CHAPTER FOURTEEN

MODE EFFECT IN LARGE-SCALE ASSESSMENT

Juhani Rautopuro and Elina Harjunen

Aims

In Finland, there is an objective to digitize traditional pen-and-paper based large-scale assessments of learning outcomes. For example, the matriculation exam (approximately 35 000 participants per year) should be completely digitized by 2019. Likewise, the Finnish Education Evaluation Centre (FINEEC) has started digitizing sample-based (approximately 10 % of the age cohort) evaluations of learning outcomes in basic education. The purpose of the assessments is to produce reliable information on how well the objectives of the national core curriculum for basic education are met, and on success in promoting educational equality. Inspired by this trend, the aim of this chapter is to concentrate on educational equality and justice. The data of this study is based on an external, national assessment of learning outcomes in mother tongue (Finnish) and literature at the end of basic education, from April 2014. Two different assessment modes were applied: a traditional pen-and-paper version and a digital environment. The assessment consisted of assignments in linguistic knowledge and writing competence. The a priori assumption was that assignments of this kind could lend themselves well to the digital environment. On the basis of the results, we discuss the similarities and differences between a printed version and a digital version of the assessment. Factors associated with the results are also discussed.

Introduction

Over the last decades, there has been an increasing demand to reform the methods of education and assessment (e.g. European Commission 2009). Towards this end, numerous computer-based environments and platforms have been constructed to develop teaching, learning, and assessment. However, large-scale, coherent, and evidence-based research on the impact of these environments is quite exiguous. Furthermore, the introduction of new technology and environments has not very often been conducted on the terms of pedagogy (Lehtinen 2006; Tossavainen 2013).

In April 2014, the Finnish Education Evaluation Centre (FINEEC) assessed learning outcomes in mother tongue and literature at the end of basic education. Together with a pen-and-paper test, a digital environment was field-tested in

an assessment of learning outcomes at the end of comprehensive school for the first time (in Finland). The experiment was quite encouraging but also a little confusing.

Transforming student assessment from a pen-and-paper to digital is a challenging task. The benefits of digital assessment are widely accepted; Digital environments provide the possibility for rapid data collection and analysis, allow assessments to take place anywhere, facilitate the use of enhanced and multiple assessment tasks, promise updated feedback, and offer various possibilities for the research of students' thinking and their misconceptions (e.g. Masters 2013, 24–27; Williams 2012). Moreover, digital environments enable an effective use of adaptive testing (e.g. Lahti et al. 2013).

The flip side of the coin is that in order to ensure the validity of the assessment, equal inward and outward circumstances should be guaranteed. From the point of view of the students, the cognitive load of the assessment environment must be considered carefully. Research has shown, for example, that students familiar with the assessment environment perform better compared to students using the environment for the first time. Therefore, teachers should introduce the tool properly, so that students can focus on the issues at hand, not on the tool. In addition, an unfamiliar environment can bring on an insurmountable cognitive load, especially for students with learning difficulties (e.g. Clariana and Wallace 2002; Ford Lawton 2014; Laakso et al. 2008; Noyes et al. 2004).

The assignments in 2014 were basically the same in both versions of the assessment. They were primarily planned for the paper (main) version of the assessment. Therefore, certain assignment types did not lend themselves as well to the digital environment, or their usability differed from that of the printed one. Similar findings have been detected in previous research (Chua and Zuraidah 2013; Kim and Huyhn 2007). For these reasons, the comparison between digital and paper-based assessment remains indicative, since even though the assignments shared the same principles, they differed in their method of execution.

The teachers, researchers, and evaluators of learning outcomes should be people with the skills of "digital assessment literacy". In practice this means that, at the most basic level, a person understands the use of digital tools in all phases of an assessment, at an intermediate level, a person has a holistic view of using alternative digital methods integrated with assessment, and, at an advanced level, a person is capable of sharing the methods for assessment and collaborate with other partners of the assessment in various ways (Eyal 2012).

