Nordic Results from the PISA Digital Reading Assessment

Tove Stjern Frønes, Eva Kristin Narvhus & Marie Cecilie Aasebo

Abstract

This article studies the Nordic results from the digital reading assessment in PISA 2009. Variations are concealed behind the key figures, and we are both looking for a common Nordic profile and studying how variation is correlated to characteristics of the texts and the tasks through a residual analysis of the Nordic countries’ achievement profile on single item level. We are also undertaking an in-depth analysis of gender differences, in which we investigate tasks with large as well as small gender differences.

Keywords: online reading comprehension, digital literacy, assessment, gender
This article presents an analysis of Nordic results from the PISA digital reading assessment, a computer-based supplementary assessment included in PISA 2009. Among the 19 participating countries, Denmark, Iceland, Norway and Sweden took part in the assessment that investigated the ability of fifteen-year-old students to read on the web. The same students participated in the PISA 2009 investigation of reading skills that was undertaken on paper – and this has given us a rare opportunity to investigate what happens when the students switch media to read web-based texts.

This article has two aims and research questions. In the first part of the article, part A Residual analysis, we will analyze groups of tasks in the digital reading test, to compare whether the countries’ variations are in step or not. If we control for the countries’ performance, which types of tasks and ways of reading can be correlated to the strengths and weaknesses in each country’s group of students? We also attempt to identify a Nordic performance profile, by way of an analysis of all the countries’ performance on individual tasks. The second part of the article, part B Gender differences, describes the Norwegian results and the differences in performance between boys and girls on single item level. In line with most reading tests, the gender differences are in the girls’ favor in most countries. Having examined in detail those textual, task-related or topical issues that characterize tasks with large and small gender differences, we will compare the Norwegian results with those of the other countries.

When compared with other OECD countries, the Nordic countries have an apparently similar performance profile in digital reading in terms of average scores, the magnitude of gender differences and the proportion of students at various proficiency levels (Frønes & Narvhus, 2011; Mejding, 2011; OECD, 2011; Skolverket, 2011). This also applies to reading on paper, if we exclude Finland, which participates only in the paper-based assessment where it outperforms all the other Nordic countries (OECD, 2010; Roe & Vagle, 2010). It is nevertheless interesting to take a closer look at what lies behind the average values, to look for significant effects associated with the students’ national background, gender and performance with regard to various web texts and types of tasks. Several Nordic studies have done so on the basis of PISA and TIMSS data (Kjærnsli & Lie, 2004; Lie & Roe, 2003; Olsen, 2005; Roe & Taube, 2003; Roe, 2012). In their country profile analysis, Lie and Roe found that Denmark, Norway, Sweden and Iceland were clearly clustered in the PISA 2000 reading assessment, while Finland had diverging strengths and weaknesses (Lie & Roe, 2003, p.151). A new analysis of the PISA 2009 data shows, however, that the Nordic countries no longer share this profile, and to the extent that the neighboring countries may share common characteristics, Norway and Sweden make up one cluster and Denmark and Finland another (Roe, 2012). In this analysis, Iceland is more similar to countries such as Greece and Italy.

Online reading comprehension

PISA measures reading comprehension with the aid of tasks accompanying a selection of authentic texts, and is embedded in a framework (OECD, 2007). The framework for reading was expanded and edited prior to PISA 2009, and digital texts and digital reading comprehension were provided with a theoretical foundation, together with the reading of texts on paper. The following definition of reading is currently used, encompassing reading in both media: «Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society» (OECD, 2007, p. 23).

Reading literacy requires a broad set of skills, such as decoding, lexical knowledge, grammar, language and text structure and knowledge about the world. A commonly used definition of reading
stems from the RAND project: “...the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (Snow, 2002, p. 11), signalling a socio-cultural and contextualized view of reading. When understood thus – and also according to PISA’s definition – reading means to realize meaning with the aid of available semiotic resources. The readers construct meaning in interaction with the text, and use a number of strategic, cognitive processes to select, organize, combine and evaluate whatever they are reading. This view of reading forms the basis of this article, along with “new literacies” perspectives on the reading of new forms of texts on new platforms. Within online reading research New literacy-studies is a central and overall perspective for numerous studies on how Internet and other ICT-technology changes the «the nature of literacy» (Coiro et al., 2008; Leu et al., 2007).

Many have pointed out that online reading is different from reading a text on paper. In their review of research literature on online reading, Afflerbach and Cho point out three areas in which the process of reading on a screen is different from reading on paper: a) the process of constructing potential texts to read, b) the necessity of establishing strategies to handle the amount of information and working memory, and c) special requirements for the self-regulatory process of readers (2009, p. 81f). The reading of hypertext presents particular challenges to students who have poorly developed strategies or only very few reading strategies in their toolbox.

