

# Associations Between Teachers' Attitudes Toward the Final Exams in the Norwegian Language Arts Subject and Perceived Washback on Instruction and Assessment: A Survey of Norwegian Secondary School Teachers

Harald Eriksen<sup>1\*</sup>, Aslaug Fodstad Gourvennec<sup>2</sup>,  
Vibeke Rønneberg<sup>2</sup> & Marte Blikstad-Balas<sup>3</sup>

<sup>1</sup>Oslo Metropolitan University, Norway; <sup>2</sup>University of Stavanger, Norway;

<sup>3</sup>University of Oslo, Norway

## Abstract

The issue of washback effects from tests on instruction has been extensively investigated. The nature and extent of washback depend on many factors, some of which are related to teachers and exams. Work to redesign the national written exams in Norwegian Language Arts (NLA) for secondary school began in 2020, placing the issue of washback effects from exams on teachers' instruction and assessment on the agenda. However, there is no systematic knowledge about the extent and nature of washback from the current exams. The present study reports on a survey of NLA teachers in lower- and upper-secondary school ( $N = 508$ ) by investigating the washback effects they perceive and their attitudes toward the existing exams, as well as exploring how perceived washback and attitudes are related using structural equation modeling (SEM). The results show that, as regards instruction, teachers believe there is a washback effect from the NLA exams. This effect on instruction is strongly associated with teachers' attitudes toward the exams. In contrast, the association between the teachers' attitudes and the washback they perceive on their own formal assessment practices is rather weak in numerical terms.

**Keywords:** *L1; backwash; writing assessment; writing instruction; national exams*

Responsible editor: Solveig Troelsen

Received: December, 2022; Accepted: October, 2024; Published: December, 2024

---

\*Correspondence: Harald Eriksen, e-mail: [haralde@oslomet.no](mailto:haralde@oslomet.no)

© 2024 Harald Eriksen, Aslaug Fodstad Gourvennec, Vibeke Rønneberg & Marte Blikstad-Balas. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

*Citation:* H. Eriksen, A. F. Gourvennec, V. Rønneberg & M. Blikstad-Balas. "Associations Between Teachers' Attitudes Toward the Final Exams in the Norwegian Language Arts Subject and Perceived Washback on Instruction and Assessment: A Survey of Norwegian Secondary School Teachers" *Nordic Journal of Literacy Research, Special issue: Temanummer om statlig måtning av läs- och skrivförmåga*, Vol. 10(2), 2024, pp. 72–92. <http://doi.org/10.23865/njlr.v10.5309>

## **Introduction**

Exams are often held at the end of a unit, semester, or school year to measure learning over the respective period, and the content and format of an exam tend to influence teaching during that period (Tsagari & Cheng, 2016). If what the exam measures is well aligned with the overall curriculum, this may not be a problem. However, if an exam covers only part of the content of the curriculum, such a “washback” effect on the instruction leading up to it may be problematic. In the case of language teaching and learning, the washback effect of tests and exams has been extensively investigated across countries (Green, 2013; Tsagari & Cheng, 2016). However, this research is not conclusive in terms of the nature and extent of the washback effect (e.g., Cheng & Curtis, 2004; Spratt, 2005). In 2020, when the Norwegian Directorate for Education and Training (NDE) initiated a revision of the national written exams in the subject of Norwegian Language Arts (NLA) held at the end of lower- and upper-secondary school, this drew increased public attention to the potential washback effect of those exams. Since this issue has been little investigated, it is difficult to have research-based discussions about what kind of written assessment NLA students would benefit from—not only on their final exams, but also considering the potential washback effect. It should be added that these exams as presently designed differ in several ways from the tests and exams typically explored in existing washback research. Hence, the current situation calls for research to be conducted in a Norwegian setting for two reasons: First, to enable informed decisions about the future of the national NLA exams and, second, to contribute knowledge to the international research base from a qualitatively different context.

The Norwegian exam system has long been known to prioritize extended writing (Blikstad-Balas et al., 2018; Hertzberg & Roe, 2016; Ongstad, 2002). When the NDE initiated revision of the exams in secondary school (Grades 10 and 13) in 2020, the proposals made included substituting one or two long writing tasks with several shorter writing tasks. This addressed both a particularly low rater reliability among NLA-exam assessors (Björnsson & Skar, 2021) and an aim to evaluate NLA competence more broadly with regard to content and genres than the current exams do. However, this proposal provoked a public debate about washback effects on instruction and assessment during the school year—of both the current exams and the suggested new ones. One of the major concerns expressed related to a possible washback effect leading to less extended writing instruction and less composition of complex texts in favor of shorter and less complex tasks (e.g., Gourvenec et al., 2021; Jambak, 2021; Valand, 2022).

The present study aims to respond to this situation by exploring the washback effects of the current exams that secondary-school NLA teachers report on their own writing instruction and assessment. It also aims to investigate whether teachers' reports of washback effects are associated with their attitudes toward the exams. We do this by exploring the following research questions:

1. What is the washback from the final written NLA exams in secondary school on teachers' writing instruction and assessment practices as they perceive it?
2. What attitudes do secondary-school NLA teachers have toward the final written NLA exams?
3. Is there an association between teachers' attitudes toward the final exams and their perceived washback on instruction and assessment?

### ***Research background***

Researchers largely agree that washback is a complex phenomenon—as demonstrated, for example, by the wide variety of factors often investigated, including both negative and positive effects of tests (Ahmed & Rahman, 2019; Kuang, 2020; Thu, 2020). Further, there is a high level of agreement in existing research that washback is context-specific (Kuang, 2020; Thu, 2020; Tsagari & Cheng, 2016). Hence, a large number of contextual factors must be considered whenever previous washback research is relied on to inform educational policy. Such contextual factors include differences between countries in teaching practices, grading, and test loads.