Data and methods

The data used in this research came from 3 345 pupils (99 schools) who participated in the paper version of the assessment and 1 799 pupils (50 schools) who participated in the digital version. The assessment took place at the end of basic education (9th grade), when most of the pupils were 15 to 16 years old. In the pen-and-paper version, 52 % of the participants were girls, in the digital version 49 %. The data was collected by using two-stage stratified random sampling. The data is a representative sample taking into account different provinces, different municipality types, and schools of different kind and size in Finland (Harjunen and Rautopuro 2015a).

The data was analysed by using various statistical methods. Basic results have been presented by using descriptive measures (e.g. percentage distributions and measures of central tendency and variation). The associations between categorical variables were examined using the traditional chi-square test. Group differences were tested using either independent samples t-test. In addition to statistically significant differences (p-values), effect size measures (Cohen's d) are also reported. The interpretations of Cohen's d-value are quite relative. In this chapter, effect size approximately 0,2 means "small effect", effect size around 0,5 "medium effect" and effect size around 0,8 or more refers to a "large effect" (Cohen 1988, 10).

Results

The assessment concentrated on two areas of mother tongue and literature: on linguistic knowledge and writing competence. All in all, the results were fairly evenly distributed across Finland. Differences between schools were also reasonably small. In linguistic competence, only about 5 % of the total variation was explained by school differences. The corresponding shares in writing were 9 % in printed assessment and 10 % in the digital assessment. Even though the differences in writing are not alarming, they are somewhat larger than those observed in PISA assessments.

Linguistic knowledge

The linguistic knowledge assignments measured how well pupils:

1. Understand that situation and purpose (context) influence the choice of linguistic expression.

2. Recognize different expressions and are able to interpret their meanings within their contexts.

3. Have a good command of standard language norms.

In the digital assignments, the average percentage of correctly completed questions was 51% of the maximum linguistic knowledge score. The score was about 5% lower than that of the printed assessment ($p < 0{,}001$; $d = 0{,}27$). Girls' average score was 56% and boys' 45%. Girls' result was 6 and boys' 5 percentage points lower than the corresponding figures in the printed assessment. The gap between girls and boys was especially notable in the digital version (11 percentage points), and boys' competence levels varied slightly more than those of girls. As in the printed version, the greatest differences between genders were observed in the assignments measuring standard language norms, where girls scored 8 and boys 6 percentage points lower than in the printed assessment (Harjunen and Rautopuro 2015a, 15, 123–125; 2015b).

When examining individual items, the difference between the results of the printed and digital version was the largest ($d = 0{,}52$) in an assignment where the pupils had to click a linguistic form within the text to underline words (a verb form in the perfect tense). Those who failed had either clicked only one of the words of the tense, or else several wrong words. In the first case, the reason for the wrong answer was likely due not only to their linguistic competence but the competence to use a keyboard, a mouse, and a precision different from the pen-and-paper method. If the student was not exact and did not check their answers, it would have been easy to click only once and think that both words were underlined. The second largest difference ($d = 0{,}38$) was in an assignment where pupils had to click a verb, which was in the passive voice in another text.

In contrast, the scores of some assignments were higher in the digital than those of the printed assessment. However, the effect size measure showed the mode effect was not very strong. The result is not a surprise. The tasks were largely the same in both versions but they were primarily planned for the paper (main) version of the assessment. Therefore, the usability of certain assignment types differed from that of the printed one. For example, pupils were not able to see the longer text on a single screen, which may have influenced their overall view of the text. However, while some sections of the linguistic knowledge assignments formed a concrete continuum, and nearly all sections were on a continuum within their context (e.g. an application for a summer job), pupils may have found it more difficult to perceive these continuums in the digital assessment, even though they were reminded of them in each assignment section

(Harjunen and Rautopuro 2015a, 121–144; 163–164). Though, results show that the single tasks worked almost equally well in both versions.