According to the “new literacy” school, the investigation of reading comprehension in new media must be undertaken with a new lens, since new skills, strategies and dispositions are required to be literate in a constantly changing technological landscape (Coiro, 2003; Coiro & Dobler, 2007; Leu et al., 2007). Leu et al. point out that whereas many conceptual models have previously focused on processes taking place internal to the reader (cognitive and linguistic processes), the model must be expanded to encompass the purpose that drives online reading, the communicative outcome of the reading process and the continuous change in the skills required to read online as a result of the dynamic nature of the tools used (2007, p. 5).

In analyses such as ours, it is an underlying assumption that the students’ personal literacy and the schools’ practices are to some extent reflected in performance, and may thus help explain the results. By this we mean: The greater the role played by the tools, genres and tasks in the students’ everyday activities, the better, we may imagine that, the students will cope with the same in an assessment. Thus the literacy of the students and the schools’ practices are indirectly brought into the investigation, even though the research design will not capture individual differences. In our analysis, we will emphasize the students’ interaction with texts and tasks, and search for traces of online reading literacy.

Data and methodology

The population for the digital reading assessment was identical to the one for the main PISA assessment in 2009: students born in 1993 who were at school in 2009. In practice, this means tenth-grade students in Norway and Iceland, and ninth-graders in Sweden and Denmark. PISA uses a school-based sample, where participating schools are drawn first, followed by a sampling of 30 students from each school. Of these students, ten were drawn to participate in the digital reading assessment also. As the sample overlapped with the sample for the ordinary paper-based assessment, and, moreover, since there is a common conceptual framework regarding the text features (OECD, 2007), there is rich potential for comparison of students’ abilities in the two assessment modes. The digital reading assessment was delivered by software on a memory stick, and the students’ responses
were stored on the same memory stick. The implementation was identical in all schools: Once the students had been presented with the log-on information, they were given ten minutes to practice using the response platform. They could review the task and response formats and the use of the navigation tools, and they were made aware of how they could monitor their own progress (time spent and number of tasks). The students could then start the assessment module and continue working until they were finished or until the application shut down automatically after 40 minutes.

Residual analysis – part A of our analysis

The first part of the analysis is based on the percentage of correct responses (p-values) to each task. Large-scale international studies such as PISA are based on psychometrical models aiming at deriving measures of some constructs (as reading), and students’ responses to single items are hypothesized to reflect their overall (latent) reading ability. One of the consequences of this perspective is that if countries were ranked according to their percentage correct for any single item (hereafter referred to as p-values), these lists should more or less be similarly sequenced across all items (within measurement errors). This is formally represented as the so-called item-by-country interactions, which is a special case of what is more generally coined as differential item-functioning (DIF). Identifying patterns of DIF is part of the international comparative studies’ procedures to ensure that the items measure the same competencies in different countries (see for instance OECD, 2012; Schulz & Fraillon, 2009 on DIF). After large-scale field trials, items with p-values deviating from the expected value are not selected for inclusion in the final assessment since they are regarded as a threat to valid interpretations of the overall scores across countries.

Still, some item-by-country variation can be observed. In the first part of the analysis such deviations represented as the difference between the expected and the observed percentages, (hereafter referred to as p-value residuals) are used in subsequent analyses. This results in a matrix of p-value residuals across countries and items. From the overall psychometrical perspective, these deviations may be seen as reflections of measurement errors, but another interpretation is that they can be viewed as reflections of relative strengths and weaknesses in countries’ digital reading literacy (about residual analysis, see e.g. Kjærnsli & Lie, 2004; Olsen, 2005; Olsen & Lie, 2011; Zabulionis, 2001).

The matrix of country-by-item residuals is then subjected to a cluster analysis where tasks (and countries) are grouped together into clusters representing the relationship between the countries, presented graphically as a dendrogram (see Figure 1). We make use of an agglomerative hierarchical clustering with between-groups linkage and using the Pearson product-moment correlation coefficient as the measure of distance (Everitt et al., 2001; Olsen, 2005).

The residual analysis allows us to investigate differences between the countries’ means in an effectual way. In the reports from the PISA digital reading assessment 2009 (Frones & Narvhus, 2011; OECD 2011), several analyses of the importance of background variables (e.g. ESCS, school effects, attitudes) are already presented. Such analysis is not our aim in this study, but we keep in mind that one of the important findings is that between-school-difference is twice the size in digital reading as the same students’ results in reading on paper (Frones & Narvhus, 2011, p.112).

In addition to the p-value residuals, information about characteristic features of the tasks is used in the analyses. Tasks are classified according to aspects of both the task and the text: the task’s target group, style, status and type, and the task’s aspect of reading, format and degree of navigation required. The categories are in use in the PISA framework (Frones & Narvhus, 2010; OECD, 2007) and in an...
MA thesis on gender differences in the digital assessment (Aasebo, 2012). We have also investigated each of the categories for gender differences in each country. The following categories have been used in the analysis:

1. **The text’s target group:** The texts describe topics that the students can be expected to be acquainted with to varying degrees, and we distinguish between tasks related to texts that target adolescents and texts that target the general adult public.

2. **Text style:** The text structure of the hypertexts varies from very simple to complex, and has additional variations in terms of vocabulary, layout, text volume and the number of texts in each text unit. A rough distinction can be drawn between texts in formal academic language and those that are written in plain everyday language.