Another relevant contextual factor is, of course, the school subject involved. In the case of language arts (LA), Anagnostopoulos (2005) investigated how tests became operative through classroom tasks in literature instruction in U.S. high schools, finding that the teachers' written tasks and classroom talk reinforced a restricted concept of terms such as “assessing” and “critiquing”. Hence the upcoming tests seemed to intensify “a narrowing of the curriculum” already present in school. The conclusion drawn was that “tests can endorse narrow definitions of reading that work against rather than support efforts to raise standards and improve students' learning opportunities” (Anagnostopoulos, 2005, pp. 59–60). In a discussion of washback on literature instruction, also in the U.S., Shelton and Brooks (2019) argue that a reliance on short text excerpts in standardized tests may fragment students' reading experiences and may discourage schools and English Language Arts (ELA) teachers from engaging with entire literary works, thus exerting a negative effect on teachers' instruction. Further, Applebee and Langer (2011) document that U.S. ELA teachers allow the high-stakes tests to have a strong influence on the writing curriculum and instruction in ELA in both middle and high school. In fact, those authors noted that very little writing was reported to take place in the high-stakes years, including writing as any kind of open-ended response, from composing single sentences to essays. Instead, teachers tended to align their own assessment practices with those of upcoming high-stakes tests emphasizing multiple-choice items and short-answer questions (Applebee & Langer, 2011, p. 18). With regard to this last finding, it is worth noting that the situation in U.S. ELA classrooms, where much of the research stems from, clearly differs from that in NLA classrooms with regard to the writing activities taking place in that Norway has historically prioritized sustained writing assignments both in classroom instruction and in the final exams (Hertzberg & Roe, 2016; Ongstad, 2002).

While numerous studies have examined washback in an international context, there is little Norwegian, or even Nordic, research. However, three studies from Sweden and Norway have shed some light on the washback effect of LA-related exams. First, Larsson and Olin-Scheller (2020) investigated the washback effects of the Swedish national tests taken in the first year of upper-secondary school. That test consists of three parts: reading, writing, and speaking. Unlike the studies in a U.S. ELA context, these authors found that the reading-comprehension test had an insignificant effect, but the written and oral tests did seem to have had a substantial washback effect. Writing instruction was affected in that teachers designed tasks imitating those of the test, both to prepare students for the test and to create opportunities for their own assessment—meaning that the washback effect also indirectly influenced the final course grade, which is set by the teacher and is independent of a student's score on the national test. Second, Hovdhaugen et al. (2018) found that Norwegian upper-secondary school teachers of NLA and mathematics interpreted any discrepancies between the final course grades set by themselves and the exam grades set by external examiners as indicating a need to reconsider, and possibly adjust, their own teaching and assessment practices. It might be claimed that this reflects an expectation on the part of teachers that the competence they test throughout the year should mirror those tested on the final exam. Third, we found in our own previous work (Gourvennec et al., 2023) that secondary-school NLA teachers report what they frame as both positive and negative washback effects on instruction from the NLA written exams. Interestingly, what they framed as positive effects—such as the exams providing directions for both students and teachers to follow within a vast and complex school subject—are closely related to those they framed as negative—such as causing teachers to prioritize writing tasks in the classroom that resemble those given on the exams rather than assigning more creative or collaborative tasks. Overall, the limited research that we have on washback from a Nordic context thus suggests that the format and types of tasks given on an exam often exert an impact on what LA teachers prioritize when planning their instruction and assessment.

### ***Background to the choice of variables in the present study***

In a review of research on washback from LA exams, Spratt (2005) identifies several factors—such as teacher-related factors, materials such as past papers and exam-related textbooks, the school, and the exam—that have proven to affect the degree and kind of washback effect. The teacher-related factors include: (i) Teachers' beliefs about the reliability, fairness, and utility of the exam, about its alignment with teaching materials and with their teaching practices, about effective teaching methods and teaching philosophy, and about the stakes and usefulness of the exam; (ii) Teachers' attitudes toward the exam; and (iii) Their teacher training. The exam factor includes the stakes involved, the status of the language tested, the purpose of the exam, the formats it employs, the weighting of individual papers, when the exam was introduced, and how familiar it is to teachers (Spratt, 2005, p. 23). One important conclusion

from Spratt's study is that teachers mostly seem to be in control of how the washback operates (meaning that they could resist it if they wanted to). This represents an argument in favor of the usefulness of investigating to what extent teachers *perceive* a washback effect in the Norwegian context.

A more recent review by Thu (2020) presents a framework for washback research with three principal areas of investigation—teaching practices (including assessment practices), perceptions of teaching, and attitudes toward the test (Thu, 2020, p. 7). In the following, we will elaborate on how these areas connect to the development of our survey.

Instruction is a multifaceted and context-sensitive field, with regard to aspects such as teaching content, materials, methodology, skills, the language of instruction, the allocation of time to different skills, collaboration between teachers, and professional development. Thu (2020) places those various aspects of instruction at the center of the framework and calls for more research on the complex connections between them. In the Norwegian context, there seems to be an overlap between writing instruction as observed in NLA classrooms (Blikstad-Balas et al., 2018) and the writing exams, with both prioritizing sustained writing, and the format of the exams has indeed been suggested as an explanation for the sustained writing often seen in Norwegian classrooms (Smidt, 2007). This is why we have included items about washback on *instruction* in our survey. We have done so by adapting one item from a study by Collins and Miller (2018) on English teachers' attitudes toward the Test of English as a Foreign Language (TOEFL). Our hypotheses are that teachers will report a high degree of washback on instruction and that the extent of washback on instruction reported will be positively correlated with attitudes toward the exams.

Washback on *assessment* seems to be less investigated within the field of washback research. In fact, it is not included in the review articles by Green (2013) or Tzagari and Cheng (2016). However, it is part of the framework developed by Thu (2020). The low prominence of assessment could be due to the fact that many studies (e.g., Ahmmed & Rahman, 2019) are conducted in a context where there already exists a national or central testing system, and where an important aim of teaching and instruction is for students to pass a specific test, meaning that research is likely to focus on the potential effect of that test on classroom instruction. However, given the close connection between instruction and assessment that exists in Norway, we wished to explore washback effects on assessment as well. To this end, we developed our own items. Our hypotheses are that teachers will report that there is a washback effect on assessment, that there will be found to exist statistical associations between perceived washback on instruction and perceived washback on assessment, and that teachers' attitudes toward the exams will be positively correlated with the washback they report on their own assessment practices.

Exploring the connection between teachers' *attitudes* toward a test and the potential impact of that test is an established approach in research on washback in language testing (Spratt, 2005; Thu, 2020). Such attitudes seem to be influenced by a

range of contextual factors. For example, teachers' typically very limited influence on test design may provoke feelings of tension and anxiety in them (Thu, 2020, p. 161). Further, the extent to which a test is deemed suitable for its purpose also seems to influence teachers' perceptions of the washback effect. Also, if a test represents "something new" and if it includes features such as a high degree of goal orientation and sufficient preparation materials, prospectively leading to more time-efficient teaching, the washback effect it causes tends to be perceived as positive (Thu, 2020, p. 162). To investigate teachers' attitudes, we translated and adapted three items from Collins and Miller (2018). Our hypothesis is that there will be a statistically significant correlation between teachers' attitudes toward the exams and the washback effects they perceive to exist on instruction and assessment.