As seen in the table below (14–1), it is easy to see that the results were generally somewhat better in the paper version of the assessment compared with the digital version. The percentage of pupils achieving upper scores was higher in the paper version. This trend seems to be quite similar for both boys and girls. Due to the large sample size, all differences are statistically significant. However, the effect size measure (Cohen's d) shows the mode effect is not very strong. Nevertheless, the mode effect was slightly more pronounced among girls than among boys.

When seperately analyzing pupils, who were aiming at upper secondary education and vocational education, the differences (effect sizes, as well) in learning outcomes in the paper version and the digital version were about the same as shown in the Table 14-1.

Previous research in Finland has shown that the educational background of pupils' parents is strongly connected to pupils' achievements in the assessments of learning outcomes (Harjunen and Rautopuro 2015a, 105–106, 142; Hildén and Rautopuro 2014, 78–81; Rautopuro 2013, 7, 115). Our results show that the higher the level of parents' education, the better the students' results. However, parents' education has no effect on differences in achievements between the paper version and the digital version of the assessment.

Even if there seems to be a seemingly significant difference between the achievements in paper and digital version of the assessments, the phenomenon is not that straightforward: the effect size measure (Cohen's d) shows that the mode effect is generally not very strong when analyzing single tasks.

Table 14-1. Results in linguistic knowledge

	Boys (%)		Girls (%)		Total (%)	
	Paper	Digital	Paper	Digital	Paper	Digital
Under 10%	0,4	1,3			0,2	0,6
≥ 10% to 20%	3,5	6,6	0,7	1,7	2,2	4,2
≥ 20% to 30%	10,6	12,8	2,4	4,8	6,6	8,6
≥ 30% to 40%	19,3	22,0	8,1	11,7	13,9	17,0
≥ 40% to 50%	17,6	16,9	11,3	16,8	14,5	16,8
≥ 50% to 60%	16,8	17,1	16,3	16,6	16,5	16,8
≥ 60% to70%	18,7	14,2	26,3	26,4	22,4	20,3
≥ 70% to 80%	9,1	7,1	20,5	14,6	14,6	11,0
≥ 80% to 90%	3,3	1,5	12,2	6,2	7,6	3,9
≥ 90%	0,8	0,6	2,2	1,1	1,5	0,8
Mean	49,6	45,2	62,6	56,6	55,9	51,0
Standard deviation	17,9	18,1	16,5	16,8	18,4	18,3
p-value	p < 0,001		p < 0,001		p < 0,001	
Cohen's d	d = 0,24		d = 0,36		d = 0,27	

Writing competence

When it came to writing competence, the pupils achieved an average of 57% of the maximum writing score in the digital assignment. Girls' average score was 65% and boys 48%. The results were some 3 percentage points lower than the corresponding figures in the printed assessment. In both versions, the competence gap between girls and boys was large, and boys' competence levels varied slightly more.

Writing skills were assessed on the basis of three assignments: writing an application for a summer job, a rejoinder to a column, and a news item or a description based on a photograph. The two first assignments were designed for this particular evaluation of learning outcomes. In these two assignments, there were no statistically significant differences between the paper and digital version. There were no differences between test modes even when the results were examined by gender and parents' educational background. The assessment criteria for the application and the rejoinder covered the typical features of these genres. Language was also assessed, with two language criteria applied to the rejoinder.

The situation was remarkably different when an old assignment (linking item) was used—The third assignment, writing a description or news item on the basis of a photograph, was included in the 2001 assessment of learning outcomes. This enabled comparisons with earlier results. This particular assignment was originally designed for a pen-and-paper assessment and has now been transformed to a digital platform. In linking items, the assignment and the criterion were to be exactly the same as in the previous assessment. In the case of boys, the 2014 assessment yielded similar results in the printed version and lower results in the digital version to the 2001 assessment. For the assignment, the pupils could make a selection between two alternatives: description and news. The results for this item are presented in Tables 14–2 and 14–3.

