competence. Additionally, adaptivity of the learning environment is important for adequately responding to the interests of the student. By offering students reading material that aligns with their interest, they will be more likely to be motivated for engaging in reading.

Taking all of this into account, digital learning environments offer number of features which influence reading motivation and engagement, and by extension, literacy competence.

For the digital learning environment to be effective, it is essential to take individual differences between students into account and play into this by giving tailored feedback and tailored reading material. Text mining can be a useful approach to achieve this, since it can aid in mapping variations, but also common features of student-written texts. One way in which this can be done is looking at the extent of positivity in a text by coding sentiments of the words. Text length is another characteristic which can tell something about the writing styles of students. The knowledge gained from this will help design a digital learning environment that can optimally play into the student's writing behaviour.

Aim of the thesis

The aim of this thesis is to explore differences between writing texts of primary school students from two European countries. By means of computational text and sentiment analysis, variations in genre and sentiment as well as differences between countries and age can be identified. Since the data used is the first available data from the AILit project, this thesis is the first attempt to study the texts retrieved from students. Therefore, this study means to answer some exploratory questions regarding the properties of the data. Ultimately, this thesis will create more insight into how adaptive digital environments can respond to student interest and characteristics to increase literacy engagement.

Methods

For the study, students writing assignments were analysed to find variations across age, countries, topics, and sentiments. Data was collected as part of an ongoing European research project (AILIT), which aims to use Artificial Intelligence to promote literacy among primary school students and shared with the researcher.

Sample

The sample of this study was obtained through the AILIT project. The participants of this study include 158 students from Ireland and 43 students from Norway, whose schools are stakeholders in the AILIT research project. The student's ages ranged from 8 to 12 years of age. Each of the student wrote one text, on which the text analysis was performed. This brings the total number of written texts to 201. Each of the students wrote the text in their native language, which were English and Norwegian, respectively. Using Google's translation API, the Norwegian texts were translated into English to facilitate effective comparison and analysis. Students received varying guidelines about their writing assignments. It is unknown what exact instructions were given to the Irish students, but it can be inferred from the dataset that each class received a different writing prompt. These include writing a procedure (e.g., 'how to become an artist'), a report, a poem, or a fictional story or narrative (e.g., about Christmas). For instance, the Norwegian students were instructed to write a story which included the word 'liker' (to like) as the second word in the sentence. The texts were afterwards categorised into genres.

Since the study uses data from minors, their parents were all informed about the purpose of the research and asked for consent to collect and use their child's data.

Two texts were removed from the Irish dataset, because characteristics such as writer ID, genre and content were missing. This brings the total number of entries to 199. Additionally, a written texts did not contain any words with sentimental value according to the AFINN lexicon. After removal, the sample consisted of 191 texts. Additionally, 24 outliers were removed by means of the Interquartile Range method. This resulted in 167 entries deemed fit for analysis.

Data Pre-processing

The data used in this study contains elements such as punctuation, capitals, and names. Contrary to humans, machine learning algorithms struggle to differentiate between meaningful and meaningless textual elements. For this reason, it is necessary to pre-process the data thoroughly. This includes standardising the text by removing texts, characters and symbols that are irrelevant for the analysis (Danubianu, 2015). This will ensure that the quality and performance of the analysis is higher. To pre-process the data correctly, a few steps were taken:

1. Estimated ages were standardised for both countries and remarks were removed. This creates the opportunity for comparing subgroups within the data.

2. The data was *tokenised* by transforming sentences into single words. Since R cannot derive meaning from sentences, tokenisation is needed for the programme to ‘understand’ the text data.

3. The data was *filtered* by removing punctuation marks, uppercase marks, and most importantly, stop words. By removing stop words, words are removed from which no meaning can be drawn, such as ‘the’, ‘he’, and ‘are’, as well as person names. To remove stop words, an open-source dictionary called ‘stop words-iso’ (Diaz, 2016) was used, as well as a dictionary with stop words which were compiled by the researcher, including numbers, person names, and some remarks in Danish. The Norwegian data was also filtered to exclude the word *like,* since children were assigned to start their text with this word. Because of this, the word did not hold any meaning and needed to be removed.

