

Building a Corpus of Written Tasks of Swedish National Tests in English: Motivation, Method and Research Applications

Rachel Allan, Irina Shaw, and Martin Shaw (Mid-Sweden University)

Abstract

This article describes a collaborative project involving the construction of a corpus of graded year 9 National Tests in written English. National Tests are standardized high stakes tests which are an important part of the Swedish education system because the results provide an indication of performance at national level, and also feed into pupils' overall assessment. The grading of National Tests in written English has been found to be problematic for teachers, and a need for assessment training identified (Erickson and Tholin 2022). By providing a searchable database of graded written texts, together with the teacher feedback, this project aims to create a resource to support pre- and in-service teachers in interpreting knowledge requirements and assessment guidelines, and providing effective feedback. The corpus will also provide a resource for research into the features of student writing at different grade levels.

To create the corpus, past papers from collaborating schools have been anonymized, digitized and coded. As a result, pupils' texts can be easily sorted by a range of criteria, for example, year, gender, education type, grade achieved on the written paper and overall grade for the National Test. Teacher feedback can be accessed similarly. We outline potential research areas provided by this resource, and demonstrate how some of these might be explored. We also give examples of how the developing corpus has already been used as a resource for English teacher training programmes, and outline future plans for the project.

Keywords: Swedish National Test in English; writing; corpus; assessment; grading; teacher training

1. Introduction

This article reports on the motivation for and construction of a corpus of graded year 9 National Tests in written English. English is a core subject in Swedish schools, with elementary and secondary school pupils required to take the National Test in English (NTE) in years six and nine (ages 13

Allan, Rachel, Irina Shaw, and Martin Shaw. 2023. 'Building a Corpus of Written Tasks of Swedish National Tests in English: Motivation, Method and Research Applications.' *Nordic Journal of English Studies* 22(2): 128–154.

and 16). These are high stakes, standardized tests of productive and receptive skills in written and spoken English, and an important component of the overall assessment of the pupil. National tests cease to be confidential three to six years after the test,¹ and at this point past papers from those years are released, presenting an opportunity to use these for research and training purposes. Here, we describe a new collaborative project between researchers at a Swedish university and schools in the region, using these National Tests as a research and training resource. In section 2, we outline the status of the NTE in Sweden and existing research in the field, before looking more broadly at research into writing assessment (section 3) and the need to develop teacher competence in this area (section 4). In section 5, we discuss the aims and potential uses of the corpus, and then in section 6 we outline its construction and the coding system used. We go on to describe some preliminary research into extracts from the database in section 7, and in section 8, the use of the database within teaching, teacher education and student research is outlined, along with plans for collaborative workshops with local teachers. Finally, we summarize what has been achieved so far, and our ambitions regarding the future of the project.

2. The Swedish National Test in English

The NTE is an important benchmark for English education in Sweden. As Erickson (2020: 1) points out, national tests have several functions, with one important aim being to encourage comparability and equity within the school system, while offering a means of evaluating performance across it. The tests aim to both encourage learning and be an indicator of learning (Erickson 2020), but they do not function as traditional exams. They give one perspective on pupils' learning, which is combined with classroom-based assessments to measure pupils' overall achievements. The tests have traditionally been paper-based, but are moving towards digital delivery, with full digitization planned for 2025 (year 6) and 2026 (year 9).² At present, national tests are not centrally assessed but graded locally in the

¹ <https://www.skolverket.se/undervisning/grundskolan/nationella-prov-i-grundskolan/forbereda-och-bestalla-prov-i-grundskolan/sekretess-pa-prov-i-grundskolan>

² <https://www.skolverket.se/om-oss/var-verksamhet/skolverkets-prioriterade-omraden/digitalisering/digitala-nationella-prov/digitalisering-av-de-nationella-proven---overgripande-information#Tidplan>.

schools where they are taken, although there are plans for external assessment and central marking to be introduced in the course of the digitization process.³ In the current system, teachers work in teams to apply the knowledge requirements provided, using detailed assessment guidance made available to them by the Swedish National Agency for Education, produced by the NAFS (National tests in foreign languages) project at the University of Gothenburg (2022).

NAFS produces the NTE along with other assessment materials in English, French, German and Spanish as foreign languages. According to the website (<https://www.gu.se/en/national-tests-of-foreign-languages>), the tests are developed in line with research-based principles, drawing on input from large groups of teachers, pupils and researchers. The material follows action-based criteria, allowing pupils to show what they know and can do, with assessment focusing on strengths rather than weaknesses, evaluating pupils' willingness and ability to understand and convey meaning in a comprehensible way. The assessment guidelines distinguish between errors that [might] disturb communication and errors that actually do disrupt communication, i.e., between errors representing different degrees of gravity. The criteria are influenced by the Common European Framework of Reference (CEFR) (Council of Europe 2001, 2020) but have not been empirically aligned to it (see Erickson 2019 for a discussion of this⁴).

³ Ordinance SFS 2021:1061 (November 2021) outlines the initial experimentation (January 18–December 23) into changes in the NTE, including central correction and external assessment, so that NTEs will not be assessed by the student's teacher. As piloting is still in progress, it is not known when this will be implemented on a broader scale.

⁴ The tasks/content level of the written content of year 9 national tests in English (such as those examined in the present study) were considered to fall between B1 and B2 of the CEFR in Erickson's (2019) holistic peer analyses. The benchmarked sample answers were considered to form the following relationships: Pass: A2+/Low B1 (ranging from A1+ to B1+); Pass with distinction: B1+ (with a range from B1 to B2); Pass with special distinction: B2 (ranging from B1 to C1). Several informants commented on the discrepancy between students' fluency and confidence, and their linguistic accuracy (Erickson 2019: 57).

My Media

The next edition of *Teen Technology* will focus on young people's thoughts about media and how they use it for communication, information or simply to relax. The magazine would like some input from teenagers living in different countries and different cultures. Your school has been invited to send in texts to the magazine.

Write a text called **My Media** including these four parts:

- **Describe** what different types of media you use – TV, radio, books, newspapers, magazines, computers, the Internet, social media. What media do you like the best? What do you use it/them for? When? How much time?
- **Explain and compare** how the use of media has changed over the last 10 years for you and for other people.
- **Discuss** how the use of media affects your life and other people's lives. What is positive and negative about it? The illustration below might give you some ideas.
- **Describe** how you think the way we use different media will change in the future?