3. **Text status:** We distinguish between authored texts with an identifiable originator and message-based texts from blogs, forums and other social media.

4. **Text type** characterizes the texts’ instrumentalities and rhetorical purpose, and we distinguish between argumentative, descriptive, expository and transactional text types. The text types in the PISA assessments are based on Wehrlich (1976), with the addition of transactional text type (Frones & Narvhus, 2010).

5. **The task’s aspect of reading:** PISA subdivides reading literacy into three aspects (ways of reading), which are also used as reporting scales for results. The aspects are: accessing and retrieving information, integrating and interpreting information, and reflecting and evaluating on the content or the form of a text. Highly sophisticated tasks involving several consecutive steps are categorized as complex.

6. **Task format:** The assessment comprises two types of tasks. Open-constructed response tasks require the students to enter a response, sometimes very short (e.g. a single word or a figure), or a longer response over three to five lines of text. Closed tasks mainly involve multiple-choice responses with four response alternatives, of which the students should tick the correct one.

7. **Degree of required navigation:** Requirements for navigation include the extent to which menus and hyperlinks must be used to solve the task. It varies whether the information requested by the task is prominent and consistent across the web pages, how many navigation tools are used and in terms of the total number of links. In our analysis, tasks are categorized as involving no navigation (0 clicks), some navigation (1–3 clicks) or extensive navigation (>3 clicks).

**In-depth analysis of gender differences – part B of our analysis**

In the second part of this article we study the Norwegian results on single task level and compare them to those of the other Nordic countries with an emphasis on gender differences. In the Nordic countries with the exception of Denmark, the gender differences have traditionally been overwhelmingly in favor of girls when it comes to reading on paper. This pattern is also discernible in the digital reading test. We have selected the twelve tasks with the largest performance gap between the genders, as well as the eleven tasks where the gender differences are smallest (based on Aasebo, 2012). The differences on the tasks with the largest gender differences are also significant. Having looked in some detail at the textual, technical and topical issues that characterize these tasks, we will compare the results with those from the other Nordic countries. For the sake
of clarity, when we refer to gender differences, these are invariably in the girls’ favor if we do not say otherwise.

Analysis part A: A Nordic profile?

Table 1 shows the results for both digital and paper based reading in the Nordic participant countries in PISA 2009 (edited from OECD, 2011). For each country, the average scores from the two tests are provided, with standard errors and standard deviations. A 95% confidence interval of the average values amounts to approximately two standard errors below and above each of the countries’ average scores. The standard deviation is shown in a separate column as a measure of the performance variation. The scale is standardized to enable comparisons across countries and over time, and during the first implementation of PISA in 2000 the OECD average was set to 500 with a standard deviation of 100.

<table>
<thead>
<tr>
<th></th>
<th>PISA test of online reading</th>
<th>PISA test of reading on paper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average score</td>
<td>Standard error</td>
</tr>
<tr>
<td>Denmark</td>
<td>489</td>
<td>2,6</td>
</tr>
<tr>
<td>Iceland</td>
<td>512</td>
<td>1,4</td>
</tr>
<tr>
<td>Norway</td>
<td>500</td>
<td>2,8</td>
</tr>
<tr>
<td>Sweden</td>
<td>510</td>
<td>3,3</td>
</tr>
<tr>
<td>OECD average</td>
<td>499</td>
<td>0,8</td>
</tr>
</tbody>
</table>

A comparison across the two reading assessments shows that Norway performs approximately equally on both assessments, Denmark performs better on paper, whereas Iceland and Sweden perform statistically significantly better on the online reading test. In all countries, the variation is smaller on the online assessment than on the paper-based test, and Norwegian and Danish students have least variation on the online reading assessment of all European countries (OECD, 2011).

With the aid of a cluster analysis of the residual matrix of all the 19 countries’ performance on all tasks in the online reading test, we can estimate the distance between the countries in several ways. To detect patterns of co-variance between countries we use agglomerative hierarchical clustering, a process that analyses between-group linkages in several consecutive steps (Everitt et al., 2001). The distance between the countries is measured by the Pearson product-moment correlation coefficient (r) and a dendrogram showing how and when in the process and at what distance the countries are related to each other. Figure 1 shows the dendrogram of how and at what distance the countries are related to each other in the online reading test. Nodes far to the left end of the figure (small distances) refer to the grouping of strongly positively correlated lower order units, while clustering towards the right end side of the figure (long distances), illustrates the merging of countries/groups that are negatively correlated. The two most similar countries are Australia and New Zealand, followed by Chile and Colombia.
Of all the Nordic countries only Iceland and Norway form an isolated group, while Denmark is included in a weakly linked group along with New Zealand, Iceland, Norway, Japan and Hong Kong. Sweden has a similarly weak kinship with France, Belgium and Poland.