Data analysis

The data analysis is conducted in the open-source statistical program R (R Core Team, 2022). The following R packages were used: *tidyverse*, *tm*, ggplot2, *janitor, lubridate, ggrepel, tidytext, stopwords, textdata, reshape2,* and *emmeans.* For removing the most common stop words from the dataset, the collection of stop words from the Stopwords ISO (Diaz, 2016) was used. This open-source collection is one of the most complete and reliable collections of stop words. For supplementing this list of stop words, an additional list was created. For sentiment analysis, the AFINN dictionary (Nielsen, 2011) was used. Sentiments are measured by coding words according to their associated positivity or negativity. These values range from -5 (very negative words) to 5 (very positive words). Consequently, the sum of the sentiments per text can provide information about the positivity or negativity of the text. Outliers for the variable *total words* were identified and removed by using the inter-quartile range method. Descriptive statistics were computed to give insight into the dataset. The two countries were compared where possible, using Welch's two sampled t-tests check for significant differences. Two ANOVA models were compared to explore how the variation in *total words* could be explained. A simpler model was compared to a model with added covariates. To test the effect of *genre* on *sentiment*, an estimated marginal means (EMM) analysis was performed to compare contrasts between genres. This type of analysis was useful for exploring the association between *genre* and *sentiment* because the average means were quite close to each other. Since the means for *total words* were more dispersed, the conclusion from the EMM analysis would be similar to what was already found, so the use of an EMM analysis was deemed unnecessary. Lastly, a visual exploration by means of boxplots was done to examine the effect of *estimated* age on *total words.*

Results

Is there an association between genre and total words?

Between the six genres of texts, variations of total words used in the texts can be seen in Table 1. The genre 'Narrative' averages the highest number of words (*M* = 206.86), and the genre 'Poem' has the lowest number (*M* = 55.63). There were two genres for which the countries could be compared. A Welch Two Sample t-test was conducted to test the following null hypothesis: 'there is no significant difference in total words between the two countries'. For the genre 'Personal', the following t-statistic was reported: $t(18.73) = -2.447$, $p = .024$. For the genre 'Fiction', the following t-statistic was reported: $t(15.57) = -5.597$, $p < .001$. For both scenarios, the null hypothesis should be rejected. It appears that the Norwegian students wrote shorter texts than the Irish students.

Table 1. *Total words used per genre.*

Genre	M	$M_{Ireland}$		M_{Norway}	
		Mean	SD	Mean	SD
Narrative	206.86	206.86	40.82		
Procedure	104.64	104.64	48.27		
Report	92.57	92.57	21.98		
Fiction	90.11	148.55	23.38	71.21	36.63
Personal	56.86	66.08	30.49	44.56	18.28
Poem	55.63	55.63	21.09		

Two models were tested to find the variables that were best able to predict *total words*. The first ANCOVA model included the predictor *genre*, as well as the covariates *sentiment*, and *estimated age*. This model can be seen in Table 2. The second ANOVA model (Table 3) included only the predictor *genre*. In both models, *genre* has a highly significant effect on *total words*. For Model 1, $F(5,144) = 46.17$, $p < 0.001$, and for Model 2, $F(5,161) = 43.31$, $p < 0.001$. In Model 1 (see Table 2), the variables *sentiment* ($p = 0.73$) and *estimated age* ($p = 0.32$) did not have a significant effect on *total words*. An Akaike's criterion test was carried out to test the performance of the models. Model 1 performed slightly better (1515.6, $k = 6$) compared to Model 2 (1675.8, $k = 2$). Similarly, for Model 1, $R^2 = .62$, $F(7,144) = 33.31$, $p < 0.001$ and for Model 2, $R^2 = .57$, $F(5,161) = 43.31$, $p < 0.001$. Based on the outcomes for Akaike's criterion test and the R-squared, it can be said that Model 1 performs better at explaining the variation in *total words*.

Table 2: *Model 1 - ANCOVA model including the predictor 'genre' and two covariates.*

Variable	Df	F-value	P
Genre	5	46.17	<.001
Sentiment	1	0.12	0.73
Estimated age	1	0.97	0.33
Residuals	144		

Table 3: *ANOVA model including the predictor 'genre'.*

	Df	F-value	P
Genre	5	43.31	<.001
Residuals	161		

How do sentiments differ per genre and country?