Table 14-2. Results in writing "description"

	Boys (%)		Girls (%)		Total (%)	
	Paper	Digital	Paper	Digital	Paper	Digital
Under 10%	3,7	23,9	0,9	7,3	4,6	14,1
≥ 10% to 20%	7,5	14,8	1,2	2,9	4,7	7,9
≥ 20% to 30%	15,7	11,7	3,3	7,6	8,9	9,3
≥ 30% to 40%	8,8	7,4	3,4	6,4	5,8	6,7
≥ 40% to 50%	17,8	14,3	11,0	11,4	13,1	12,6
≥ 50% to 60%	16,9	10,9	14,7	14,3	16,2	12,8
≥ 60% to70%	7,5	5,7	8,1	7,9	8,4	7,1
≥ 70% to 80%	11,2	6,1	20,1	15,7	16,4	11,7
≥ 80% to 90%	8,8	4,8	23,9	16,9	14,6	11,9
≥ 90%	2,2	0,4	13,5	9,6	7,4	5,9
Mean	51,8	33,5	61,4	58,2	57,2	48,0
Standard deviation	24,5	26,0	24,8	27,2	25,2	29,4
p-value	p < 0,001		N.S		p < 0,001	
Cohen's d	d = 0,73		N.S		d = 0.34	

Table 14-3. Results in writing "news"

	Boys (%)		Girls (%)		Total (%)	
	Paper	Digital	Paper	Digital	Paper	Digital
Under 10%	6,3	8,4	4,9	2,7	5,7	5,8
≥ 10% to 20%	5,1	7,4	3,9	3,0	4,6	5,2
≥ 20% to 30%	10,3	14,1	5,4	4,8	8,1	9,7
≥ 30% to 40%	6,7	6,9	5,0	5,8	6,0	6,4
≥ 40% to 50%	17,2	17,2	13,9	15,2	15,7	16,2
≥ 50% to 60%	17,5	15,7	15,8	16,1	16,8	15,9
≥ 60% to70%	9,5	9,0	7,5	9,7	8,6	9,3
≥ 70% to 80%	13,5	12,6	18,1	17,5	15,5	14,9
≥ 80% to 90%	10,7	7,5	18,4	17,7	14,1	12,4
≥ 90%	3,3	1,1	7,1	7,6	5,0	4,2
Mean	51,7	46,4	59,5	61,1	55,2	53,3
Standard deviation	24,5	26,0	24,8	27,2	24,8	24,7
p-value	p < 0,001		N.S		p < 0,05	
Cohen's d	d = 0,22		N.S		d = 0,08	

It is easy to see from Tables 14–2 and 14–3 that the results from the third writing assignment are different than those from linguistic competence (Table 14-1). First, a statistically significant difference was not found among girls between the paper and digital versions of the assessment. Second, in the task of writing a description, the mode effect (d = 0,73 vs. 0,24) for boys is remarkably stronger in comparison with the linguistic competence. The difference was quite strong (d = 0,73) only among boys in the task of writing a description, but small (d = 0,22) among boys in the task of writing a news item. In the description task, the mean result was nearly 20 percentage units lower in the digital version compared to the paper version. Almost one fourth of the boys fell under 10 % achievement in the digital version, while in the paper version less than 4 % of the boys performed so poorly. Also, the percentage of good achievements was remarkably lower in the digital version among boys.

Based on the data of this assessment it is difficult to explain the differences between boys and girls in the digital environment. However, research has shown (e.g. Kaarakainen et al. 2013) that, while the activity to use ICT in Finland is quite the same between boys and girls, there are differences in how they use ICT. Girls use significantly more social media and blog posting than boys. Boys, on the other hand, concentrate more on internet games and programming.

The assessment criteria for the linking item were intended for the pen-and-paper method, for that reason all criteria did not fully suit the writing assessment within the digital environment. Further analysis shows that one of the criterion, based on the demand to plan the text in a separate part of the screen before writing, was not suitable for use with a word processor (d 0.35). The planning process seems to be different in the digital environment: Pupils start to write the text without planning, despite the teacher's instructions. They plan and process the text at the same time they are writing it on the screen (Nordmark 2014, 191–192). The screen seems to function as a space for the externalization of thoughts, and the pupils uses the screen and the keyboard to organize them (Åkerfeldt 2014b).