The studies presented above together constitute the background of the development of our latent variables, which are teachers' perceived washback on instruction and on assessment, as well as their attitudes toward the exams.

### ***The Norwegian educational setting***

In Norway, mandatory schooling consists of Grades 1–10, with students starting in Grade 1 in the fall of the calendar year in which they turn six. Lower-secondary school consists of Grades 8–10 (students aged 13–16) and is administered by municipalities. Upper-secondary school, which consists of Grades 11–13 (students aged 16–19), is administered by regions. Although upper-secondary school is optional, approximately 95% (Statistics Norway, 2022) of the students who complete Grade 10 go on to enroll in upper-secondary school, where approximately half the students attend vocational programs and the other half attend academically oriented programs.

In Norwegian secondary schools, students are graded in all subjects. Their report card at the end of Grades 10 and 13 includes three grades in NLA: one in each of the two written forms of standard Norwegian (*Bokmål* and *Nynorsk*)<sup>1</sup> and one in oral Norwegian. *Bokmål* and *Nynorsk* are both taught to all students throughout their schooling years. However, in secondary school, each student is free to choose which form is to be his or her first choice. The grades on the report card, which are set by classroom teachers, make up the major share of the students' final certificate. In addition, the students are required to take some exams, and the grades obtained on those constitute the rest of the diploma.

In Grade 10, all students take a written exam in either NLA, mathematics, or English (decided by lot). The NLA exam covers two days, one devoted to the student's first-choice form of written Norwegian and one to his or her second choice. In Grade 13, all students are required to pass a written NLA exam (pertaining to their first-choice form of Norwegian). They may also have to take an exam in their

---

<sup>1</sup> The two forms are very closely related, but they have different origins: *Bokmål* can be seen as an adapted form of Danish, whereas *Nynorsk* is built on Norwegian dialects. In 2023–2024, 11.4% of students in Grades 1–10 had *Nynorsk* as their first-choice form.



second-choice form (again decided by lot). Their results on the exam (or these exams) influence the overall grade on the final diploma.

All grades on the diploma are to reflect the student's proficiency at the end of grade 10 or 13, relative to the definitions of competence set out in the syllabus for each subject. However, whereas a grade on the report card is informed by various assessment situations over time, and is mostly decided by the classroom teacher, the Ministry of Education and Research stresses that the exam grades are based on a more limited assessment and represent a single occasion (Ministry of Education and Research, 2016, p. 63). Furthermore, these grades are set by examiners who are not familiar with the students. Obvious examples of how the basis for the two types of grades may diverge include differences in text types and in genres represented, as well as in opportunities for process writing.

As noted above, students are given three report-card grades in NLA. This in itself means that NLA grades strongly impact students' final-diploma grades and hence choices for upper-secondary school and higher education and career opportunities. Since the NLA exam is the only exam taken by those tenth-graders who are assigned that subject by lot, and since thirteenth-graders take one or two NLA exams among a limited number of exams, NLA exams also have a strong impact on the diploma. Hence, they may be characterized as *high-stakes* exams.

Up to the time of our survey, the written NLA exams have long been designed in the following manner. For Grade 10, both exam days include the composition of two texts—one mandatory short text and another rather long text (typically several pages of connected text), where students may choose from several different themes and genres. For Grade 13, the design is similar but comprises either one long text of choice or a combination of a mandatory short text and a long text—each of these two designs is used for either the first-choice or the second-choice form of written Norwegian (with the distribution changing from one year to another). For their long texts, students may typically produce informative, argumentative, expository, or creative texts, or literary interpretations.

## **Method and material**

### ***Sample***

A total of 508 Norwegian teachers in lower- and upper-secondary schools responded to our survey. Of those, 110 in fact completed the pilot survey, but they were included in the final sample because no closed items were added and only a few items were slightly adjusted. The other 398 participants completed the main survey. The pilot study was conducted in the county of Rogaland in April–May 2022. The main survey included three additional counties (Agder, Viken, and Trøndelag) and was conducted in June 2022. A total of 296 of the participants taught in upper-secondary school and 212 taught in lower-secondary school. The participants' mean age was 45.7 years and their average length of teaching experience was 16.3 years.

To recruit NLA teachers to participate participants for the survey, we reached out to a combination of rural and urban public lower-secondary schools and to all public upper-secondary schools in the four counties offering an academically oriented study program. To be able to measure the response rate (see Table 1), we asked each school to provide us with the email addresses of their NLA teachers. Reminders were sent to schools that did not respond. An invitation to the survey was sent to all the email addresses provided. Reminders were sent to teachers who did not respond to the survey after the first invitation. Teachers provided their consent to participate in the study on the first page of the survey. Email addresses were only handled by the data manager and were deleted once the main survey was closed and the data had been checked for duplicates. Out of a total of 779 invitations, a response rate of 65.2% was obtained. Compared with the average online-survey response rate of 44.1% identified by Wu et al. (2022), we see this as a rather strong response rate.

*Table 1.* Overview of the recruitment process and sample

<b>Pilot survey</b>	<b>Total</b>	<b>Upper secondary</b>	<b>Lower secondary</b>	<b>Main survey</b>	<b>Total</b>	<b>Upper secondary</b>	<b>Lower secondary</b>
Schools invited	57	19	38	Schools invited	220	107	113
Schools that provided email addresses	43	17	26	Schools that provided email addresses	112	61	51
Email addresses received	231	133	98	Email addresses received	548	301	247
Respondents	110	71	39	Respondents	398	225	173

### ***The survey instrument***

As part of the development of the survey, we conducted a qualitative pilot study involving one teacher from a lower-secondary school and one from an upper-secondary school. This study was conducted digitally on the Zoom platform as a “think-aloud”, where the teachers responded to the survey and explained how they understood the questions. The teachers shared their screens with the researchers during the pilot. A third person—a faculty member with extensive teaching experience from upper-secondary school—also completed the survey and provided written feedback. Based on the three participants’ feedback, we adjusted some of the items.