Figure 1. Dendrogram for country clustering in PISA 2009 digital reading

Figure 1 clearly shows that there is no basis for identifying a distinct Nordic profile of relative strengths and weaknesses across the tasks in the PISA online reading test, with the exception of a strong similarity between the performance profiles of Norwegian and Icelandic students. Based on data from the same study, Roe also found that a Nordic profile could no longer be detected with regard to reading on paper, although the relationships between the countries were somewhat different from what they are here (Roe, 2012). It is interesting to note that a change seems to have occurred in a relatively short time where the notion of a Nordic unity is no longer present, but it is not possible from these analyses to identify possible reasons. As referred to above, we believe that the students’ familiarity with digital media and online reading is an important basis for their performance in the assessment, combined with similarities and differences in school practices related to online reading. We will now continue with a more detailed study of the performance level in each Nordic country.

Country profiles

Table 2 shows the p-value residuals for tasks in each of the categories and for each of the Nordic participant countries, as well as values for boys and girls. Positive values imply that the tasks were easier for the students than might have been expected. Corresponding negative values show that the tasks were relatively more difficult for the students. For the residual value to be significantly
above or below what would be expected from the item difficulties and the countries’ average performance, as a rule of thumb, it should exceed 2–3 percentage points in a positive or negative direction. This is not to be considered statistically significant, as these standardized values can best be seen in relation to each other, but as above or below a level of conceptual significance.

Table 2. P-value residuals for average scores in Denmark (DNK), Iceland (ICL), Norway (NOR) and Sweden (SWE), as well as residuals for gender, distributed over each of the seven categories of texts and tasks in the analysis.

<table>
<thead>
<tr>
<th>Text target group</th>
<th>DNK</th>
<th>ICL</th>
<th>NOR</th>
<th>SWE</th>
<th>Denmark</th>
<th>Iceland</th>
<th>Norway</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adolescent</td>
<td>-1.7</td>
<td>2.2</td>
<td>2.3</td>
<td>0.5</td>
<td>-1.2</td>
<td>-2.1</td>
<td>1.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Adult</td>
<td>0.9</td>
<td>-1.3</td>
<td>-1.2</td>
<td>-0.2</td>
<td>0.6</td>
<td>1.1</td>
<td>-0.8</td>
<td>-1.4</td>
</tr>
<tr>
<td>Test style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Everyday</td>
<td>0.1</td>
<td>1.3</td>
<td>1.8</td>
<td>0.8</td>
<td>-0.4</td>
<td>0.6</td>
<td>1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Academic</td>
<td>-0.1</td>
<td>-1.9</td>
<td>-2.9</td>
<td>-1.3</td>
<td>0.6</td>
<td>-0.9</td>
<td>-1.8</td>
<td>-2.4</td>
</tr>
<tr>
<td>Test status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Message-based</td>
<td>1.5</td>
<td>1.7</td>
<td>1.9</td>
<td>1.6</td>
<td>0.9</td>
<td>2.3</td>
<td>1.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Authored</td>
<td>-0.8</td>
<td>-0.9</td>
<td>-1</td>
<td>-0.8</td>
<td>-0.5</td>
<td>-1.1</td>
<td>-1</td>
<td>-0.8</td>
</tr>
<tr>
<td>Argumentative</td>
<td>0.1</td>
<td>1.2</td>
<td>1.7</td>
<td>1.3</td>
<td>-0.2</td>
<td>0.4</td>
<td>1.8</td>
<td>0.5</td>
</tr>
<tr>
<td>Descriptive</td>
<td>-0.4</td>
<td>-1.5</td>
<td>0</td>
<td>0.9</td>
<td>0.1</td>
<td>1.5</td>
<td>-1.4</td>
<td>0</td>
</tr>
<tr>
<td>Expository</td>
<td>-0.9</td>
<td>-1.3</td>
<td>-2.4</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1.9</td>
<td>-3.1</td>
</tr>
<tr>
<td>Transactional</td>
<td>2.5</td>
<td>1.3</td>
<td>3.0</td>
<td>3.1</td>
<td>3.7</td>
<td>4.4</td>
<td>1</td>
<td>3.8</td>
</tr>
<tr>
<td>Aspects of reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access</td>
<td>1.8</td>
<td>-1.1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>2.8</td>
<td>-2.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Interpret</td>
<td>1.8</td>
<td>-1.2</td>
<td>-0.8</td>
<td>1.4</td>
<td>1.9</td>
<td>1.6</td>
<td>0.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>Reflect</td>
<td>-1.5</td>
<td>-1.3</td>
<td>-0.1</td>
<td>-1.7</td>
<td>-4.6</td>
<td>-7</td>
<td>-1.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>Complex</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
<td>0.1</td>
<td>0.6</td>
<td>1</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Task format</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple choice</td>
<td>1.2</td>
<td>0.5</td>
<td>0.5</td>
<td>-0.1</td>
<td>1.2</td>
<td>1.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Open-constructed</td>
<td>-3.7</td>
<td>-0.7</td>
<td>1.6</td>
<td>-1.1</td>
<td>-3.3</td>
<td>-4.2</td>
<td>-0.9</td>
<td>-0.5</td>
</tr>
<tr>
<td>Degree of navigation required</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>-2</td>
<td>0.2</td>
<td>1.5</td>
<td>0.7</td>
<td>-2.2</td>
<td>-1.8</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>Some</td>
<td>0</td>
<td>0.5</td>
<td>0.7</td>
<td>-0.6</td>
<td>-0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Extensive</td>
<td>1.2</td>
<td>-0.5</td>
<td>-1.4</td>
<td>0.1</td>
<td>1.4</td>
<td>-0.8</td>
<td>-0.2</td>
<td>-1.5</td>
</tr>
</tbody>
</table>