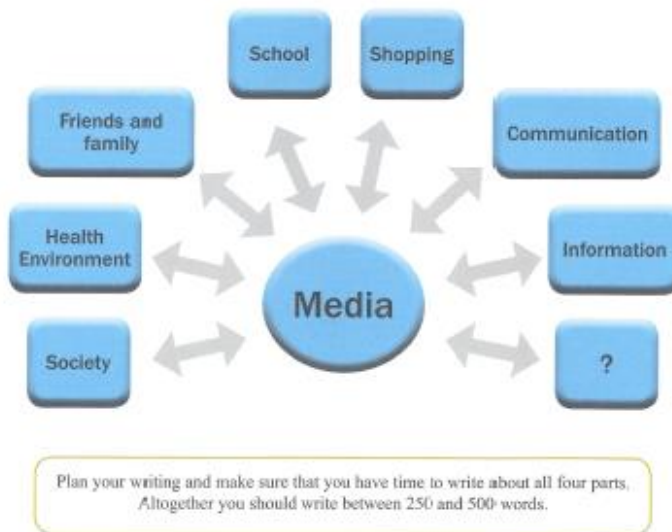


Figure 1: Writing task for English 9 NTE (2015/16)

The NTE includes both receptive and productive skills, with reading and listening tests as well as spoken and written tasks. In this article, the discussion will focus on the written task as this is the main area of interest in the present phase of the project, although we hope to extend this to other

test components in the future. Pupils are asked to write a response to a single task, which requires them to produce a text for a particular context up to a given word count. For the year 9 tests, the focus of our corpus, pupils are asked to write between 250 and 500 words in a given genre, with support provided both through specific questions they should respond to, and a diagram with ideas, as shown in Figure 1. The questions encourage different linguistic strategies, e.g., *describe*, *explain*, *compare*, *discuss*, and the content guidance prompts the use of past, present and future tenses/constructions. Aside from this task, pupils are also given the opportunity to comment on the task at the end of the test, which offers an insight into their abilities to express their thoughts in writing, as well as providing feedback on their experience of the test.

<p>E</p> <p>In written presentations of various kinds, the student formulates himself simply, comprehensibly and relatively coherently.</p> <p>In interaction in different contexts, the student expresses himself easily and comprehensibly and to some extent adapted to the purpose, recipient and situation. In addition, the student uses strategies that to some extent facilitate and improve interaction.**</p>	<p>C</p> <p>In written presentations of various kinds, the student learns with some variation, relatively clearly and relatively coherent. The student formulates himself with some fluency and to some extent suited to purpose, recipient and situation.</p> <p>In interaction in different contexts, the student expresses himself relatively clearly and with some fluency and to some extent adapted to the purpose, recipient and situation. In addition, the student uses strategies that facilitate and improve interaction.**</p>	<p>A</p> <p>In written presentations of various kinds, the student formulates himself with some variety, clearly and coherently. The student also formulates himself with fluency and some adaptation to the purpose, recipient and situation.</p> <p>In interaction in different contexts, the student expresses himself clearly and with fluency as well as with some adaptation to the purpose, recipient and situation. In addition, the student uses strategies that facilitate and improve the interaction and move it forward in a constructive way.**</p>
--	--	--

** Not specifically focused, but the subtest provides an opportunity for assessment of this ability.

SUBTEST C, GRADE 9, ENGLISH

Figure 2: Assessment guidelines for written production and interaction for English 9 NTE

To assist with the grading of the NTE written task, NAFS supplies detailed assessment guidelines for each level (see Figure 2 for year 9, for example), focusing on communicative ability. NAFS also supplies a range of benchmarked tests prepared in collaboration with experienced teachers for further support. Teachers are encouraged to work collaboratively to rate the tests to encourage inter-rater consistency (for detailed examples of assessment discussions on NTE speaking tests, see Sandlund and Sundkvist 2019, 2021). However, this has been an area of weakness, as studies by the Swedish Schools Inspectorate have indicated that inter-rater agreement for the NTE has fallen (Erickson 2020: 5), raising concerns

about reliability. While reliability may be improved by the eventual centralization of assessment, in the meantime there is a need for training in interpreting and applying criteria to improve the test's reliability for both pre-service and in-service teachers. Even once central marking of the national tests is underway, test washback (i.e., the positive or negative effects of a test on teaching and learning) implies that teachers will still need to have a full understanding of the grading system in order to teach the syllabus effectively.⁵

3. Research into the NTE

To date, most published research on the NTE (and other foreign languages) has originated from NAFS (2021), with an emphasis on the purpose and design of the tests. Erickson et al. (2022) describe the collaborative nature of national tests, detailing the contributions made by the main stakeholders, the test-takers and teachers. Regarding the test-takers' feedback, this is considered essential, as 'students are the true experts, and [...] test developers need their assistance to optimize the quality of the materials' (Erickson et al. 2022: 8). Such feedback includes targeted questionnaires at the pre-testing stage (e.g., Olsson et al. 2018), and also the comment box on the test booklet that test-takers and teachers are invited to fill in after taking the test. Student comments feed into the construction of tests, giving insights into the effectiveness of instructions, texts and tasks, helping to avoid ambiguity, obscurity and bias (Erickson et al. 2022: 9). Similarly, teachers make an important contribution, taking an active part in many aspects of test development, giving input on tasks, and by taking part in rating and benchmarking activities in addition to administering the tests and responding to questionnaires (Erickson et al. 2022: 14; see also Sandlund and Sundqvist 2019, 2021).

Like test-takers, teachers are reported to be generally positive towards the NTE, although critical of system-related aspects, such as workload and the regulations regarding the relationship between national test results and final grades (Erickson et al. 2022: 9). Assessment standards have also been flagged, particularly for the year 6 test. In their study of teachers' perceptions of the year 6 writing component of the NTE, Erickson and Tholin (2022) found that teachers were critical of the level required,

⁵ <https://www.skolverket.se/om-oss/press/pressmeddelanden/pressmeddelanden/2022-02-15-sa-kan-central-rattning-fungera-i-praktiken>.

finding the assessment of the tests too lenient. This may account for the conflict between NTE grades and classroom assessment in year 6. Although NTE results should ‘be taken into special consideration’ (Skolverket 2018) when awarding grades, there has been a consistent pattern of lower final grades being awarded than those indicated by the national test results (Erickson and Tholin 2022: 4–5). The study reported uncertainty regarding grading expressed in the teachers’ comments, with requests for more explicit grading criteria and clearer guidelines to avoid subjective and arbitrary assessments (Erickson and Tholin 2022: 13–14). Their study highlights the importance of continued research in direct collaboration with active teachers to increase mutual understanding of assessment and grading processes, with the aim of improving reliability.