In the next phase, 110 teachers answered a pilot survey that consisted of items covering the following latent variables: washback on instruction, washback on assessment, and perceptions of the validity, reliability, and fairness of the exams. Item development was partially informed by previous research as described above. In addition, the survey included items designed to explore the teachers’ backgrounds and writing-assessment practices. To enhance the quality of our measurement, we included three reversed items to identify self-contradictory responses (and found



none). We used a seven-point Likert-like scale to measure teachers' perceptions; the indicators contained phrases ranging from 1 = "Strongly agree" to 7 = "Strongly disagree", with 4 as a neutral middle value ("Neither agree nor disagree"). Since the constructs in this study are complex and could not be measured adequately with single items, we employed a multi-item approach. We performed analysis of variance, skewness and kurtosis, and scale reliability to ensure sufficient variance and to test the measurement of the underlying constructs (Pallant, 2020). Results showed that nine items had a range of 1–7 while two had a range of 1–6. The standard deviation (SD) was overall above 1 and below 1.5. All values of skewness and kurtosis were within the rule-of-thumb values of 1/-1 and 3/-3, respectively. Scale analysis showed that Cronbach's Alpha (CA) for perceived washback on instruction was 0.84, perceived washback on assessment 0.71, and attitudes towards the exams 0.89. Given these findings, we deemed it sound to proceed with the survey. As noted above, the changes made following the pilot survey were so minor that we included its participants in the study.

For the final phase, we still wanted to measure the underlying constructs of washback on instruction, washback on assessment, and teachers' attitudes toward the current exams. All items are presented in Table 2.

Table 2. Overview of theoretical constructs and items

Construct	Item (in Norwegian)	English translation
Washback on instruction	SU_12: Eksamensrettet undervisning stemmer godt overens med de overordnede målene i norskfaget.	Teaching to the exam aligns well with the primary goals of the NLA subject.
	SU_13: Eksamen støtter meg i å velge innhold og oppgaver i undervisningen.	The exam supports me in choosing content and tasks for instruction.
	SU_14: Eksamen viser meg hvilke ferdigheter det er viktig å fokusere på i undervisningen.	The exam shows me what skills are important to focus on in my teaching.
Washback on assessment	SU_04: Vurderingssituasjonene gjennom semesteret ligner eksamen.	The assessment situations during the semester are similar to the exam.
	SU_06: Å gjøre elevene eksamensforberedt er et svært viktig hensyn når jeg utformer vurderingssituasjoner på 10. trinn/vg3.	Preparing students for the exam is a very important consideration when I design assessment situations in Grade 10/Grade 13.
	SU_08: Jeg har større grad av frihet til å velge oppgavetyper i vurderingssituasjoner på 8. og 9. trinn/vg1 og vg2 enn i avgangsåret.	I have more freedom when selecting tasks for assessment situations in Grades 8–9/11–12 than during the final years [Grade 10/13].
Attitude toward exams	EK_13: Eksamenskarakteren er et korrekt mål på norskfaglig kompetanse.	The exam grade provides a valid measure of NLA proficiency.
	EK_14: Eksamenskarakteren gir et mer korrekt bilde av elevens norskfaglige kompetanse enn standpunkt karakteren gjør.	The exam grade provides a more valid indication of the student's NLA proficiency than the report-card grade does.
	EK_15: Eksamenskarakteren gir et presist bilde av elevenes norskfaglige kompetanse.	The exam grade provides an accurate indication of students' NLA proficiency.

There were no missing data, since we made it compulsory for participants to answer each item in order to be able to proceed with the survey. It was possible to indicate a neutral answer to each question and to add additional comments through six open-ended questions.

### ***Data analysis***

We first performed a descriptive analysis in which we inspected the range, mean, standard deviation, skewness, and kurtosis for each survey item, as well as response-option frequencies. Next, we conducted an exploratory factor analysis (EFA) to obtain insight into the underlying factor structure (Pallant, 2020). The results from the EFA were used to determine latent factors, which in turn were analyzed using structural equation modeling (SEM) for confirmatory factor analysis (CFA); this method is suitable for the analysis of complex datasets (Kline, 2016). We used IBM SPSS Amos 29 for the structural equation modeling.

### **Results**

When reporting our results below, we will first provide descriptive statistics for the items included in the study and then report the findings related to each research question in turn.

#### ***Descriptive statistics***

Descriptive statistics for the nine items included in this study are presented in Table 3. Skewness and kurtosis are measures of the distribution of responses. As a rule of thumb, values above or below one and three, respectively, indicate a deviation from normality in the distribution. None of the items reached either of these levels. Further, the descriptive statistics also demonstrate that each of the three latent variables (corresponding to the constructs in Table 2) showed sufficient variability to be included, and they indicate the extent to which the respondents' answers are aligned with a normal distribution.

For the first latent variable, "washback on assessment", the scale ( $\alpha = 0.63$ ) indicates a moderate level of internal consistency. Cronbach's Alpha values are sensitive to the number of items (Pallant, 2020), and the small number of items involved here could partially explain the relatively low value; with five or more items instead of three, a higher value would be more likely. For the second latent variable, "washback on instruction", the scale ( $\alpha = 0.79$ ) indicates an acceptable internal consistency. Finally, for the items used to operationalize the last latent variable, "attitude toward exams", the scale ( $\alpha = 0.86$ ) shows a strong internal consistency.

Table 3. Descriptive statistics (N = 508)

Items	Range	Mean	SD	Skewness	Kurtosis	Cronbach's Alpha
SU_12: Teaching to the exam aligns well with the primary goals of the NLA subject.	1–7	3.4	1.4	0.3	–0.4	0.79
SU_13: The exam supports me in choosing content and tasks for instruction.	1–6	2.9	1.3	0.6	–0.0	
SU_14: The exam shows me what skills are important to focus on in my teaching.	1–7	2.5	1.2	0.9	1.4	
SU_04: The assessment situations during the semester are similar to the exam.	1–7	2.8	1.4	0.4	0.3	0.63
SU_06: Preparing students for the exam is a very important consideration when I design assessment situations in Grade 10/Grade 13.	1–6	2.0	1	0.9	1.1	
SU_08: I have more freedom when selecting tasks for assessment situations in Grades 8–9/11–12 than during the final years [Grade10/13].	1–7	2.2	1.3	0.9	0.4	
EK_13: The exam grade provides a valid measure of NLA proficiency.	1–7	3.5	1.3	–0.3	–0.2	0.86
EK_14: The exam grade provides a more valid picture of the student's NLA proficiency than the report-card grade does.	1–7	5.1	1.5	–0.9	0.4	
EK_15: The exam grade provides an accurate picture of students' NLA proficiency.	1–7	4.1	1.5	0.0	–0.5	

We conducted an EFA indicating three separate factors based on principal factor extraction with Oblimin rotation (see Table 4). In the empty sections, factor loadings below 0.3 were omitted (Comrey & Lee, 2013). The main loadings are highlighted in bold.