To examine the differences in sentiments per genre, the sentiment means were calculated for every genre. Sentiments means are the sum of the numeric values which were assigned to words in the texts. The values range from -5 for words with a negative association, to 5 for words with a positive association. In this way, the negativity or positivity of a text can be calculated. The outcomes in Table 4 show that the genre 'Poem' was reported to be the most positive and the genre 'Fiction', the most negative. The Norwegian texts were more negative for 'Personal' compared to the Irish texts ($M = 2.78$ compared to $M = 3.42$, respectively) but more positive for 'Fiction' ($M = 3.00$, compared to $M = -1.72$, respectively).

Welch’s t-test was conducted to test the whether the following null-hypothesis could be rejected: ‘there is no significant difference between the two countries compared’. For the genre ‘Fiction’, a t-statistic of $t(15.22) = 2.93$, $p < .05$ was reported. For the genre ‘Personal’, a t-statistic of $t(13.88) = .32$, $p > 0.05$ was reported. Therefore, the null hypothesis could not be rejected for both instances.

Table 4. *Sentiment means per genre*

Genre	Mean	$M_{Ireland}$		M_{Norway}	
		Mean	SD	Mean	SD
Poem	7.75	7.75	6.15		
Procedure	6.08	6.08	4.27		
Report	5.21	5.21	2.04		
Personal	3.25	3.42	3.60	2.78	5.02
Narrative	2.07	2.07	8.9607		
Fiction	-1.73	-3.36	6.45	3.00	5.61

Is there an association between genre and sentiment for the Irish texts?

An estimated marginal means (EMM) analysis was performed on a linear regression model using sentiment as a dependent variable and genre, country, estimated age and total words as independent variables. Based on the contrasts between genres from the EMM analysis, significant differences in sentiment were found between several genres. These can be found in Table 5. For the Fiction genre, there were significant contrasts between this genre and ‘Narrative’, ‘Poem’ and ‘Procedure’. There is no significant difference in sentiment between Fiction and the ‘Personal’ and ‘Report’ genre. These outcomes show that the sentiments between Fiction and Narrative genres or Fiction and Poem genres do not remarkably differ from each other, however, there are quite some differences in sentiment between Fiction and Narrative, Fiction and Poem and Fiction and Procedure. Similar differences can be found between other genres, as can be seen in Table 5.

Table 5. *Contrasts found in the estimate means analysis. The significant contrasts are highlighted in the first column.*