In the case of the description task, problems occurred with the length of the text: some texts were really short—only one sentence or two—and could neither compose a complete description (d = 0,52) nor illustrate creative vision (d = 0,34). If the text is short, the writer cannot succeed in some other criterion as well, such as "descriptive adjectives, substantives, or verbs", appealing "to at least two senses (e.g. sight and hearing)", or "one's own creative impression". Only 28% of the boys chose the description (38% of the girls).

In both texts, the difference was strong in the criterion "the layout of the text" (d = 0,74 in description; d = 0,44 in news). Many texts were really short or consisted only of a string of sentences, though the platform created possibilities for the different sizes and styles of fonts or separated sections. One reason for this result can be the habit—at least at school—to print out texts and check their lines of thought, structure, and linguistic forms as well as layout on a physical paper copy. Therefore, a digital environment may have presented new challenges to writing insofar as all pupils were not equally familiar with writing texts on a computer (e.g. Pommerich 2004).

It is worth considering that genres, and the relationships between them, change over time, with the conventions of each genre shifting as well. For instance, contemporary news need not only be an event (e.g. an accident)—that always contains a clear description of the situation, time, space of action, reasons and consequences—as the criterion suggested. If a pupil wrote news of a research result, such a text could not fulfill all criterion. The situation would be the same if the news concentrated of a new exhibition. It is noteworthy that if the linking item were disregarded in the overall writing achievement, leaving the application and the rejoinder as the only assignments to be considered, the printed and electronic writing assessments would have practically yielded the same results.

In general, among those pupils who were aiming at upper secondary school, there were no statistically significant differences between the two test modes. However, among those pupils aiming at vocational studies, those that participated in the paper version performed significantly ($p < 0,001$) better in both assignments compared to those participating in the digital version. In the task of writing a description, the effect was quite strong (d = 0,70).

The association of parents' educational background with learning outcomes in two test modes differed from linguistic competence. In the writing description test, there was no difference between the two test modes if both parents had completed the matriculation examination. If only one or neither of the parents had not completed the examination, the pupils who completed the paper version performed significantly ($p < 0,001$; d = 0,40) better. In writing the news, there was a similar ($p < 0,05$) difference between test modes in the group of students whose parents had not completed the matriculation examination.

When more precisely examining the description and news writing assignments, the differences in performances between the paper and digital version varied across the different provinces of Finland. In writing the description, the differ-

ence between the two test modes in favor of the paper version was the largest (23 percentage points), and both statistically ($p < 0{,}001$) and remarkably ($d = 0{,}98$) significant in Southwestern Finland. The difference was also statistically significant in Eastern Finland, and Western and Inland Finland ($p < 0.001$), though the mode effect (difference from 10 to 13 percentage points) was a little smaller ($0.38 < d < 0.48$). In writing the news, the only statistically significant difference ($p < 0{,}001$) was found in the province of Southwestern Finland. The average score for pupils working with the paper version of the assessment was 10 percentage points higher compared with the pupils using the digital version ($d = 0{,}46$).

There could be many reasons for the differences between regions in Finland. There is some evidence that the resources and the use of ICT in schools vary (between regions and within regions). However, systematical research on this has not been done yet.

Discussion

Our study shows that it is very challenging or even impossible to use linking assignments and transfer their criterion directly into digital form if they were originally designed for a pen-and-paper assignment. The use of digital tools highlights and requires a different kind of approach, skills, and competencies than the use of paper and the pencil. The architecture of the assessment tools has to be meticulously designed and developed as well (Neal 2011; Reilly and Atkins 2013). Åkerfeldt (2014a, 87) underlines that digitalization challenges the notion of competence: what kind of skills will be recognized as competencies—or digital literacies (Poe 2013)?