Other NAFS research into NTE results at upper secondary level has uncovered gender differences in English proficiency (Börjesson and Nilsson 2018), finding that male pupils tend to perform better in receptive skills. Previously, this was offset by better productive skills in female pupils, but this has levelled out over recent years, leading to higher results overall for male pupils. One explanation put forward for this is that male pupils tend to have a higher vocabulary level than female pupils, perhaps due to more reading or engagement with English through extracurricular activities (see Sundqvist 2009; Sundqvist and Sylvén 2012). The focus of this study was on upper secondary school pupils, and the authors note that this trend has not been apparent in lower secondary, i.e., year 9 NTE results (Börjesson and Nilsson 2018: 1). Results published for the 2021–2022 academic year show a fairly even performance between genders in the year 9 NTE sub-tests, with the exception of the writing sub-test, where girls outperform boys by almost 2% (Skolverket 2022). However, it is important to keep any potential gender difference under review at all stages in national test performance.

A further study, Olsson’s (2018) investigation into NTE writing tasks, shows a close relationship between year 9 and upper secondary results. This study found that in a sample of 71 writing tasks in the NTE across four different levels (lower secondary years 6 and 9, and upper secondary English 5 and 6), there were no significant differences in text length, word variation, word length and use of more unusual vocabulary between texts written in year 9 and in the two high school courses, whereas there were significant differences in these areas for texts written in year 6. This

suggests that year 9 writing tasks may be able to give us useful insights into potential performance at higher levels.

4. Assessment literacy in teachers

Erickson and Tholin (2022) suggest that a lack of assessment training in teachers at year 6 level may account, at least in part, for their difficulties in interpreting the knowledge requirements. Rater insecurity is addressed in Sandlund and Sundqvist's (2019, 2021) investigations into assessment of the speaking part of the NTE, and they find that inexperience and insecurity leads teachers to rate more severely, severity being seen as 'more professional' than leniency (Sandlund and Sundqvist 2021: 158). A review of the literature suggests that the situation in Sweden is far from unusual. Studies published internationally indicate that language assessment literacy (LAL), defined as 'the knowledge skills and principles that stakeholders involved in assessment activities are required to master in order to perform assessment tasks' (Inbar-Lourie 2017: 257), has been neglected in teacher training, despite an increasing emphasis on language testing and assessment in recent years (see Fulcher 2012 for an overview). For example, Tsagari and Vogt (2014) report on a survey of 853 language teachers, including primary, secondary and tertiary levels from seven European countries, concluding that LAL is poorly developed, with insufficient training both at pre-service and in-service level, with most teachers learning on the job. A further study, Roslan et al. (2022), revealed a similar picture among teachers in Japan and Brunei. As a counterpoint, data from a small qualitative study by Yan et al. (2018) suggests that learning on the job is not necessarily negative; involvement in local assessment practices can encourage teachers to develop relevant knowledge and skills, which can be consolidated through repeated practice and self-reflection. However, as Tsagari and Vogt (2014: 391–392) point out, a lack of training leaves teachers insufficiently prepared and may result in perpetuating inappropriate assessment methods.

Looking specifically at second language writing assessment, Crusan et al.'s (2016) survey, involving 702 university-level teachers from 41 countries on five continents, showed a more positive, although sometimes contradictory picture of LAL. Teachers reported feeling inadequately prepared to assess writing, particularly in the use/creation of grading criteria, yet nearly half of those surveyed believed that these criteria are an important tool in helping students understand why they receive a specific

grade (Crusan et al. 2016: 53). Similarly, 66% of the informants believed that the scoring of writing is always inaccurate, yet when asked whether scoring was a subjective process, almost 60% of them disagreed (Crusan et al. 2016: 49–50). Inter-rater reliability was considered difficult to achieve by 55% of the instructors, yet 80% believed that rater training is not helpful (Crusan et al. 2016: 50). These findings point to a higher level of LAL among teachers in some respects, but suggests there is still a need for reinforcement through in-service support. This is highlighted by their finding of an inverse relationship between teaching experience and knowledge of assessment, suggesting that ‘expertise is not guaranteed by the number of years of teaching experience’ (Crusan et al. 2016: 53).

Training is particularly important because the research literature shows that teachers are more likely to use assessment procedures they have been trained in (Tzagari and Vogt 2014: 376). For example, Reynolds-Keefer (2010) found that pre-service teachers who used grading criteria as students may be more likely to use them in their own teaching. There has been a move to address training needs in language assessment, with initiatives such as the Teachers’ Assessment Literacy Enhancement (TALE) Project. Following a major needs analysis among English language teachers and learners in four EU countries, the TALE project produced an online training programme and practical handbook in response (all freely available at www.taleproject.eu), providing an accessible resource to promote a broad-based understanding of assessment principles and procedures. However, as the research suggests (e.g., Yan et al. 2018: 167), there is also a need for assessment training for language teachers tailored to their context, which takes current assessment practices into consideration. It has been proposed that collaborative forms of in-service teacher training, with teachers and experts such as teacher trainers and assessment specialists working together, may be a useful way for identifying and responding to these needs in context (Tzagari and Vogt 2014: 392; Sandlund and Sundqvist 2021). Thus, through the NTE project, we aim to respond to these recommendations for contextually-situated and collaborative training, as discussed below.

5. The NTE corpus: Aims and purpose

The aim of the project is to construct a digital database, initially of year 9 writing tasks from the NTE (up to 2016) gathered from municipal and

independent lower secondary schools in the municipality.⁶ In this area, the policy is to preserve all national tests after the reuse time has expired (Gallringsråd för kommuners, landstings och regioners utbildningsväsende 2011: 10), providing a potentially valuable resource for analysis. Our project is a collaboration with schools who consent to release these papers for inclusion in the database, following the usual ethical considerations regarding confidentiality and anonymity and in accordance with the the European code of conduct for research integrity (ALLEA 2023). After the anonymised papers have been collected from the schools, they are processed to construct a coded digital database that can be easily searched to show patterns in pupil achievement and assessment. Through this project, we plan to use the corpus in a variety of ways to promote professional development among practicing English teachers and student teachers. The corpus will also offer opportunities for more general research into pupil performance, both in the context of the NTE and at a broader level. Both of these aspects are discussed in detail below.