As can be seen from the table, the countries show varying degrees of significance and deviation from expectations when we control for performance levels and the degree of complexity of the tasks. In the next section we will highlight interesting findings and draw parallels with previous research, and visualize the results from Table 2 in four bar charts (Figures 2–5).

Text target group

The online reading assessment includes a wide range of texts, which describe topics that can be expected to be known to the students to some degree. In Table 2, we can see that students in Sweden and Denmark perform approximately as expected on texts written for adolescents and texts for adults, whereas Icelandic and Norwegian students perform better than expected on texts for adolescents. If we look at boys and girls separately, we can see that the Icelandic girls and Norwegian boys contribute to this profile in particular. The Danish girls differ from those in the other Nordic countries in performing more poorly on the texts written for adolescents, relatively speaking.

In previous PISA studies we have seen that the Norwegian students perform better on texts written for adolescents than on texts that target an adult audience (Roe & Vagle, 2010). Students in the other Nordic countries also performed better on texts that target adolescents, with the exception of Finnish students, who performed better on texts written for adults.

© Universitetsforlaget, Nordic Journal of Digital Literacy, Vol. 8, 2013, Nr 01-02
Text style

When texts and tasks are categorized according to their style of language, we can see from Figure 2 that students in all the Nordic countries perform approximately as expected on texts written in everyday language, with the exception of Norwegian girls. They perform significantly better on these texts. The Norwegian boys and girls, accompanied by the Swedish girls, perform significantly worse than expected on tasks related to academic texts. This applies mostly to the Norwegian girls, who at the same time deliver relatively the best performance on tasks related to texts written in everyday language, and who also have the widest performance spread of all the Nordic countries, with a difference of more than five percentage points.

![Figure 2. Nordic students’ performance on tasks related to texts written in different styles (academic or everyday language), by gender](image)

In previous PISA studies we have seen that Norwegian girls outperform the boys on transactional texts written in an academic style, and also on tasks that require advanced language and textual skills. Seen as a group, Norwegian girls possess a wider set of written-language skills than the boys (Roe & Vagle, 2010, p. 91). The online reading assessment shows us another picture. Even though the Norwegian boys perform relatively less well on tasks related to academic texts than those written in an everyday language, the Norwegian girls here stand out more distinctively.

Text status

The texts included in the online reading assessment are also subdivided into two main types: authored texts and message-based texts. Authored texts are static and have an identifiable origin, for example an official website, whereas the message-based ones are dynamic and open to influence by the reader, such as those found in blogs, forums and other social media. Texts categorized as message-based will to some extent overlap with those that use everyday language and that are written for adolescents. This observation and the fact that message-based websites account for a significant proportion of young people’s use of the Internet, and that of girls in particular, give rise to expectations of especially strong performance levels (Egeberg et al., 2012; Frønes & Narvhus, 2011;...
Medietilsynet, 2012). From PISA survey data we know that girls are the most frequent participants in social media such as web communities and chat rooms (Frones & Narvhus, 2011, p. 91).

Despite these expectations, the Nordic students do not perform significantly better on message-based than on authored texts, with the exception of Danish and Norwegian girls. They perform better than expected on message-based texts, see Table 2, and this might be explained both by the access to equipment and the great amount of time spent in online forums by these students (Frones & Narvhus, 2011, p. 97).

**Text type**

Text type characterizes the text’s instrumentalities and rhetorical intention, and the online reading assessment distinguishes between expository, argumentative, descriptive and transactional text types. Expository texts are characterized by having a heavy information content describing complicated issues, and the students must locate very detailed information that can be hard to find and may be accompanied by many other competing information parts. From previous paper-based tests, we know that this presents challenges to the students (Roe, 2010; Roe & Vagle, 2012). An example from the assessment could be the text about a disease, which comprises a web page from a medical centre, a page from an online medical encyclopedia and a website for a medicinal drug. Figure 3 shows that Norwegian and Swedish students perform more poorly than expected on tasks associated with expository texts. These results are consistent with results from reading on paper in PISA 2009 (Roe & Vagle, 2010).

Both Danish and Swedish students have tasks related to transactional texts as their forte. These comprise, for example, e-mails, discussion forums and applications for ordering products online.

![Figure 3. Nordic students' performance on tasks related to texts of various types: transactional, expository, descriptive and argumentative texts](image-url)
When comparing expository and transactional text types we can see a distinctive profile, and tasks related to transactional texts reveal interesting gender differences in the countries concerned.