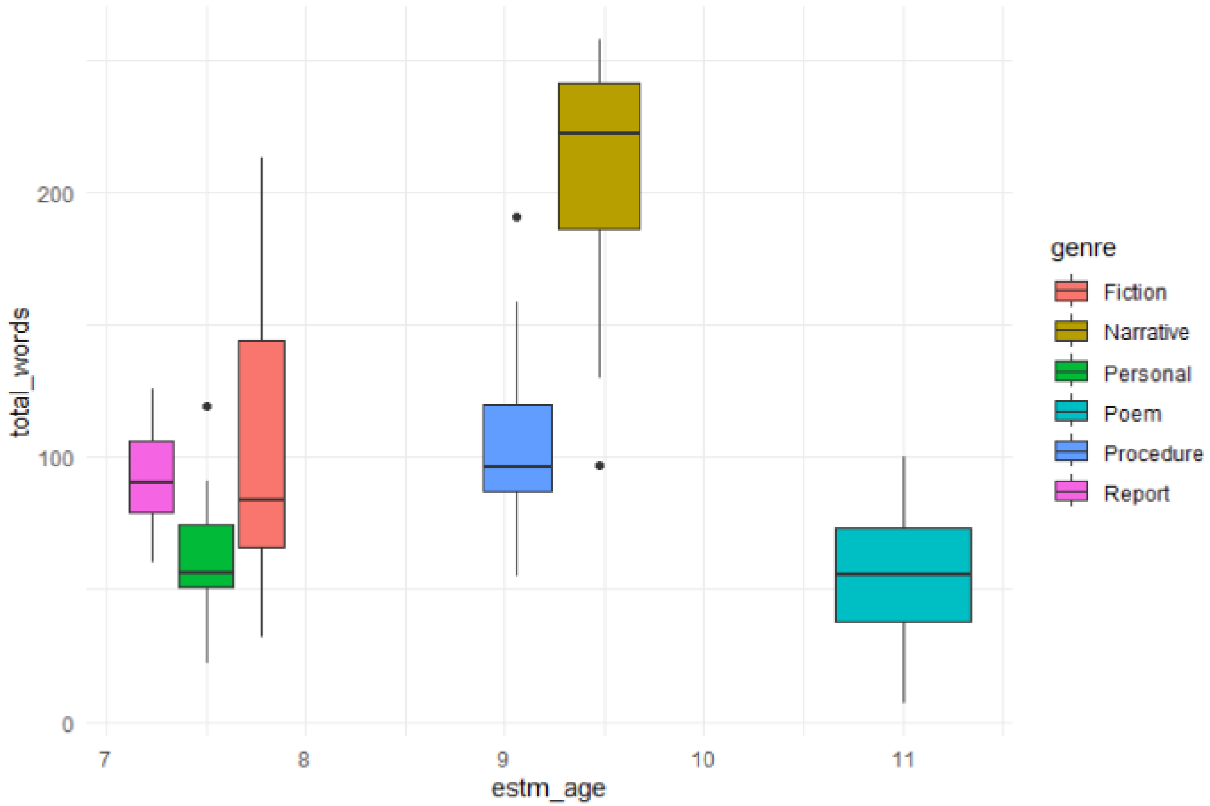
Genre	Estimated means	Standard error	*t*(df)	P-value
Fiction – Narrative	-16.386	5.25	-3.388 (143)	.021*
Fiction – Personal	0.541	2.55	-.045 (143)	1
Fiction – Poem	-38.618	8.39	-4.289 (143)	<.001**
Fiction – Procedure	-18.822	3.61	-4.951 (143)	<.001**
Fiction – Report	-0.825	2.97	-.529 (143)	1
Narrative – Personal	16.927	6.81	2.597 (143)	.214
Narrative – Poem	-22.233	4.43	-4.109 (143)	<.001**
Narrative – Procedure	-2.437	2.91	-.032 (143)	.963
Narrative – Report	15.561	7.18	2.259 (143)	.291
Personal – Poem	-39.159	9.76	-3.679 (143)	<.001**
Personal – Procedure	-19.364	4.91	-3.623 (143)	.026*
Personal – Report	-1.366	2.16	-.675 (143)	.998
Poem – Procedure	19.796	5.32	3.409 (143)	.013*
Poem – Report	37.793	10.24	3.364 (143)	.014*
Procedure – Report	17.997	5.34	3.056 (143)	.045*

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

Is there an association between estimated age and total words per text (corrected for genre) and how does this differ per country?

A linear regression was carried out to explore the effect of age on total words per text. It appears that the occurrence of virtually all genres is contained to a certain age. For example, all 11-year-old students wrote poems, and all 10-year-old students wrote texts within the genre ‘Narrative’. However, some variation can be seen for the other ages. For age 7, the majority wrote a report, but some wrote a ‘personal’ text. For age 8, most students wrote ‘fiction’, but a few chose to write ‘personal’ texts. Lastly, for age 9, most students wrote procedures, but a few wrote ‘narrative’ texts. It becomes clear that the difference between countries accounts for the difference in genre across age. It seems like the texts per genre are divided by different teacher groups. Therefore, no conclusions can be drawn about correlations between age and genre.

Figure 1: *Effect of age on total words, taking genre into account.*



Discussion

The association between genre and total words is best explained when the covariates estimated age and sentiment are considered. It is not surprising that the genre ‘Poem’ has fewer words than the other genres. It is notable that ‘Narrative’ is much longer than the other genres, also compared to ‘Personal’ or ‘Fiction’, even though these seem more similar to each other than fiction and report, for example. Since ‘Narrative’ texts were assumingly written by one division of students, it might be possible that these students had more time to write. Alternatively, it might be possible that other genres require more guidelines with regards to formatting or choosing a writing subject. Therefore, students might spend more time thinking about what they must write, or looking up correct information, for the genre ‘Procedure’, than for the genre ‘Narrative’.