As indicated in sections 3 and 4, teacher assessment training and practices in the Swedish context have received relatively little attention by researchers. However, Erickson and Tholin's (2022) study and research into assessment literacy in general points to the need for teachers to develop their competence in assessment and grading practices. This holds true regardless of the future organization of the NTE, whether teachers are assigned a role in grading for partner schools or not; a clear understanding of the assessment procedure is required to teach their own pupils effectively. The NTE corpus will be an invaluable tool for trainers in teacher education programmes, giving student teachers an opportunity to operationalize the theoretical principles taught on the course. They will be able to view multiple responses to different task types, at different levels, rather than just one or two sample answers, allowing them access to a wide range of authentic text features and how these are viewed by practicing teachers. In view of the collaborative nature of the project, the resource will also be used to bring together educators from university and schools to explore issues around assessment and grading. A series of practical workshops are planned to introduce and familiarize teachers with the corpus and demonstrate its use and potential applications. This will also

⁶ We aim to extend to other municipalities as the project develops.

present an opportunity to gain insights into inter-rater discussions (see Sandlund and Sundqvist 2021). A further purpose of the project is to offer the university's student teachers the opportunity to use the database as material for their independent work in subject didactics with English in focus.

In addition to providing data for student research, the NTE corpus offers many potential areas for higher level research. The previous studies (outlined in section 3) demonstrate the benefits of analysing previous test performance in order to assess and respond to gaps that may develop between genders or test years. Aside from Olsson's (2018) study, which analysed a limited number of papers, very little specific research has been carried out into year 9 writing tasks. With regard to the assessment of student performance, the relationship between various linguistic features of task performance and grade awarded can be explored, supplementing and extending existing research such as Olsson (2018). Male and female achievement can also be assessed in the context of productive skills, complementing Börjesson and Nilsson's (2018) study. Areas where no large-scale research currently exists, such as the effect of task and task instructions on pupil achievement in the writing component of the NTE, can also be considered. The other main area the corpus will offer insights into is teacher feedback. It can be used to inform us about which aspects of the task response (e.g., content, range, accuracy) generate comments; which aspects of the knowledge requirements generate feedback and which do not; what kind of comments are given, and how they are phrased. These areas constitute fields of research in which there is little concrete, data-based research available.

The corpus will also offer the opportunity to examine learner language at a broader level. As has been noted, the proficiency levels in the current Swedish national foreign language syllabi have not been empirically adapted to the six levels of CEFR (Erickson 2019; Erickson and Pakula 2017), although there are associations between them. Through an in-depth, large-scale analysis of year 9 written production allowed by the corpus, such associations can be made clearer. Concrete examples of language production from the corpus can be mapped against knowledge requirements. A study of this kind has been carried out at an international level with the EnglishProfile project (Saville and Hawkey 2010), which used the Cambridge Learner Corpus (a corpus of Cambridge English test

papers at different levels), resulting in a website⁷ which illustrates which aspects of English are typically learned at each level of the CEFR (Council of Europe 2001). EnglishProfile does this with both lexical and grammatical features; a lexical example is that *tell* is typically produced at A1 level, whereas *tell somebody off* is most likely to be produced at B2 level. A corpus of NTE written texts will allow us to similarly concretize the lexical and grammatical profile of a year 9 learner, both in terms of the NTE knowledge requirements and the CEFR.

6. Construction of the corpus

The corpus is envisaged as a monitor corpus, with additional material to be added from further years and regions as it becomes available. The initial version of the corpus was constructed in collaboration with schools in the region. The project team liaised with schools to explain the purpose of the proposed project and invite them to join. The mutually beneficial nature of the project was highlighted; the schools willing to contribute to it through the provision of materials would benefit from their involvement, not only from the resulting corpus but from a programme of associated workshops. Three schools, one private and two public, with approximately 900 to 1150 students in years 1–9, initially agreed to take part in this collaboration. The participating schools agreed to provide the anonymized test papers from their year 9 classes within the period indicated (i.e., up to 2016).

The material included in the corpus consists of the pupils' written response to the writing task (see the task example in Figure 1), any teacher comments written on the response, and the pupil and/or teachers' comment in the comment box (see Figure 3). Although texts were fully anonymized, i.e., all identifying features were deleted from the task response included in the body of the corpus, we coded each file with information that would enable division into subcorpora, as discussed in 6.1. Contributions from individual schools were recorded as shown in Table 1. The material was then scanned into digital text files. As Table 1 shows, in most cases, the written response to the task was typewritten, making it straightforward to scan into digital format and convert into an editable text file using software inbuilt in most modern computers. A limited number of texts had teacher markings (in pencil) included, which mainly highlighted spelling errors or

⁷ <https://www.englishprofile.org/>.

punctuation errors. These were scanned twice, first in their original format and second with the markings erased for conversion to text files. Handwritten texts were mainly limited to the short texts in the comment box, such as the one shown in Figure 3, and these were manually transcribed.

Table 1: Sample of papers collected from one school in the region

Class	Pupils (n)	Male (n)	Female (n)	Exams (n)	Typed	Comments
9A	30	17	13	30	30	8 with teacher pencil markings (limited)
9B	31	15	16	31	31	10 with teacher pencil markings (limited)
9C	31	11	20	31	31	4 with teacher pencil markings (limited)
9D	32	16	16	32	32	6 with teacher pencil markings (limited)
Total	124	59	65	124	124	

To be filled in by the teacher		To be filled in by the teacher and the student after the test													
Delprov A - Focus: Speaking <table border="1"> <tr><td>F</td><td>E</td><td>D</td><td>C</td><td>B</td><td>A</td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td>X</td></tr> </table>		F	E	D	C	B	A						X	Comments: All three parts went well. Both listening and reading was very long, especially listening, it became hard to focus on the more complicated speakings (in the end), because the "tasks" in the beginning was so slow. The writing part was fine, but I think I wrote a bit too much.	
F	E	D	C	B	A										
					X										
Delprov B - Reception B 1 - Focus: Reading B1a (16) 15 pts B1b (19) 19 pts B1c (22) 19 pts B 2 - Focus: Listening B2a (10) 9 pts B2b (15) 15 pts B2c (23) 23 pts Total (105) 100 pts															
Delprov C - Focus: Writing <table border="1"> <tr><td>F</td><td>E</td><td>D</td><td>C</td><td>B</td><td>A</td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td>X</td></tr> </table>		F	E	D	C	B	A						X		
F	E	D	C	B	A										
					X										
TEST GRADE <table border="1"> <tr><td>F</td><td>E</td><td>D</td><td>C</td><td>B</td><td>A</td></tr> <tr><td></td><td></td><td></td><td></td><td></td><td>X</td></tr> </table>		F	E	D	C	B	A						X		
F	E	D	C	B	A										
					X										

Figure 3: Grading and comment box

6.1. Corpus annotation

In order to be able to easily extract individual components from the database, a coding system was applied to the text files. As further data will be added to the corpus, it was important to adopt a system that would allow for further expansion. The first code identifies the year; in this initial phase the NTE corpus focuses on year 9, indicated by Y9. This will enable us to easily extend the coding to incorporate other years, e.g., Y6. The second code indicates the year the test was written in, i.e., 2016. The third and fourth codes provide information about the student, i.e., their individual number in the class, and their gender, limited here to female and male in accordance with the NTE categories. Next, information about grades is indicated, with G representing the grade awarded for the writing task, i.e., from A–F, and FG shows the final grade awarded for the NTE. S indicates the school number, with each school given a unique number, and then the education type is recorded, i.e., private or public school, represented by 1 and 2 respectively. Then, the class number is listed. Finally, the teacher number is given, and their gender is indicated. A summary of the coding system is shown in Table 2.