Figure 4 shows that the Swedish and Danish girls perform significantly better than expected on transactional texts. Of the boys, only the Swedish perform better than expected on this type of text.

A different profile appears when we look at the expository texts: Girls in Norway and Sweden perform more poorly than expected, and the Norwegian girls deviate most from their expected level. The boys in the Nordic countries perform approximately as expected on this type of text.

The tasks’ aspect of reading

In PISA, reading literacy is subdivided into three aspects, which are also used as reporting scales for results: *accessing and retrieving information, integrating and interpreting information,* and *reflecting and evaluating on the content or the form of a text.* Distinguishing between these three can be difficult, and there is to some extent a hierarchical relationship between them.³
Figure 5. Nordic students’ performance on tasks related to the aspects of reading of accessing information, integrating and interpreting information, and reflecting and evaluating on the content or the form of texts, as well as complex tasks.

Figure 5 shows how Nordic students perform better or worse than expected on tasks related to each aspect. The Danish students stand out in performing far more poorly than expected on the tasks that require reflection, and the girls poorest of all (see Table 2). Swedish girls also score relatively poorly on tasks that involve reflection.

PISA 2000 and PISA 2009, both of which focused on reading proficiency, fail to provide an unambiguous picture of the Nordic countries’ aspect profile (Roe & Vagle, 2010, p. 71). In 2000, Norwegian and Danish students performed equally well on average in each of the three aspects, while in 2009, students in both countries performed relatively better on tasks requiring them to find and retrieve information than on tasks involving interpretation or reflection. We cannot yet conclude whether this is part of a trend, and we look forward to new analyses resulting from the next round of the survey to focus on reading proficiency, PISA 2018.

Task format

The online reading assessment includes two task formats: open-constructed response item and multiple-choice. PISA 2009, like previous PISA studies, has shown that the gender differences in the girls’ favor are largest with regard to tasks requiring the students to formulate an answer in the form of an explanation or justification, but are somewhat smaller when the students need only to provide a short response or tick a box (Roe & Vagle, 2010). Analyses of the PISA 2009 reading assessment on paper showed that Norwegian and Swedish students had no particular strengths or weaknesses associated with the task format (Roe & Vagle 2010). Icelandic students score relatively better on tasks requiring short answers, while Danish students are relatively weaker when it comes to responses requiring longer formulations with explanations or justifications.

In Table 2, we can see that Danish students stand out in their responses to the two task formats in the online reading test; both girls and boys perform significantly less well than expected on the
open-constructed tasks. Students from the other Nordic countries perform approximately as expected in both task formats. Norwegian girls, however, score better than expected in the open format, exactly as in previous studies.

Tasks with and without navigation

As mentioned above, the tasks are categorized into three types: those that require no navigation whatsoever (0 clicks), those that require some navigation (1–3 clicks) and those that require extensive navigation (>3 clicks). It appears as though the requirements for navigation have little effect on performance levels. Danish boys perform more poorly than expected on tasks with no need for navigation, while Norwegian girls perform better, see Table 2. In light of the boys’ previously dominant role in the use of digital media, we had expected to find larger gender differences, but this picture is constantly changing. One factor that obviously has a bearing on this outcome is the fact that the reading assessment does not include tasks that require advanced digital skills, since it is designed as a traditional reading assessment subject to the reading framework.

Analysis part B: Gender differences

In all the PISA countries, the gender differences are in the girls’ favor when it comes to reading on paper. However, this imbalance was largely expected to be redressed when the assessment was implemented digitally, because the boys have traditionally been dominant in the use of digital media. The results show, however, that on average the girls continue to perform better than the boys in all countries, even though the distance between the genders is smaller than in the paper-based assessment (Frønes & Narvhus, 2011, p. 32f). Seen from a Nordic perspective, the gender differences in the Nordic countries are larger than average for the OECD, with the exception of Denmark, where the gender difference amounts to only 6% in the girls’ favor.

Results from Denmark show that as many as 19 tasks (of the total of 29 tasks) produce striking small gender differences. Sweden has 14 tasks with remarkably small gender differences and Norway has six. Norway and Iceland have the largest number of tasks with significant gender differences. In the Nordic countries, the gender differences tend to be small for tasks related to texts written in academic language – both girls and boys have trouble with these. Moreover, and with the exception of the Danish results, it appears as though girls perform better when the tasks relate to social media and communication genres. We have therefore chosen to split the analysis into two parts: We will look at, on the one hand, tasks associated with texts written for adults in an academic style, and, on the other, tasks related to texts drawn from social media and other communication genres.

We will first look at the twelve tasks that produced significant gender differences in Norway on single task level, three of which were associated with academic-language texts and nine with everyday-language texts. We will also see how the other Nordic countries performed on these twelve tasks.