Another finding is that Norwegian texts are significantly shorter than Irish texts. The exact reason for this remains unclear since little is known about the classroom conditions. It might be the case that Norwegian students had less time to write their texts, or that teacher expectations differed between countries. A study by Morin and colleagues (2012) suggests that even style of handwriting can influence writing speed. Another possible explanation for the shorter texts of the Norwegian students can be found in the PISA 2018 (OECD, 2019) results, where Irish students score higher on reading competence than Norwegian students. Since writing is a cognitively effortful task (Peverly, 2006), higher student literacy may indicate higher writing ease, which may lead Irish students to produce more words than Norwegian students in a similar timeframe.

With regards to the differences in sentiments per genre and country, the comparison between Ireland and Norway was obstructed by the limited available data. One finding requires some attention: sentiments seem to be lower on average for the ‘story-like’ genres such as ‘Narrative’, ‘Fiction’ and ‘Personal’. When looking at the Irish dataset, topic seems to be a substantial factor. A considerable part of the Irish students was instructed to write a Halloween story, which affected the type of words used and the average sentiments of some of the genres. This is a likely cause of the large differences that were observed. Additionally, the estimate means analysis described significant differences in sentiments between the genres. It is notable that the genre ‘Poem’ is significantly more positive when the sentiments are compared with the other genres. These findings contrast with research by Hipson (2019), who discovered similar ratios between positive and negative sentiment in poems compared to novels. Possibly, this was again because of the instruction the students received. At first sight, the genres ‘Fiction’, ‘Narrative’ and ‘Personal’ genres seem to overlap a great deal. It is probable that the students’ teachers instructed students to write a certain type of text (e.g., ‘write a fictional story’ or ‘write a report’) and afterwards coded the texts as such. However, it is not exactly clear on the basis on what criteria the different genres were established, and if it is necessary to distinguish between ‘Fiction’, ‘Narrative’ and ‘Personal’, since these genres contain similar elements of storytelling.

When examining the boxplot, it becomes evident that the differences between genres are due to the different divisions of students. For this reason, no correlation between genre and age can be found.

Limitations and recommendations

One limitation of this study relates to the type of lexicon used for sentiment analysis. The AFINN dictionary (Nielsen, 2011) has the advantage of being easy to use when quantitative statistical analyses are performed. However, its library of words is quite small, and it is possible that not all words containing positive or negative affect were identified and scored. Therefore, the measured text sentiments may not be an accurate representation of the actual positivity or negativity of the text because not all the relevant words were taken into account. It would be recommended that future research uses a different dictionary, to gain a different, more elaborate insight into text sentiments. A non-numeric lexicon such as the NRC lexicon (Mohammad & Turney, 2013) might be suitable. This lexicon even offers the opportunity for analysing texts in the original language, which omits the necessity of using a translation service. In this way, possible translation errors can be prevented.

Further limitations of this study are mainly caused by a lack of data and the variations in conditions under which the texts were written. To develop an adaptive recommendation system for reading and writing effectively, it is crucial that student characteristics can be examined in their texts. In future data collection, it would be advisable to examine other factors that may influence writing behaviour, such as gender differences in writing. A study among Estonian students suggest that girls are more interested in reading fiction compared to boys (Uusen & Müürsepp, 2012), so investigating gender differences for writing tasks may be helpful in establishing an adaptive digital environment. Additionally, higher reading competence is associated with motivation for literacy-related tasks (Lepola, 2005). Therefore, it would be suggested to consider measures of reading performance for each student. Since there seem to be developmental trends in poetry writing (Hipson, 2019), with negative affect being higher for older students compared to young students, the current sentiment level may not be representative, since the sample currently entails only 11-year-old students who wrote a poem. It can be expected that sentiment levels would be higher if younger students were instructed to write a poem. As such, it is recommended to gather more poems from all ages to examine sentiment levels and interests which will be relevant for designing an adaptive recommendation system. In the current dataset, the genre is imposed by the teacher through instructing the students. From the teacher's perspective, this is an understandable choice, since primary school students seem to have trouble differentiating between genres (Gillespie et al., 2013). However, this raises a few problems. Because students of the same class have written the same genres, variations in genre are constrained to different student divisions, no conclusions can be drawn about age-related preferences for genres. Allowing students to choose from a few options (e.g., a report, a poem, or a fictional story), so that they can write according to their own preferences, would help tackle this issue. Further, it remains unclear what criteria were used to characterise and distinguish between the genres: What distinguishes a narrative story from a fictional story, for instance, and what are the implications of this distinction? Instead of determining the genre of texts through a set of criteria, using humans to judge the texts, it could be interesting to identify patterns and variations by machine learning algorithms instead. Possibly, this will lead to grouping texts according to different, yet to be defined, 'rules'. It would be interesting to study the application of these categorisations instead of the classic genres.