Table 2: Coding system used for the NTE database

Code	<Y9><D2016><SN1><SGF><GA><FGA><S1><ET1><C9A><T1><TM>
Key	Y = year in school (e.g., 9) D = date (e.g., 2016) SN = student number (e.g., 1) SG = student gender (e.g., F = female) G = grade on writing task (e.g., A) FG = final grade on NTE (e.g., A) S = school number (e.g., 1) ET = education type (e.g., 1) C = class (e.g. 9A) TN = teacher number (e.g., 1) TG = teacher gender (e.g., M = male)

In the individual files, we also indicated the different components of the text, i.e., ‘<partC>’ for the student text, ‘<grade>’ for the grading page (including grades on speaking and reading, illustrated in Figure 3), ‘<reflection>’ for the students’ self-reflection, also shown in Figure 3, and ‘<feedback>’ for comments from the teacher. If there were student notes, this was also indicated together with a brief description, e.g., ‘<notes> [see pdf – draft answer]’. Some students had written plans, others had drawn diagrams and doodles, making them difficult to transcribe.

One of the limitations faced in the construction of this corpus relates to information regarding social variables among the test-takers and assessors. Gender is the only social variable recorded on the NTE papers, and this is limited to traditional binary male/female roles. While this is useful, allowing for the investigation of its potential effect on test-taker performance following on from previous research (see section 3), it offers only one perspective. Many other factors, such as pupils' first language, ethnic background and socio-economic situation also have a role to play; similarly, for teachers, these factors, plus education level and experience are likely to have an effect. However, as papers were collected retrospectively, and with the anonymity measures in place, it was not possible to obtain information on such variables.

7. Preliminary research

The project is still in its early stages and input for the database is still being gathered and processed, but we have begun to explore its potential for future research. As noted in section 5, one way we plan to use the database is to examine the relationship between the completed tasks and the grades awarded. As a starting point for this research, we took sample data sets from the database and examined these from two different perspectives. First, as outlined in 7.1, we looked at the aspects teachers explicitly noted in their feedback, which is assumed to have informed their grading decision. Second, as described in 7.2, we profiled the lexical content of the responses at different grade levels, to see if there is any evidence of a relationship between vocabulary use and grade.

7.1 Teacher feedback

Teacher feedback on a sample of 52 papers, from two classes from the same educational institution, was examined. Each writing task is given brief feedback by the grading teacher. In this sample, there were three teachers grading. Feedback in this sample is a short comment, typically structured with an evaluative remark on what was achieved, generally relating to content, and a focus on an area for improvement, e.g., *Great content and details – focus on your spelling*. The feedback was compiled in a text file and the AntConc application (Anthony 2022) used to construct a word list, allowing us to see which areas of performance were most frequently mentioned. As shown in Table 3, the areas for comment

prioritized were *content* and *spelling*, each mentioned in 26 of the 52 comments. There were much fewer mentions on *structure*.

Table 3: Words relating to performance in teacher feedback, number of occurrences in parentheses

Content	Accuracy	Structure
Content (26)	Spelling (26)	Structure (5)
Task (3)	Grammar (19)	Paragraph (3)
Topic (3)	Edit*(7)	
Points (4)	Proofread* (12)	
	Capitalise* (6)	
	Punctuation (3)	

* includes different forms/spellings of the word

Most of the comments including the word content were positive; those on task, topic and points tended to point out insufficiencies, e.g., *Make sure you cover all the points* (Grade D+) and were associated with lower grades (D and below). Comments on accuracy were included at all levels, and were generally framed as an area for improvement, e.g., *Work on improving grammar and spelling* (Grade C-), and this was the case even at higher grades, e.g., *keep working on your spelling* (Grade A). However, there were four positive comments relating to grammar and spelling, all for tasks awarded grade A, e.g., *great job with the grammar and spelling*. Text structure received less attention, with a mix of positive comments, e.g., *nice structure to your writing* and pointers for improvement, e.g., *make sure you structure your writing so you cover all points in detail*. There were a few areas that received little to no attention. There was one comment on style, *Very strong examples and voice when writing*, and no comments relating to the range of grammatical structures or vocabulary used.

Reviewing the assessment guidelines shown in Figure 2, the main areas of focus are *variety, clarity, coherence, fluency, adaptation to purpose, recipient and situation*. There is clearly a relationship between these comments and the assessment guidelines; occasionally this is explicit (*your writing was fluent*), but for the most part it is, understandably, couched in terms pupils are more familiar with. Higher grades received praise for their content, which encompasses each of these aspects of the guidelines. Clarity and coherence were addressed in comments about accuracy, from the frequent comment to *work on spelling*, to improve clarity, to a more serious comment relating to disturbances in

content communication, i.e., *some of the mistakes you made changed the meaning of the sentence*. Adaptation to purpose could be considered to be subsumed in comments on structure. There were no comments on other areas relating to some of the criteria, such as *variety* of grammatical structures and vocabulary used, or *adaptation* in terms of register consistency (i.e., style/formality level). Of course, this does not necessarily imply that they were not considered in the grading, but in the teachers' comments, accuracy is prioritized over range. These comments represent only a very small sample, but they demonstrate one area that is likely to be fruitful for further investigation as the database grows. Issues around the transparency of the knowledge requirements and how broadly they are being interpreted and applied have been highlighted in previous research (e.g., Erickson and Tholin (2022): see section 3), and the NTE database can be used to provide further perspectives on this. As indicated in section 8.3, it is an area of research that students have already shown interest in investigating.

7.2 Use of vocabulary

As noted in section 7.1, there was no explicit mention of range of vocabulary used in the feedback comments. To see if there was any relationship between range and grading, a sample of student responses were profiled to assess whether there was any evidence that some relationship existed. Vocabulary profiling is a way of assessing the level of vocabulary used; high frequency vocabulary is usually learnt first, with research suggesting that knowledge of the most common 2,000–3,000 words is sufficient for basic communication in English,⁸ with lower frequency vocabulary usually suggesting a more advanced level. Although this is rather a crude measure, it can give an initial insight into the vocabulary levels involved here, and pave the way for more sophisticated analysis.