There were major differences in the quality and availability of ICT equipment between the sample schools, and this could account for the differences, explaining some of the confusing results. The students in Eastern-Finland, for example, achieved about the highest results in writing in the pen-and-paper version. In the digital version, the students from that area achieved the lowest scores. Poe (2013) insists that the issues of fairness are particularly important in large-scale digital writing assessments concerning e.g. parents' socio-economical status, educational level and ethnicity, pupil's gender, disabilities, access to and use of digital tools and technology, previous experiences and attitudes towards these tools.

In general, the level of ICT use at school was somewhat low. Teachers in the sample schools chiefly used ICT equipment in the classroom for information

retrieval and the editing and laying out of texts. This fact may explain the result that peer-reviewed scoring of the teacher had a good correspondence in the pen-and-paper version, but the correspondence was much lower in the digital version. The level of the use of ICT does not necessarily imply that the teachers are unwilling to use ICT. The availability and the quality of ICT varies remarkably between schools. It is not the case that all the pupils and teachers have a possibility to use computers and internet in their classrooms. Instead, they have to use separate almost fully booked computer labs. Moreover, the learning materials designed for digital learning environments are diverse.

Unfortunately, very little is known about studying writing in the digital classroom when teaching mother tongue (Finnish) in lower secondary education (Kauppinen et al. 2015). In Finland as well as in Sweden (Nordmark 2014, 247), the research on teaching Finnish has focused on text as a product of writing, not on the processes of digital writing.

All in all, digital assessments are coming to education to stay, and based on our results, there is no reason to fear them. Nevertheless, in the future, pupils must be guaranteed uniform opportunities to use ICT equipment across Finland, in different types of municipalities, and in all schools. Hardware and its use should be reformed and standardized, providing the basis for equality. Teacher training must offer prospective teachers the capabilities for using ICT in teaching and for assessing texts produced online (multimodal texts). Teachers should be offered continuing education in this field. ICT skills are part of modern civics.

References

Åkerfeldt, A. "Didaktisk design med digitala resurser. En studie av kunskapsrepresentationer I en digitaliserad skola". *Doktors avhandlingar från Institutionen för pedagogic och didaktik* 32. Stockholm: Stockholm universitet, 2014a.

Åkerfeldt, A. "Re-shaping of writing in the digital age. A study of pupils' writing with different resources." *Nordic Journal of Digital Literacy* 4, 3 (2014b), 172–193.

Chua, P. Y., and M. D. Zuraidah. "Effects of computer-based educational achievements tests on test performance and test takers' motivation". *Computers in Human Behaviour* 29 (2013): 1889–1895.

Clariana, R., and P. Wallace. "Paper-based versus computer-based assessment: key factors associated with the test mode effect". *British Journal of Educational Technology* 33, 5 (2002): 593–602.

Cohen, J. *Statistical Power Analysis for Behavioral Sciences*. Second edition. New Jersey: Lawrence Erlbaum Associates, 1988.

European Commission 2009. National Testing of Pupils in Europe: Objectives, Organisation and Use of Results. Bryssel: Education, Audiovisual and Culture Executive Agency. http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf.

Eyal, L. "Digital Assessment Literacy—the Core Role of the Teacher in a Digital Environment". *Educational Technology and Society* 15, 2 (2012): 37–49.

Harjunen, E., and J. Rautopuro. *Kielenk*äytön ajattelua ja ajattelun kielentämistä: Äidinkielen ja kirjallisuuden oppimistuokset perusopetuksen päättövaiheessa 2014: Keskiössä kielentuntemus ja kirjoittaminen. Julkaisut 2015:8. Helsinki: Kansallinen koulutuksen arviointikeskus, 2015a.

Harjunen, E. and J. Rautopuro. "Thinking about language and language thinking. Learning outcomes in mother tongue and literature at the end of basic education in 2014". Summary. Helsinki: Finnish Education Evaluation Centre, 2015b.