Table 4. Exploratory factor analysis (N = 508)

Item	Factor 1	Factor 2	Factor 3
SU_12	0.45		<b>0.69</b>
SU_13	0.46	0.42	<b>0.92</b>
SU_14	0.42	0.49	<b>0.63</b>
SU_04		<b>0.56</b>	0.31
SU_06		<b>0.95</b>	0.35
SU_08		<b>0.42</b>	
EK_13	<b>0.83</b>		0.41
EK_14	<b>0.92</b>		0.59
EK_15	<b>0.83</b>		0.50

As shown in Table 3, the results reveal that the three latent variables are separate from one another but share some common variance. A scree-plot analysis shows two clear factors where the initial eigenvalue percentages of variance are 42.2% and 18.6%, and then a third one at 9.6%. The close theoretical and practical connection between assessment and instruction in the Norwegian context might explain the shared variance between the factors.

***Perceived washback effects from the written exams on teachers' writing instruction and assessment***

In the following section, we report the findings related to the first research question. Frequencies are presented in Tables 5 and 6.

Table 5. Frequencies for the “washback on instruction” latent variable

	Teaching to the exam aligns well with the primary goals of the NLA subject.	The exam supports me in choosing content and tasks for instruction.	The exam shows me what skills are important to focus on in my teaching.
<b>Strongly agree</b>	6.7	12.6	19.9
<b>Agree</b>	19.7	26.4	28.7
<b>Somewhat agree</b>	29.5	35.2	36.2
<b>Neither agree nor disagree</b>	22.2	13.4	9.8
<b>Somewhat disagree</b>	12.6	7.1	2.2
<b>Disagree</b>	7.5	5.3	2.4
<b>Strongly disagree</b>	1.8	0	0.8

For the items intended to measure washback on instruction, teachers tended to agree with the statements “The exam supports me in choosing content and tasks for instruction” (with 74.2% answering that they “strongly agree,” “agree,” or “somewhat agree”) and “The exam shows me which skills are important to focus on in my teaching” (with 84.8% answering that they “strongly agree,” “agree,” or “somewhat agree”). However, the pattern for the statement “Teaching to the exam aligns well with the primary goals of the NLA subject” was somewhat different, with more teachers disagreeing. Even so, 55.9% either “somewhat agree,” “agree,” or “strongly agree” with it. Overall, the results indicated a moderate to clear degree of agreement with the statements to the effect that classroom instruction was influenced by the NLA exam.

The results show that most teachers responded that they “strongly agree,” “agree,” or “somewhat agree” with the statements in the items related to washback on assessment. For the statement “The assessment situations during the semester are similar to the exam,” 80.1% of the teachers reported that they “strongly agree,” “agree,” or “somewhat agree”. A similar pattern was found for the statement “Preparing students

Table 6. Frequencies for the “washback on assessment” latent variable

	<b>The assessment situations during the semester are similar to the exam.</b>	<b>Preparing students for the exam is a very important consideration when I design assessment situations in Grade 10/Grade 13.</b>	<b>I have more freedom when selecting tasks for assessment situations in Grades 8–9/11–12 than during the final years [Grade 10/13].</b>
<b>Strongly agree</b>	13.8	33.9	37.0
<b>Agree</b>	31.3	36.2	24.8
<b>Somewhat agree</b>	35.0	23.0	21.7
<b>Neither agree nor disagree</b>	8.9	5.1	10.2
<b>Somewhat disagree</b>	8.1	1.0	4.1
<b>Disagree</b>	2.8	0.8	1.8
<b>Strongly disagree</b>	0.2	0	0.4

for the exam is a very important consideration when I design assessment situations in Grade 10/13,” with which 70.1% claimed to “agree” or “strongly agree.” Finally, for the item “I have more freedom when selecting tasks for assessment situations in Grades 8–9/11–12 than during the respective final years [Grade 10/13],” 83.5% of teachers claimed to “strongly agree,” “agree,” or “somewhat agree” with that statement. Overall, the responses to these three items indicated that Norwegian teachers perceive that there is a clear washback effect on their assessment practices.

***Attitude toward the final written exams***

In the following section, we report the findings related to the second research question. Frequencies are presented in Table 7.

Table 7. Frequencies for the “attitude toward exams” latent variable

	<b>The exam grade provides a valid measure of NLA proficiency.</b>	<b>The exam grade provides a more valid indication of the student’s NLA proficiency than the report-card grade does.</b>	<b>The exam grade provides an accurate indication of students’ NLA proficiency.</b>
<b>Strongly agree</b>	4.3	3.1	2.6
<b>Agree</b>	15.4	5.3	10.6
<b>Somewhat agree</b>	32.5	4.1	23.0
<b>Neither agree nor disagree</b>	22.2	16.3	18.1
<b>Somewhat disagree</b>	17.7	26.2	30.9
<b>Disagree</b>	5.3	29.3	8.3
<b>Strongly disagree</b>	2.6	15.6	6.5

The responses to the items intended to measure secondary-school NLA teachers' attitude toward the final written exams indicated that most teachers "somewhat agree" or "neither agree nor disagree" with the statement "The exam grade provides a valid measure of NLA proficiency." Further, most teachers did not agree with the statement "The exam grade provides a more valid indication of the student's NLA proficiency than the report-card grade does": 15.6% of all teachers claimed to "strongly disagree" with this statement, 29.3% to "disagree" with it, and 26.2% to "somewhat disagree" with it. Finally, for the item "The exam grade provides an accurate indication of students' NLA proficiency," the teachers tended to choose the middle values on the scale: 72% of them claimed to "somewhat agree", "neither agree nor disagree", or "somewhat disagree" with that statement. In sum, the results revealed variation in the teachers' attitudes toward the ability of the final written exams to measure students' proficiency, with the majority centering around a neutral stance.