To profile the vocabulary, first, the texts were separated by grade. This was a high-scoring sample, and there were many more papers at grades A and B than at grade C and below.⁹ At grades E and F, there was only one paper at each level, while at grade D, there were eight papers, with nine

⁸ See Schmitt and Schmitt 2014 for a full review.

⁹ Because of the small sample size, +/- (only awarded at C, D, E and F grades) were not taken into consideration.

available at grade C and 18 and 25 at grades B and A respectively. For the preliminary analysis, we decided to look only at eight randomly selected papers from grades A to D, and disregard grades E and F due to the limited data available. The data analysed is outlined in Table 4. As Table 4 shows, the higher graded papers tended to be longer, with an average length of 776 words (tokens) for texts graded at A, compared to 572 for texts graded at D. It is interesting to note that the requirement was to write 250–500 words, so all of the students had written longer texts than required, even those that were awarded lower grades.

Following the process used in Walker and Allan (2018: 199–200), we first extracted the words used in the task instructions (Figure 1) from the words used in each grade set of papers, to ensure that we were only including lexis that students retrieved from their own knowledge. This was done using the Text Lex Compare interface (http://www.lex tutor.ca/cgi-bin/tl_compare/index.pl) available in Compleat Lexical Tutor (Cobb 2023). The total number of unique words (types) occurring at each level after the words from the task instructions had been removed is shown in Table 4; this includes off-list words. It demonstrates that a wider range of types was used at Grade A (900), than Grade B (764), while at Grade C and D, a similar number of unique words were used (702 and 703 respectively).

Table 4: Length of texts and range of types at different grade levels

Grade	A	B	C+/-	D+/-
Tasks (n)	8	8	8	8
Total text (tokens)	6204	4687	5217	4582
Text length (mean)	776	586	652	572
Unique words (types)	900	764	702	703

The resulting word lists were then uploaded in turn into the VocabProfiler VP-Compleat (<http://www.lex tutor.ca/vp/comp>), using the BNC-COCA-25 option. This allocates the words to different frequency levels based on their occurrence both in the British National Corpus and the Corpus of Contemporary American English, in a total of 25 lists represented as K1 to K25, K1 representing the most frequent thousand words, K2 the next most frequent thousand words, etc. Any words in the students' word lists that do not occur in the 25 lists are placed in an off-list file: these are words that do not occur, or occur with low frequency in the BNC/COCA corpora, and include proper nouns or adjectives, together with newer words which

were not widely used when the corpora were built, such as *Instagram*, *google* and *Netflix*. It is also helpful in filtering out misspelt words, although it should be noted that some misspelt words are incorrectly attributed to other word lists, for example, *trough* occurs in the K7 word list, when it is actually a misspelling of *through*, which would be in the K1 word list. After completing this process, we found that approximately 99% of the word types (i.e., unique words) in these lists were on the K1 to K9 word lists, plus the off-list words, as demonstrated in Figure 4. Each grade level uses mainly K1 and K2 words, as would be expected, with fewer lower frequency words used. However, we can see there is a slight progression, with Grade A texts using more lower frequency vocabulary than the other grades, and Grade D using fewer lower frequency words. It is also noticeable that there is a smaller percentage of off-list words occurring at Grade A than at all of the other levels, notably Grade D.

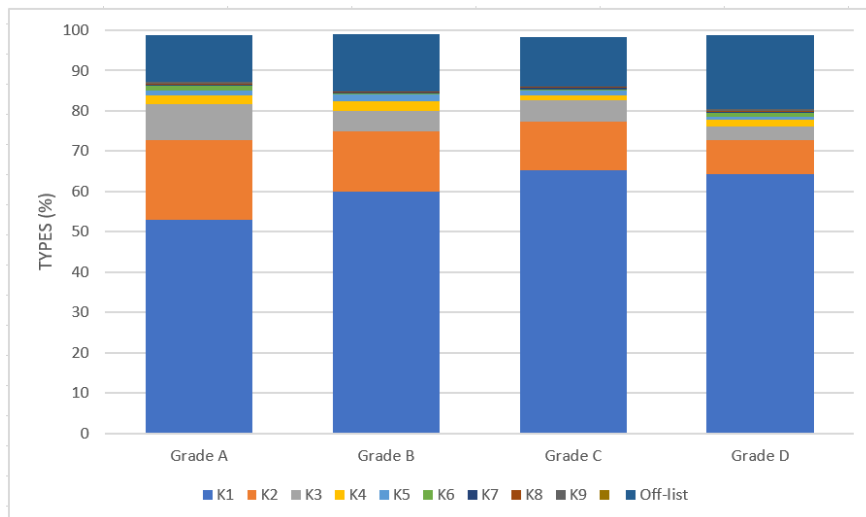


Figure 4: Percentage of K1 to K9 level and off-list words in student texts, according to grade awarded

A more qualitative analysis of this data was needed to assess performance at each grade. At Grade A, almost all the lower frequency words (K4 to K9) are correctly attributed to their level. Only two words are misspellings, *witch* for *which* and *jut* for *just*. In the off-list words, we find *cyberbully*, *digitalized*, *english*, *google*, *instagram*, *ipads*, *medias netflix*,

scamming, *smartphones* and *vlogs*, making 11 of the 61 off-list words (18%) correct. In contrast, at Grade D, many of the lower frequency words are misspellings. In the off-list, 11 of the 129 words (9%) are correct (*eriksson*, *gamer*, *gamers*, *instagram*, *ipad*, *ipads*, *samsung*, *sweden*, *texted* and *webcam*). The same pattern follows for Grade B and Grade C; at Grade B, 14 out of 91 (15%) off-list words are correct, and at Grade C, 9 of the 84 (11%) off-list words are correct. Space precludes a fuller qualitative analysis of contextual use of the lower frequency vocabulary, but this is also expected to shed light on achievements at different levels.

Although this preliminary examination is based on only a small data sample, it points to a relationship between grade and both the range and accuracy of vocabulary used. We aim to deepen our understanding of this relationship with further analysis on a larger scale as the database grows. A fuller qualitative analysis of how the words are used in context will be needed to complement a future analysis on a larger scale. We also see further opportunities in exploring gender-related performance, following on from Börjesson and Nilsson's (2018) study (see section 3), and the impact of task on performance.

8. Applications of the NTE database

As indicated in section 5, we see a number of applications for the corpus beyond our own research. It has already been used various ways as outlined in 8.1–8.3 below, and in 8.4 we outline our plans for an upcoming series of workshops with the teachers at schools we have collaborated with.