Hildén R. and J. Rautopuro. Ruotsin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013. Julkaisut 2014:1. Tampere: Kansallinen koulutuksen arviointikeskus, 2014.

Kaarakainen M.-T., K. Osmo, T. Katja. "Kouluikäisten tietoteknologian vapaa-ajan käyttö". *Nuorisotutkimus* 2/2013: 20–33.

Kauppinen M., J. Pentikäinen, M. Hankala, P. Kulju, E. Harjunen and S. Routarinne. "Systemaattinen katsaus perusopetusikäisten kirjoittamisen opetusta ja osaamista koskevaan tutkimukseen". *Kasvatus* 46, 2 (2015): 160–174.

Ford Lawton, D. "Beyond Bubble Sheets and Number Two Pencils: Assessment in the Digital age". *The Delta Kappa Gamma Bulletin. International Journal for Professional Educators*. Volume 81, 1 (2014): 53–58

Kim, D.-H., and H. Huynh. "Comparabilitys of Computer and Paper versions of Algebra and Biology Assessments". *The Journal of Technology, Learning and Assessment* 6, 4 (2007): 1–31.

Laakso, M.-J., T. Rajala, E. Kaila, and T. Salakoski. "The Impact of Prior Experience in Using a Visualization Tool on Learning to Program". *Proceedings of CELDA 2008, Freiburg, Germany*: 129–136.

Lahti, J., S. Heinonen, E. Siira and M. Lattu. Korkean panoksen sähköiset kokeet maailmalla. Digabiprojektintyöpaperi. 2013. https://digabi.fi/wordpress/wp-content/uploads/2014/02/Korkean-panoksen-s%C3%A4hk%C3%B6iset-kokeet-maailmalla.pdf.

Lehtinen, E. "Teknologian kehitys ja oppimisen utopia". *Oppimisen teoria ja teknologian opetuskäyttö*. Eds. S. Järvelä, P. Häkkinen, E. Lehtinen. Helsinki: WSOY, 2006.

Masters, G. N. "Reforming Educational Assessment: Imperatives, principles and challenges". *Australian Education Review* no 57. Australian Council for Educational Research. Camberwll, ACER Press, 2013.

Neal, M. R. *Writing Assessment and Re Revolution in Digital Texts and Technologies*. New York: Teachers College, Columbia University, 2011.

Nordmark, M. "Digitalt skrivande I gymnasieskolans svenskundervisning. En ämnesdidaktisk studie av skrivprocessen". Örebro studies in Education 45. Örebro Studies in Educational Sciences an Emphasis on Didactics 9. Örebro: Örebro universitet, 2014.

Noyes, J., K. Garland, and L. Robbins. "Paper-based versus computer-based assessment: is workload another test mode effect?" *British Journal of Educational Technology* 35, 1 (2004): 111–113.

Poe, M. "Making digital writing assessment fair for diverse writers". *Digital Writing. Assessment and Evaluation.* Eds. H. A. McKee and D. N. DeVoss. Computers and Composition Digital Press/Utah State University Press, 2013. http://ccdigitalpress.org/dwae/index.html.

Pommerich, M. "Developing Computerized Versions of Paper-and-Pencil Tests: Mode Effect for Passage-Based Tests". *The Journal of Technology, Learning and Assessment* 2 (6) (2004): 1–44.

Rautopuro, J. (ed.). *Hy*ödyllinen pakkolasku. Matematiikan oppimistulokset peruskoulun päättövaiheessa. Koulutuksen seurantaraportti 2013:3. Tampere: Opetushallitus, 2013.

Tossavainen, T. Tekniikka ei saa olla kouluissa itsetarkoitus. Helsingin Sanomat, Vieraskynä 24.1.2013.

Weigle, S. C. *Assessing writing.* Cambridge: Cambridge University Press, 2011 [2002].

Williams, P. J. "Maximising the potential of ICT to provide authentic summative assessment opportunities". *Computers in New Zealand Schools*, 24, 2 (2012): 137–155.