### ***Associations between teachers' attitudes and perceived washback***

In order to answer the third research question, we used structural equation modeling for confirmatory factor analysis (CFA) to investigate the strength of the relationships among the latent variables. CFA is suitable for estimating relationships between observed indicators and non-observed latent variables. In our case, we tested a measurement model, and we used acknowledged fit indices to assess the alignment between the data and model. The following indices were used to assess the fit between our model and the data: chi-square, the root mean square error of approximation (RMSEA), the Tucker–Lewis Index (TLI), and the comparative-fit index (CFI). According to Kline (2016), the values of chi-square  $p > 0.05$ , RMSEA  $< 0.05$ , TLI  $> 0.95$ , and CFI  $> 0.95$  indicate a good fit, while RMSEA  $< 0.08$ , TLI  $> 0.90$ , and CFI  $> 0.90$  indicate an acceptable fit.

Figure 1 presents the estimated model with three latent variables: "perceived washback on assessment" and "perceived washback on instruction," and "attitudes toward exams" as the dependent variable. The circles represent measurement errors, the squares represent measured variables, and the ovals represent the latent variables.

### **Standardized estimates**

**CHI = 84.671; df = 24; p-chi = 0.000; RMSEA = 0.071; CFI = 0.967; TLI = 0.951**

The fit values (of RMSEA, TLI, and CFI) indicate that the model in Figure 1 has an acceptable fit;  $p$  is not significant, and the null hypothesis cannot be rejected. However,  $N$  (508) easily yields a non-significant test (Kline, 2016, p. 270); therefore, we emphasize the values of the other fit indices to evaluate the model. The arrows between the latent variables reflect the measured strength of each connection; the strength increases with the numerical value. Washback on instruction is positively related to "attitudes toward exams" (0.7), as we hypothesized. However,



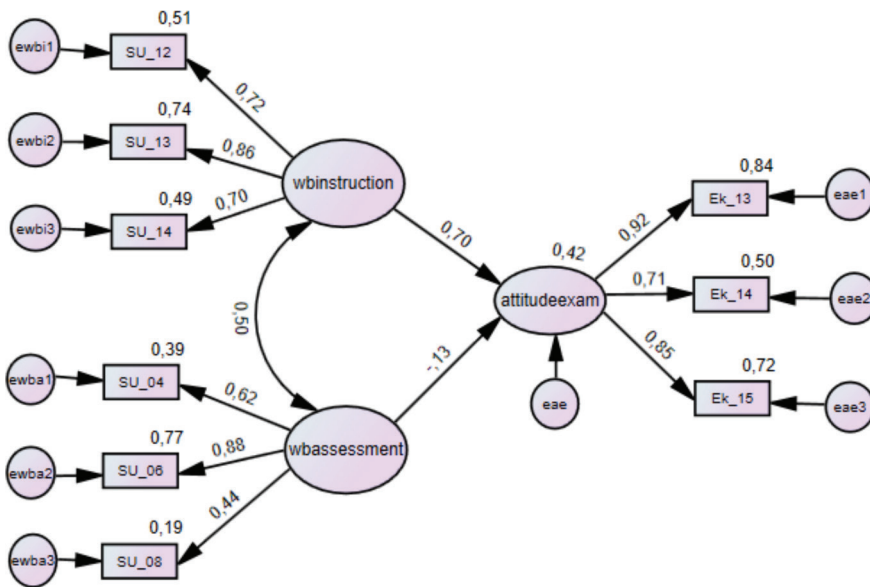


Figure 1. Model of attitudes toward exams, washback on instruction, and washback on assessment ( $N = 508$ )

the correlation between “attitudes toward exams” and “washback on assessment” is not strong ( $-0.13$ ) (Kline, 2016), which is contrary to our hypothesis. Finally, the value of 0.5 between the two independent variables (“washback on instruction” and “washback on assessment”) indicates a clear connection between these factors, as mentioned above. It is also worth noting that two of the factor loadings for the latent variable “washback on assessment” are rather weak (0.62 and 0.44). In sum, it seems as though teachers’ attitudes toward the exams do not play a prominent role in how likely they are to report a perceived washback on their assessment.

## Discussion

Situations in which traditions previously taken for granted are put at stake often trigger discussions about the nature and consequences of those traditions. This has been the case in Norway, where proposals to change the final written NLA exams have been extensively debated. Unlike many other countries, Norway has historically prioritized sustained writing opportunities for students, both in instruction and in final exams. Given that the exams in their current form and the writing instruction typically provided in Norwegian classrooms both differ substantially from the characteristics of the school contexts from which much of our knowledge about the washback effect of tests stems, this is an interesting case. A situation where a final exam may be radically changed represents a natural experiment that could yield new knowledge

about whether, and if so how, actual changes to a high-stakes test situation change teachers' understanding of their subject and their teaching plans.

The present study identifies a perceived washback effect previously undocumented in Norway. As indicated in the results section, the participating teachers agree that the exams influence how they teach and, in particular, how they assess their students. This is an important finding, because while there is research indicating that many NLA teachers tend to prioritize extended writing in their classrooms and adopt a process-oriented approach to writing (Blikstad-Balas et al., 2018; Hertzberg & Roe, 2016), we do not know whether they would still do so even if the final exams did not explicitly require all students to produce lengthy texts. Hence, the washback effect documented here raises questions about whether the longstanding tradition of sustained writing in NLA instruction is vulnerable to changes in the exam format.

Another important finding is that the teachers' reported attitudes toward the ability of the exams to measure students' proficiency are somewhat neutral. In this connection, it is interesting to note that, while the teachers report a slightly positive attitude toward the validity of the exams in relation to students' NLA proficiency, they are instead neutral or slightly negative as regards the accuracy of the exams. One possible interpretation of this apparent discrepancy might be that, whereas the validity item (EK\_13) clearly focuses on the exam construct, the accuracy item (EK\_15) explicitly brings attention to the student. Hence, we might assume that the responses to the former item reflect the teachers' conviction that the exams measure something that is central to proficiency in the NLA subject. The more negative responses to the latter item might then reflect the teachers' awareness that students' proficiency in NLA manifests in multiple ways which cannot all be captured in a single five-hour writing session.

As regards the associations between teachers' attitudes toward the exams and the washback perceived by them, it is not surprising that the washback effect on instruction is correlated with the attitudes, as depicted in Figure 1. In other words, teachers with a more positive attitude are more likely to report that the exams align well with the primary goals of the NLA subject, that the exams support them in choosing content and tasks for instruction, and that the exams show them which skills are important to focus on in their teaching.