8.1 Teaching

The written part of NTEs constitutes a didactic resource for exploring L2 learners' proficiency in English in terms of language choices (style) and accuracy. A clear majority of the tests represented in the database have been awarded an A and B on the written production part (approximately 75 percent of the total grades), despite the incorrectly used contractions, i.e., *they're* instead of *their*, *it's* instead of *its*, lack of capitalization (with the pronoun *I*, proper nouns), colloquial expressions, spelling errors, and the conjunctions *or*, *and*, *but* used sentence-initially. These features are representative of spoken English, rather than a formally written test, and thus require closer examination in terms of the purpose of the text and the

intended audience (Skolverket 2022). In line with McCarthy et al.'s (2010) approach, which places different text types in English on a continuum of 'informal spoken' to 'formal written' English, Lundahl (2021: 400) emphasizes the importance of the learner's personal touch in their written output in favor of the 'unnecessarily complicated texts' (our translation). However, this view of L2 learners' writing may pose a challenge for student teachers as they tend to interpret the curriculum descriptors and support materials (i.e., assessment guidelines) in a literal way, which results in a harsher approach to grading overall (Lundahl 2022; Siegel 2022: 84). This view may also extend to practicing teachers (see section 7.1).

8.2 Assessment and grading

The database of English national tests has been an important tool in the subject's courses for continuing education for teachers (with a focus on grades 7–9). The student teachers have worked with the learners' written texts and analysed how these relate to the assessment guidelines and the knowledge requirements for the school subject English. The students have emphasized the importance of having access to the authentic material in the form of the pupils' texts and the active teachers' interpretation of the National Agency for Education's instructions and guidelines. Their assignment using the database comprised two parts, namely:

1. Carry out an analysis of the writing task of the NTE, focusing primarily on how the instruction text, including the mind map provided, can be linked to the knowledge requirements for English (year 9);
2. Choose one of the pupils' texts provided and grade/assess it in terms of how it corresponds to the knowledge requirements for English year 9. Also, discuss the function of NTEs in relation to the types of assessment (covered in the course literature).

The assignment was first completed individually and then discussed in pairs in a forum in the virtual learning platform Moodle. Interestingly, most of the students' comments addressed the pupils' accuracy in English, such as typos, subject-verb agreement mistakes, lack of punctuation, etc. This contrasted with the comments we have so far examined by teachers

represented in the NTE database (see section 7.1). These first focused on the overall impression of the pupils' texts, followed by a comment on language, while the student teachers' main comments concerned language-related details.

8.3 Student research

We planned that the database would be used for research purposes, at faculty and student level, and even at this early stage it has been used as source material for two of the English subject's student teachers' independent projects at undergraduate and advanced level this semester (Spring 2023). The undergraduate project focused on the students' linguistic competence in English based on their sentence structure, vocabulary and requirements fulfillment regarding the adaptation of written texts (Skolverket 2022). The advanced level project (Schalander 2023) examined the grading teachers' approaches to the assessment of national tests, finding that that form-focused and content-focused types of feedback are almost equally represented in the NTE data analysed. The teachers' individual styles were also compared, and *indirect* feedback was found to be the most frequent way of framing criticism.

8.4 Collaborative workshops

A workshop series for the staff of the participating schools is planned from Autumn 2023. The overarching theme of this series will be assessment and grading, and these are envisaged as practical sessions focused around texts drawn from the NTE database. In light of the new curriculum/syllabus for the English subject being launched recently, the key terms featured in the steering documents require additional scrutiny, including their practical application in the L2 teaching classroom and impact on teacher autonomy. Due to the generally vague nature of policy texts (Siegel 2022: 84), the teacher's role in education when it comes to interpreting and implementing these documents has been seen as a form of weakness (Biesta et al. 2015: 624). There is a general tendency towards making the education system more centralized (controlled at the national level) (Lundahl 2019: 64), which could be partly reflected in the proposed digitization of national tests (see section 2). Access to experienced teachers' assessment and feedback practices in the form of a digital database allows for 'evidence-based and data-driven approaches', advocated for by the proponents of

centralization of education (Biesta et al. 2015: 624), while boosting individual teachers' confidence regarding their grading and assessment competences rather than undermining their role in the teaching process. A hands-on approach to assisting the in-service teachers in their interpretation of the steering documents could be a model for breaking down the core content/knowledge requirements (Skolverket 2022) into a number of specific *grading criteria* designed for a particular task (Lundahl 2019: 427; see also Erickson and Tholin 2022), such as writing a letter on media to a magazine, featured in the NTEs examined. Following the lead of Sandlund and Sundqvist's (2019, 2021) studies into rater attitudes and perceptions, we anticipate that these sessions will also generate useful data in this area.

9. Summary

Considering the importance of the NTE in the school careers of Swedish pupils and their teachers, surprisingly little research has been carried out into it. The collaborative project we outline here aims to encourage further studies in this area. By providing an accessible, easily manipulated database of graded written papers, the corpus represents a valuable resource for research into test-takers' performance. The range of texts at different grades also provides material for improving confidence and competence in assessment procedures for teachers at all levels, whether trainee, novice or experienced. Existing studies into the NTE and assessment literacy in general have indicated that further research and training needs to be done. We aim to engage with the other stakeholders in this project to address some of these areas. The NTE project is still in its very early stages, but even the limited database compiled so far has demonstrated its value, and we look forward to building on this. As the project grows, we anticipate further collaboration with schools and universities in other regions, as we work towards creating a larger, more representative national database.

Acknowledgements

The NTE project has received funding from the national project ULF (Utbildning, Lärande, Forskning), administered at Mid-Sweden University in collaboration with Umeå University.¹⁰

References

- ALLEA. 2023. *The European Code of Conduct for Research Integrity*. Rev. ed. Berlin: ALLEA (All European Academies), doi:10.26356/ECOC.
- Anthony, Laurence. 2022. AntConc (Version 4.2.0) [Computer Software].
- Biesta, Gert, Mark Priestly, and Sarah Robinson. 2015. The role of beliefs in teacher agency: Teachers and teaching. *Theory and Practice* 21(6): 624–640.
- Börjesson, Lena, and Sofia Nilsson. 2018. Könnskillnader i engelsk språkfärdighet. In *Att bedöma språklig kompetens: Rapporter från projektet Nationella prov i främmande språk*, edited by Gudrun Erickson, 107–118. <https://gupea.ub.gu.se/handle/2077/57683>.
- Cobb, Tom. 2023. *Compleat Lexical Tutor*. <https://www.lexutor.ca/>.
- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. 2020. *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume*. Strasbourg: Council of Europe Publishing. www.coe.int/lang-cefr.
- Crusan, Deborah, Lia Plakans, and Atta Gebriel. 2016. Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28: 43–56. doi.org/10.1016/j.asw.2016.03.001.
- Erickson, Gudrun. 2019. Holistic peer analyses of national tests in relation to the CEFR. In *Developments in language education memorial volume in honour of Sauli Takala*, edited by Ari Huhta, Gudrun Erickson, and Neus Figueras, 49–66. Jyväskylä: University Printing House.