Conclusion

By examining the texts, a few results become clear about students writing characteristics. Large significant differences between number of words and genre could be found. It is not too surprising that poems contain the fewest words, on average. However, one interesting finding relates to the relation between country and number of words. Norwegian students wrote significantly shorter texts than Irish students. Some significant differences were found for sentiment per genre, and even between countries. Unfortunately, nothing could be said about genre and age. It will benefit the digital adaptive environment if future research will consider more extensive data collection and standardising teacher instruction. This way, more in-depth insight into the student's writing behaviour can be gained.

Some limitations include the lack of data that can be compared to identify main student interests and characteristics. Main recommendations for further research include gathering more data, establishing a methodology for text instruction that can be applied for every classroom and teacher, allowing students to write according to their preferences, and attempting data-driven text categorisation. If these points are considered in future research, individual student differences can be studied more thoroughly.

This thesis aimed to explore differences in writing behaviour of young students by studying variations across countries, sentiments, and genres. Effectively, the gained insights into sentiment, genre, and number of words across texts from Irish and Norwegian students, contribute to the development of an adaptive digital environment by the AILIT project. The engaging properties of this digital environment will be able to promote higher reading competence among primary school students.

References

Bates, C.C., D'Agostino, J.V., Gambrell, L., & Xu, M. (2016). Reading Recovery: Exploring the effects on first graders' reading motivation and achievement. *Journal of Education for Students Placed at Risk, 21*(1), 47–59. https://doi.org/10.1080/10824669.2015.1110027

Binkley, M. et al. (2011), Defining Twenty-First Century Skills, in Assessment and Teaching of 21st Century Skills. Springer Netherlands. http://dx.doi.org/10.1007/978-94-007-2324-5_

Cunningham, A. & Stanovich, K. (1997) Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, Vol. 33*(6), 934-945. https://doi.org/10.1037/0012-1649.33.6.934.

Danubianu, M. (2015). Step by Step Data Pre-processing for Data Mining: A Case Study. International Conference on Information Technologies, Bulgaria.

Gene Diaz. 2016. Collection of Stopwords for multiple languages. Retrieved from: https://github.com/stopwords-iso/stopwords-iso.

Elliott, S. W. (2017). *Computers and the Future of Skill Demand. Educational Research and Innovation Series*. OECD Publishing. http://dx.doi.org/10.1787/9789264284395-en

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59–109. https://doi.org/10.3102/00346543074001059.

Graham, S., Berninger, V., and Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology*. *32,* 516–536. https://doi.org/10.1016/j.cedpsych.2007.01.002

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil,

P. B. Mosenthal, P. D. Pearson,& R. Barr (Eds.), *Handbook of reading research* (3rd ed., pp. 403 – 422). Longman.

Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In *Handbook of research on student engagement* (pp. 601–634). Springer.

Hipson, W. E. (2019). Using sentiment analysis to detect affect in children's and adolescents' poetry. *International Journal of Behavioral Development, 43*(4), 375–382. https://doi.org/10.1177/0165025419830248.

Johnson, D. W., & Johnson, R. T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, *38*, 365-379. https://doi.org/10.3102/0013189X09339057.