Tasks with texts written in academic language

Only three tasks associated with academic-language texts produced significant gender differences. One of the tasks had an open-response format, and the students needed to use information from the online texts to give reasons for their answers. Another task was related to a text in which continuous as well as non-continuous text parts were described in a diagram. The final task required
the students to navigate extensively to find the correct answer, and we may assume that this type of assessment also measures the students’ concentration and perseverance.

Results from the other Nordic countries show that the task involving interpretation of the bar graph produces a parallel gender profile. Sweden and Denmark have large gender differences for the navigation task, but diverge from the Norwegian results in their smaller gender differences on the open-response task. Iceland has a gender profile, which is fairly similar to Norway’s for all tasks.

Tasks with texts written in everyday language

Of all the tasks that produced significant gender differences, nine were related to texts written in everyday language. All these texts were drawn from the online genres social media or web communication. The latter includes high-frequency digital genres such as blogs, online forums, online learning forums and e-mail. For example, the students would read a blog entry and identify information explicitly referred to in the blog’s text. The task for the online forum involved reading vacancy announcements and identifying which of four jobs would be best suited for a student. The task for the online learning forum involved identifying plain information. The e-mail tasks included reading a long and complicated exchange to agree on a shared activity and order a product. All these tasks are characterized by being associated with texts containing a large amount of information, and they would often require the students to combine information from several websites to provide a correct answer. Four of the tasks required an open-format response. Moreover, four of them involved a complex aspect of reading requiring the students to perform numerous labor-intensive operations.

Among the Nordic countries, Danish students stand out in terms of their striking small gender differences related to six of these nine tasks. The only task for which the gender differences are pronounced in Denmark is one involving e-mails. Sweden has striking small gender differences on the task related to the online forum and one of the e-mail tasks. The Icelandic results are near-identical to the Norwegian ones.

In the next part of the analysis we will take a look at eleven tasks that produced striking small gender differences in Norway, six of which involved texts written in academic language and five in everyday language. We will also see how the other Nordic countries performed on these eleven tasks.

Tasks with academic-language texts

Six tasks related to texts written in an academic style produced striking small gender differences. For most of these tasks, the Norwegian performance levels are below or approximately equal to the OECD average. The tasks included reading complicated factual texts about a medical topic, as well as websites of organizations with descriptions of complicated topics.

All the Nordic countries returned small gender differences from several of these tasks. Iceland stands out in terms of larger gender differences than the others for two of the tasks. Denmark has large gender differences for one of the tasks.

Tasks with everyday-language texts

Of the eleven tasks that returned very small gender differences, five were related to texts written in everyday language. All the tasks were associated with texts belonging to the genres of social media.
and web communication. Three of the tasks have certain features that are especially interesting. The topics addressed in them can be assumed to kindle the boys’ interest in particular. One deals with crime, and another with a desire for a future job as a web designer. Moreover, the former of these produced outcomes in the boys’ favor in both Norway and Sweden, even though the gender difference is not significant. The tasks included the type “complex multiple-choice”, which tests combinatorial skills. This was the only task of this type, and it appears to have caused problems for boys and girls alike. Another task tests the students’ skills in critical assessment of key factors of web use. In this task the students cannot read the answer directly from the text, but need to use their prior knowledge of necessary precautions when using the web in order to respond.

Like Norway, the other Nordic participant countries return small gender differences from the tasks dealing with the precautions that are necessary when using the web. In addition, they have small gender differences on the task dealing with crime. This applies also to the other Nordic countries with the exception of Iceland. Iceland stands out in terms of significant gender differences in the girls’ favor for three of the tasks, whereas the other Nordic countries have small gender differences. This applies to the task dealing with crime, the complex multiple-choice task and the one dealing with the desire for a future job as a web designer. The last task also returns significant results in the Swedish girls’ favor.

Discussion

In the introduction to this article we pointed out two assumptions that are inherent in analyses such as these: We expect that the personal literacy of the students as well as school practices will be reflected to some extent in the performance levels, and thus contribute to explaining the outcomes. Other systematic reasons for differences like curriculum and culturally grounded literacy practice differences might also be considered. However, the research design of large-scale studies such as PISA does not enable us to draw conclusions about these matters, but we may indicate some possible correlations.

In our investigation of the residual matrix for all countries, we failed to detect an isolated Nordic profile, but we did find similarities between the performance levels of Norwegian and Icelandic students. Studies of the students’ access to and use of computers at home and at school may corroborate this finding, since Norwegian and Icelandic students are reported to have rather similar habits of computer use (Frones & Narvhus, 2011, p. 86ff). Danish students also report extensive access to and frequent use of computers, including at school. In the light of this, it is unexpected that Denmark should not be part of a Nordic unit, but this is most likely due to the complexity of the causal factors. On the whole, Swedish students report having less access to computers at home and at school – and using computers less frequently for school work – and the distance between Sweden and the other Nordic countries revealed by the cluster analysis is more in step with expectations.