Another important finding is that the reported washback effect is stronger for assessment than for instruction. The participating teachers report that, when they design assessment situations that resemble the final exams, they do so because they consider it important for students to be prepared for the exams, and they also report being less free to select assessment tasks in the grades in which there is a final exam (Grades 10 and 13). This suggests that changing the format of the final exams that students take may have potentially significant consequences in that students might be assessed in a new manner not only on their final high-stakes exam but also in assessment situations throughout the year preceding it. In turn, this may affect what students consider to be important skills in the NLA subject. As mentioned in the introduction, one significant concern in Norway has been whether placing less

emphasis on sustained writing in the final exams will lead to less emphasis being placed on such writing in everyday teaching practice. Our findings suggest that this concern may be legitimate, as teachers clearly confirm the presence of a perceived washback effect on their assessment practice.

Furthermore, our model estimates a weak connection between teachers' attitudes toward the exams and the washback effect on assessment that they perceive. This suggests that teachers' views on the exams are less important when it comes to explaining the presence or absence of a perceived washback effect. We propose two possible explanations for this. The first is that teachers are, to a certain extent, held responsible if there is a large discrepancy between the report-card grades they decide themselves and the exam grades set by external assessors (Hovdhaugen et al., 2018). Because of this, teachers might be influenced by the exams in designing assessment situations irrespective of their attitudes toward the exams. A second possible explanation is that teachers believe that they have a moral obligation to prepare their students for the final exam, given that it represents a high-stakes situation that will have a strong impact on the students' future. In other words, regardless of what they think of the final exam, they see it as an important task for an NLA teacher to prepare his or her students for that exam.

In Norway, there is a close connection between instruction and assessment in that teachers themselves determine the content, nature, and frequency of the writing assignments that they set their students, as well as the criteria for assessing students' performance on those assignments throughout the school year. This could explain the shared variance among the related factors in the EFA. Further, the fact that washback on assessment is operationalized using two items with relatively low factor scores could be used as an argument in favor of revising the instrument or reducing the number of items used. In fact, the numerical value of the path coefficient between the two independent variables supports a revision of the instrument. On the other hand, having fewer items will reduce the empirical basis for the construct, and we argue that even if instruction and assessment in Norwegian classrooms are connected, they are also separate phenomena. Indeed, the scores for the fit between the theoretical model and the data support investigating them as two latent variables, not as a single one. However, it is clear that there is a need for more work to validate the instrument both empirically and theoretically; this would be an appropriate topic for future research.

## **Limitations**

The participants in this study answered a survey voluntarily. Hence, it could be claimed that the teachers who responded were the ones with the strongest interest in the topic. If this were so, it would reduce the level of external validity (Shadish et al., 2002). However, we have no actual indication of a selection bias, even though we must acknowledge that it might exist. On the other hand, a relatively large number of respondents ( $N = 508$ ), along with a fairly high response rate suggest that this problem might be negligible.

Tsagari and Cheng (2016) call for more experimental designs to be used in research on the washback effect in order to deepen our understanding of how washback works. However, given the lack of previous research in the Norwegian context, where the characteristics of both writing instruction and the final exams differ qualitatively from those underpinning most previous research on washback, we found it pertinent to begin by exploring the extent to which teachers perceive a washback effect on two different but interrelated aspects of teaching: Instruction and assessment. The close connection between these two latent variables made operationalization a demanding task. In fact, different item wordings might well have yielded better support for both the theoretical and empirical dimensions of our study. This could be seen as a threat to construct validity (Shadish et al., 2002). Nevertheless, the moderate-to-good fit of our model provides a useful starting point for further research to identify improved ways of measuring the washback effect on other aspects of teaching, such as teacher–student relationships, teacher self-efficacy, perceived teacher autonomy, and motivational dimensions.

We assume that the associations we have found between the constructs are based on causal processes, but our research design is not capable of providing any insights as regards causality. In our view, this calls for further research. We also acknowledge the complexity of washback and are fully aware that our model explores only a few aspects of it. A further limitation of this study is that it measures *perceived* washback effects, which might well differ from actual effects on how teaching plays out in the classroom. More research is needed to investigate the relationship between the dimensions reported in our survey and actual classroom practice.

Furthermore, we might raise a critical question as to whether our first two variables primarily measure perceived washback from the exams on instruction and assessment or rather the alignment between the exams, instruction, and assessment. Specifically, this issue applies mainly to one item in our first variable (SU\_12: “Teaching to the exam aligns well with the primary goals of the NLA subject”). It is indeed possible to argue that this item focuses on the alignment between the exams, the instruction, and the curriculum. However, the other five items in the first two variables focus more clearly on what the teachers believe about how the final exams influence their choices during instruction and in assessment situations that they are in charge of. Even so, if the items developed for the present study are to be used in another setting similar to that of NLA instruction and exams, consideration should be given to how the first item could be adapted to measure washback in a more unambiguous way. Having said that, however, it must be stressed that, in a context where the exam construct has been rather stable for decades, the task of disentangling washback and alignment remains a complex one.

## **Conclusion**

In the present study, we found that teachers reported a high level of washback on both writing instruction and assessment, and that they had a neutral or slightly negative

attitude toward the ability of the final written NLA exams to measure students' proficiency in NLA. Further, we found that attitudes toward the exams were strongly connected to perceived washback on instruction but more weakly connected to perceived washback on assessment. Taken together, our results suggest that it may be appropriate to use a cautious approach when changing the format of exams in Norway. This is because the associations found between the perceived washback from the present exams on instruction and that on assessment suggest that an alternative exam design will have a strong impact on how students are assessed by their teachers—regardless of the degree to which teachers believe that exams based on that design provide a valid and reliable assessment of NLA competence.

### Authors' note

Harald Eriksen and Aslaug Fodstad Gourvenec have contributed equally to this work.

### Author biographies

**Harald Eriksen** is Associate Professor in the Faculty of Education and International Studies at Oslo Metropolitan University. He publishes research in the field of assessment in education and specializes in the assessment of writing in language arts and feedback.

**Aslaug Fodstad Gourvenec** is Associate Professor in Literacy at the Norwegian Centre for Reading Education and Research at the University of Stavanger. She publishes research in the broad L1/Language arts research field and specializes within literature education, oracy, assessment, and teachers' and students' beliefs about the L1 subject.

**Vibeke Rønneberg** is Associate Professor in Literacy at the Norwegian Centre for Reading Education and Research at the University of Stavanger. She publishes research in the field of writing.