Koenka, A. C., & Anderman, E. M. (2019). Personalized feedback as a strategy for improving motivation and performance among middle school students. *Middle School Journal*, *50(*5), 15–22.

Lepola, J., Poskiparta, E., Laakkonen, E., & Niemi, P. (2005). Development of and Relationship Between Phonological and Motivational Processes and Naming Speed in Predicting Word Recognition in Grade 1. *Scientific Studies of Reading*, *9*(4), 367-399. https://doi.org/10.1207/s1532799xssr0904_3.

Lutz, S.L. Guthrie, J.T., & Davis, M.H. (2006). Scaffolding for Engagement in Elementary School Reading Instruction, *The Journal of Educational Research*, *100*(1), 3-20, https://doi.org/10.3200/JOER.100.1.3-20.

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*, 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*, 19–35. https://www.jstor.org/stable/1002159.

Mohammad, S., & Turney, P. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, *29*(3), 436-465.

Morin, M. F., Lavoie, N., & Montésinos-Gelet, I. (2012). The effects of manuscript, cursive or manuscript/cursive styles on writing development in Grade 2. *Language and literacy*, *14*(1), 110-124. https://doi.org/10.20360/G21S3V.

Murphy, R. F. (2019). Artificial intelligence applications to support K–12 teachers and teaching: A review of promising applications, challenges, and risks. RAND Corporation.

Naeghel, J. de, Keer, H. van, & Vanderlinde, R. (2014). Strategies for promoting autonomous reading motivation: A multiple case study research in primary education. *Frontline Learning Research*, *2*(3), 83-102. https://doi.org/10.14786/flr.v2i1.84 .

Naeghel, J. de, Keer, H. van, Vansteenkiste, M., & Rosseel, Y. (2012). The relation between elementary students' recreational and academic reading motivation, reading frequency, engagement, and comprehension: A self-determination theory perspective. *Journal of Educational Psychology*, *104*(4), 1006-1021. https://doi.org/10.1037/a0027800

Nielsen, F. A. (2011). A new ANEW: A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings, 93-98.* https://doi.org/10.48550/arXiv.1103.2903.

Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and research in Education*, *7*(2), 133-144. https://doi.org/10.1177/1477878509104318.

Peverly, S. T. (2006). The importance of handwriting speed in adult writing. *Developmental Neuropsychology*, *29*(1), 197-216. https://doi.org/10.1207/s15326942dn2901_10.

OECD (2019). *PISA 2018 results (Volume I): What students know and can do.* Paris: OECD Publishing.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2(2021), 100020

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 373–404). Academic Press.

Rogiers, A., Keer, H. van, and Merchie, E. (2020). The profile of the skilled reader: an investigation into the role of intrinsic reading enjoyment and student characteristics. *International Journal of Educational Research, 99*, 1–14. https://doi.org/10.1016/j.ijer.2019.101512.

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.

Smedt, F. de, Keer, H. van, & Merchie, E. (2016). Student, teacher, and class-level correlates of Flemish late elementary school children's writing performance. *Reading and Writing*, *29*(4), 833-868. https://doi.org/10.1007/s11145-015-9590-z.

Teng, L.S., Zhang, L.J. (2018). Effects of motivational regulation strategies on writing performance: a mediation model of self-regulated learning of writing in English as a second/foreign language. *Metacognition Learning*, *13*, 213–240 https://doi-org.vu-nl.idm.oclc.org/10.1007/s11409-017-9171-4.

Uusen, A., & Müürsepp, M. (2012). Gender differences in reading habits among boys and

Girls of basic school in Estonia. *Procedia-Social and Behavioral Sciences*, *69*, 1795-1804. https://doi.org/10.1016/j.sbspro.2012.12.129.

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*, *36*(4), 502-513. https://doi.org/10.1111/jcal.12416.

Wigfield, A., Guthrie, J.T., Perencevich, K.C., Taboada, A., Lutz Klauda, S., Mcrae, A., & Barbosa, P. (2008). Role Of Reading Engagement In Mediating Effects Of Reading Comprehension Instruction On Reading Outcomes. Psychology in the Schools, *45*(5), 432-445. https://doi.org/10.1002/pits.20307.