Throughout the analysis of the categories of texts and tasks we have compared the findings to previous PISA studies of reading on paper. This is interesting for three reasons. First, the sample used for the online reading assessment is representative of the students who participated in the paper-based assessment in PISA 2009. Second, we observe a growing intra-Nordic distance and increasing diversity from PISA 2000 to PISA 2009, and it is interesting to see whether the online assessment points in the same direction. Third, it is interesting to compare reading performance in the two media to understand how these two types of reading differ from each other (cf. Leu et al.,
We can claim to have made two important findings: Our analysis of the online reading assessment confirms previous findings that Nordic students perform better than expected when they are given tasks related to texts that have adolescents as their target group, and where the topic and the text appeal to them. This is confirmed by the fact that the students perform equally well when facing a transactional text type. We find no corresponding results for tasks associated with message-based texts, or for texts written in everyday language, which is surprising, since these texts have a number of traits in common.

Returning to the issue of the students’ personal literacy as an expository factor behind the findings, we would like to emphasize in particular the way in which students in nearly all the Nordic countries perform better than expected on tasks associated with transactional text types, texts that are written specifically for adolescents and message-based texts. The Danish students’ mastery of tasks related to adolescent texts is an exception. The in-depth analysis of tasks that return small versus large gender differences confirms this finding. On the whole, the results show that the small gender differences stem from the girls having performed as poorly as the boys, and not from the boys having risen to the girls’ level.

The Nordic girls cope with genres in social media and web communication better than the boys, which concurs with research showing that girls in particular spend a lot of time on social media (Enjolras & Segaard, 2011; Tømte, 2011; Taylor & Keeter, 2010). The gender differences are smaller, however, when the texts are authored and static, for example with enterprises as originators. The topics dealt with in the various texts also appear to have an impact on the gender differences, depending on whether they appeal to boys or girls in particular.

The gender differences also appear to diminish when the information that the students must identify is easily accessible in the text. However, when the students are required to combine information from a website or retrieve complex information from various websites, the boys perform more poorly than the girls. The results indicate that the boys know where to go in the hypertext, but will not invariably know what to retrieve, interpret or assess in order to be able to respond to the task. On the other hand, both genders perform less well on tasks for which prior knowledge is an advantage in order to navigate effectively.

Conclusion

Access to computers and the Internet in schools has for many years been a key concern for the ICT programs in Nordic schools. PISA 2009 shows that the access to PCs and the Internet is currently at a high level – without this necessarily having a decisive impact on the students’ performance (Frønes & Narvhus, 2012; Mejding, 2011; Skolverket, 2011). The purposes for which the computers are used in a private context (leisure as well as school-related activities) are of major importance. Frequent use of the web for “serious” activities, such as searching for information, searching for practical advice and using reference material, also proves to have an impact on the students’ performance in the assessment (Frønes & Narvhus, 2012, p. 71). There are also reasons to be concerned by the huge gender gap that we found in the PISA 2009 assessment, similar to the gap we have known for a decade in traditional reading assessments. If this trend is not diminished when coming of age, a generation of boys might have problems with solving complex tasks using technology and also participating in society as peers.
Seen in this light, our findings are not surprising – the students are relatively best at the things they engage in most frequently, such as communicative texts written in plain everyday language and used in social media. It is though problematic that we see few traces of frequent, systematic school activity and training in the material, as we should expect both from the connectivity rate in Nordic schools and the national curricula (e.g. K06 in Norway (Utdanningsdirektoratet, 2013); Erstad, 2010), and this might be the reason for the earlier mentioned finding of huge between-school-differences. The findings nevertheless send a warning that the school system faces challenges related to provision of relevant training in a number of other types of texts and tasks that are even more essential for social participation, such as expository texts written in a less accessible style, describing adult and “boring” topics. There also seems to be a need for systemic approaches in both teacher and school leadership educations to meet the challenges in school of digital technology in rapidly changing and transforming the literacy practices needed for modern citizenship.

Acknowledgments

The PISA project in Norway is funded by the Norwegian Ministry of Education and Research and the Norwegian Directorate for Education and Training. The authors would like to thank the editors of Nordic Journal of Digital Literacy, the two anonymous reviewers, PISA National Project Manager Marit Kjærnsli and also colleagues Rolf V. Olsen and Astrid Roe for their comments on an earlier draft of this article. We would also like to thank the Norwegian Centre for ICT in Education and our research group Measurement and Evaluation of Student Achievement (MESA) at UiO for funding parts of the research project.

References


Coiro, J., & Dobler, B. (2007). Exploring the online comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. Reading Research Quarterly, 42(2), 214–257.


Olsen, R.V. (2005). *Achievement tests from an item perspective. An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students’ knowledge and thinking in science* (Doktoravhandling). Universitetet i Oslo: Unipub.


---

1 The Programme for International Student Assessment (PISA) is an international study of the educational systems in various countries which is repeated every three years. The study measures the skills of fifteen-year-olds in the disciplines of reading, science and mathematics. The study is undertaken under the auspices of the Organisation for Economic Co-operation and Development (OECD), and is implemented in nearly 70 countries.


3 See PISA’s framework for an account of the relationship between the aspects (OECD, 2007, p. 34ff).