**Marte Blikstad-Balas** is a professor in the Department of Teacher Education and School Research, University of Oslo, and Professor II in the Department of Teacher Education, Norwegian University of Science and Technology. She is also the Vice Director of the Nordic Centre of Excellence on Quality in Nordic Teaching, QUINT. She publishes research on how teachers teach in the L1 subject, especially writing and reading.

### References

- Ahmed, R., & Rahman, K. A. (2019). Review of empirical studies of washback in language testing. *British Medical Journal*, 3(1), 150–162.
- Anagnostopoulos, D. (2005). Testing, tests, and classroom texts. *Journal of Curriculum Studies*, 37(1), 35–63. <https://doi.org/10.1080/0022027042000229350>

## Associations Between Teachers' Attitudes Toward the Final Exams

- Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *The English Journal*, 100(6), 14–27.
- Björnsson, J. K., & Skar, G. B. (2021). *Sensorreliabilitet på skriftlig eksamen i videregående opplæring* [Rater reliability on written exams in upper-secondary school]. University of Oslo. <http://urn.nb.no/URN:NBN:no-88263>
- Blikstad-Balas, M., Roe, A., & Klette, K. (2018). Opportunities to write: An exploration of student writing during language arts lessons in Norwegian lower secondary classrooms. *Written Communication*, 35(2), 119–154. <https://doi.org/10.1177/0741088317751123>
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing. Research contexts and methods* (pp. 3–17). Routledge.
- Collins, J. B., & Miller, N. H. (2018). The TOEFL (ITP): A survey of teacher perceptions. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 22(2), 1–13.
- Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press.
- Gourvenec, A. F., Blikstad-Balas, M., Rønneberg, V., & Eriksen, H. (2023). Lærerstemmer om eksamens påvirkning på norskundervisningen [Teachers' voices on the impact of exams on Norwegian language instruction.]. *Norsklereren. Tidsskrift for Språk, Litteratur og Didaktikk*, 1, 22–27.
- Gourvenec, A. F., Jølle, L., & Valand, S. S. (2021). Tilbakemelding fra Landslaget for Norskundervisning (LNU) på forslag til ny eksamen i norsk [Response from the Association for Norwegian Language Arts Teachers (LNU) on the proposed new exam in Norwegian Language Arts]. *Norsklereren. Tidsskrift for Språk, Litteratur og Didaktikk*, 2, 38–41.
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39–51. <https://doi.org/10.6018/ijes.13.2.185891>
- Hertzberg, F., & Roe, A. (2016). Writing in the content areas: A Norwegian case study. *Reading and Writing*, 29(3), 555–576. <https://doi.org/10.1007/s11145-015-9607-7>
- Hovdhaugen, E., Proitz, T. S., & Seland, I. (2018). Eksamens- og standpunkt karakterer – to sider av samme sak? [Exam grades and report-card grades—Two sides of the same coin?]. *Acta Didactica Norge*, 12(4), Article 17. <https://doi.org/10.5617/adno.6276>
- Jambak, T. (2021). *Utdanningsdirektoratets endringer i eksamen moter sterke reaksjoner* [The revisions to exams made by the Norwegian Directorate for Education and Training meet with strong reactions]. *Utdanningsnytt*. <https://www.utdanningsnytt.no/eksamen-thom-jambak-utdanningsdirektoratet/utdanningsdirektoratets-endringer-i-eksamen-moter-sterke-reaksjoner/305069>
- Kline, R. (2016). *Principles and practice of structural equation modeling*. The Guilford Press.
- Kuang, Q. (2020). A review of the washback of English language tests on classroom teaching. *English Language Teaching*, 13(9), 10–17. <https://doi.org/10.5539/elt.v13n9p10>
- Larsson, M., & Olin-Scheller, C. (2020). Adaptation and resistance: Washback effects of the national test on upper secondary Swedish teaching. *The Curriculum Journal*, 31(4), 687–703. <https://doi.org/10.1002/curj.31>
- Ministry of Education and Research (Norway). (2016). *Fag—Fordypning—Forståelse. En fornyelse av Kunnskapsløftet* [Subject—Specialization—Comprehension. A renewal of the Knowledge Promotion Reform]. Report to the Storting (Meld. St. 28). Kunnskapsdepartementet. <https://www.regjeringen.no/contentassets/e8e1f41732ca4a64b003fca213ae663b/no/pdfs/stm201520160028000dddpdfs.pdf>
- Ongstad, S. (2002). Positioning early research on writing in Norway. *Written Communication*, 19(3), 345–381. <https://doi.org/10.1177/074108802237749>
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.
- Shelton, S. A., & Brooks, T. (2019). “We need to get these scores up”: A narrative examination of the challenges of teaching literature in the age of standardized testing. *Journal of Language & Literacy Education*, 15(2), 1–17.
- Smidt, J. (2007). Positioning, responsibility, and relevance in MTE research. In W. Herrlitz, S. Ongstad, & P.-H. van de Ven (Eds.), *Research on mother tongue education in a comparative international perspective. Theoretical and methodological issues* (pp. 211–226). Rodopi.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5–29. <https://doi.org/10.1191/1362168805lr152oa>
- Statistics Norway. (2022). *Upper secondary education*. <https://www.ssb.no/en/utdanning/videregaende-utdanning/statistikk/videregaende-opplaering-og-annen-videregaende-utdanning>
- Thu, D. M. (2020). Washback to language teachers: A review of models and empirical research in and beyond Vietnam. *VNU Journal of Foreign Studies*, 36(4). <https://doi.org/10.25073/2525-2445/vnufs.4582>



- Tsagari, D., & Cheng, L. (2016). Washback, impact, and consequences revisited. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 1–14). Springer International Publishing. [https://doi.org/10.1007/978-3-319-02326-7\\_24-1](https://doi.org/10.1007/978-3-319-02326-7_24-1)
- Valand, S. S. (2022). *Forslaget om flere oppgaver til vg3-eksamen i norsk bør revurderes* [The proposition of more tasks in the Grade 13 Norwegian language arts written exam should be reconsidered]. Utdanningsnytt. <https://www.utdanningsnytt.no/eksamen-landslaget-for-norskundervisning-norsk/forslaget-om-flere-oppgaver-til-vg3-eksamen-i-norsk-bor-revurderes/317001>
- Wu, M.-J., Zhao, K., & Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports*, 7, Article 100206. <https://doi.org/10.1016/j.chbr.2022.100206>