¹⁰ <https://www.miun.se/mot-mittuniversitetet/samverkan/run/praktiknara-skolforskning/ulf-utveckling-larande-forskning/om-ulf/>.

- Eriksson, Gudrun. 2020. National assessment of foreign languages in Sweden. Gothenburg: University of Gothenburg. <https://www.gu.se/nationella-prov-frammande-sprak/information-om-projektet-nafs>.
- Erickson, Gudrun, and Heini-Marja Pakula. 2017. Den gemensamma europeiska referensramen för språk: Lärande, undervisning, bedömning: ett nordiskt perspektiv. [The common European framework of references for languages: Learning, teaching, assessment: A Nordic perspective]. *Acta Didactica Norge* 11: 1–23. <https://doi.org/10.5617/adno.4789>.
- Erickson, Gudrun, Linda Borger, and Eva Olsson. 2022. National assessment of foreign languages in Sweden: A multifaceted and collaborative venture. *Language Testing* 39(3): 474–493. doi.org/10.1177/02655322221075067.
- Erickson, Gudrun, and Jörgen Tholin. 2022. Overall, a good test, but...: Swedish lower secondary teachers' perceptions and use of national test results of English. *Languages* 7(1): 64. doi.org/10.3390/languages7010064.
- Fulcher, Glenn. 2012. Assessment literacy for the language classroom. *Language Assessment Quarterly* 9(2): 113–132. [doi:10.1080/15434303.2011.642041](https://doi.org/10.1080/15434303.2011.642041).
- Inbar-Lourie, Ofra. 2017. Language assessment literacy. In *Language testing and assessment. Encyclopedia of language and education*, edited by Elana Shohamy, Iair Or, and Stephen May, 257–270. New York: Springer. doi.org/10.1007/978-3-319-02261-1_19.
- Gallringsråd för kommuners, landstings och regioners utbildningsväsende. 2011. *Bevara eller gallra 2*. 4th ed. SKL Kommentus Media.
- Lundahl, Bo. 2019. *English language education: Texts, communication, language development*. 4th ed. Lund: Studentlitteratur.
- Lundahl, Bo. 2021. *Engelsk språkdidaktik: Texter, kommunikation, språkutveckling*. 5th ed. Lund: Studentlitteratur.
- Lundahl, Bo. 2022. Language teaching in the intersection between theory and practice. Presentation at Mid-Sweden University, 14 April 2022.
- McCarthy, Michael, Christian Matthiessen, and Diana Slade. 2010. Discourse Analysis. In *An introduction to applied linguistics*, edited by Norbert Schmitt, 2nd ed., 53–69. London: Hodder Education.

- NAFS Project. 2022. Ämnesprov i engelska för årskurs 9 [English National Test Year 9]. <https://www.gu.se/nationella-prov-frammande-sprak/prov-och-bedomningsstod-i-engelska/engelska-arskurs-7-9/nationellt-prov-i-engelska-for-arskurs-9>.
- Olsson, Eva. 2018. Vokabulär och bedömning av skriftlig förmåga. In *Att bedöma språklig kompetens: Rapporter från projektet Nationella prov i främmande språk*, edited by Gudrun Erickson, 139–161. <https://gupea.ub.gu.se/handle/2077/57683>.
- Olsson, Eva, Sofia Nilsson, and AnnaKarin Lindqvist. 2018. Test-taker feedback i utvecklingsprocessen av nationella prov i engelska. *Acta Didactica Norge* 12(4): 1–27. doi.org/10.5617/adno.6288.
- Reynolds-Keefer, Laura. 2010. Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research, and Evaluation* 15(1), Article 8: 1–9. doi.org/10.7275/psk5-mf68.
- Roslan, Roslinawati, Yuri Nishio, and Rosmawijah Jawawi. 2022. Analyzing English language teacher candidates' assessment literacy: A case of Bruneian and Japanese universities. *System* 111: n. pag. doi.org/10.1016/j.system.2022.102934.
- Sandlund, Erica, and Pia Sundqvist. 2019. Doing versus assessing interactional competence. In *Teaching and testing L2 interactional competence: Bridging theory and practice*, edited by M. Rafael Salaberry and Silvia Kunitz, 357–396. Abingdon and New York: Routledge.
- Sandlund, Erica, and Pia Sundqvist. 2021. Rating and reflecting: Displaying rater identities in collegial L2 English oral assessment. In *Assessing speaking in context: Expanding the construct and its applications*, edited by M. Rafael Salaberry and Alfred Rue Burch, 132–161. Bristol: Multilingual Matters. doi:10.21832/9781788923828-007.
- Saville, Nick, and Roger Hawkey. 2010. The English Profile Programme: the first three years. *English Profile Journal* 1(1), E7: 1–14. doi:10.1017/S2041536210000061.
- Schalander, Sofia. 2023. Examining the examiners: A study of two secondary English teachers' written feedback on the 9th grade national test in English. MA dissertation, Mid-Sweden University.

- Schmitt, Norbert, and Diane Schmitt. 2014. A reassessment of frequency and vocabulary size in L2 vocabulary. *Language Teaching* 47: 484–503. doi:10.1017/S0261444812000018.
- Siegel, Joseph. 2022. *Teaching English in secondary school: A handbook of essentials*. Lund: Studentlitteratur.
- Skolverket. 2022. *Resultat på de nationella proven i årskurs 3, 6 och 9. Läsåret 2021/22*. Stockholm: Skolverket. <https://www.skolverket.se/publikationsserier/beskrivande-statistik/2022/resultat-pa-de-nationella-proven-i-arskurs-3-6-och-9---lasaret-2021-22>.
- Tsagari, Dina, and Karin Vogt. 2014. Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly* 11: 374–402. doi.org/10.1080/15434303.2014.960046.
- Walker, Terry, and Rachel Allan. 2018. Bridging the gap between university and upper secondary school studies: The ULE Project. *ICAME Journal* 42: 191–212. doi:10.1515/icame-2018-0009.
- Yan, Xun, Cong Zhang, and Jason Jinsong Fan. 2018. ‘*Assessment knowledge is important, but ...*’: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System* 74: 158–168. doi.org/10.1016/j.system.2018.03.003.