National assessment of foreign languages in Sweden

Gudrun Erickson
University of Gothenburg, Sweden
gudrun.erickson@ped.gu.se

This text was originally published in 2004.
Gradual revisions and additions have been made in May 2007, April 2009, October 2012, August 2015, November 2017, and February 2020.

There is a long tradition of assessment at the national level in Sweden. In connection with the current goal- and criterion-referenced system, assessment materials of various kinds, covering different subjects, are offered throughout the school system. Different universities are commissioned by the Swedish National Agency for Education (NAE) to take responsibility for test development and research, for example the University of Gothenburg for foreign languages – English, French, German and Spanish.

Facts about Sweden

The Swedish school system comprises preschool, compulsory preschool class, nine-year compulsory school and three- year upper secondary school, each with its own national curriculum. For compulsory and upper secondary education, there are syllabuses for individual subjects. Objectives and core content, as well as performance standards, referred to as 'knowledge requirements', are defined nationally, whereas detailed content, materials and methods are to be decided locally. Personal development dialogues between students, teachers and guardians are to be held at least once a term, and written reports are issued from primary school up to school year six as part of each student's individual development plan. Formal grades, however, are not awarded until school year six, when pupils are around 12 years old.

Teachers are responsible for the evaluation and grading of their own students' achievements. There are no formal examinations, but an extensive system of national assessment materials and tests at different levels is aimed to support teachers in their decisions concerning individual students' competences in relation to the national objectives and performance standards. Consequently, the national assessment system can best be characterized as advisory. As from 2018, however, when certain changes and modifications were decided at the system level, it is stated that results from national tests are to be taken into 'special consideration' – so far, however, not further defined – when teachers decide on individual students' grades. Moreover, it needs to be emphasized that, at present, no central marking takes place at the national level, but is sometimes arranged at local level. Usually, however, teachers take responsibility for the marking of their own students' national tests. There is a strong recommendation, however no formal demand, that this should be done in collaboration with colleagues. [Changes to the current system are being discussed, and have in some cases been decided; this will be commented on in the Post scriptum section.]

Why national assessment?

The main function of the system of national assessment is to support and advise teachers, and, to a certain extent, students as well, in their decision-making concerning diagnosing, planning and grading. Importantly, the system aims at enhancing comparability and equity within the school system, something that is increasingly emphasized. However, the system is sometimes also used, implicitly, to clarify and exemplify the view of knowledge and language expressed in the national curricula and syllabuses. Results may also be used in local and regional

evaluations and, to some extent, in national evaluation projects of the school system at large. Thus, there are several more or less implicit uses of and aims to the system, however with the one regarding equity the only one expressed explicitly. Following this, a systemic framework for all national tests, aimed at further strengthening and standardizing procedures and products, was introduced in 2017 (Skolverket, 2017).

There is a fair amount of consensus around the ambition to maintain a system in which assessment is regarded as an integrated part of the educational process, which, ideally, should work *for* learning, as well as being a reliable indicator *of* learning, and of course never go *against* learning. This means that it ought to cover as much as possible of the construct in focus and generate results that are trustworthy and stable over time. Students should be offered a variety of tasks, as authentic as possible, and results should be presented in ways that help each student gain insights about strengths and weaknesses in his/her individual profile of competence, and to plan, together with the teacher, how learning can be optimized.

What is assessed and how?

There is a fairly long tradition of communicatively oriented language teaching and assessment in the Swedish system, clearly articulated in the national curricula as from the early 1980s. The current Swedish national syllabuses for foreign languages (Lgr11 and Lgy11) are to a considerable extent influenced by, and to some extent comparable to, the Common European Framework of Reference (CEFR), although the seven steps of proficiency defined have not been fully empirically aligned to the six levels of the CEFR. Areas focused upon are receptive, productive and interactive competences, as well as intercultural communicative competence. Furthermore, strategic competence and adaptation to purpose, recipient and situation are explicitly defined as learning outcomes. Subsystems like vocabulary, grammar and pronunciation are considered important prerequisites but not as goals *per se*.

The national assessment materials do not cover all areas of the syllabuses. One reason is that they are not to be regarded, or used, as final examinations, but as supplementary, advisory materials. Another reason is that, due to their character, some objectives can be better evaluated continuously, in the learning- and teaching process. This applies in particular to the one aimed at intercultural, communicative competence.

Altogether, at present, full national assessment materials of foreign languages are provided for six different stages of English, and three of French, German and Spanish. All of them include tasks aimed at testing receptive competence and oral as well as written production and interaction. Aspects of culture are reflected in the materials, mainly in the choice of texts, and in topics for oral and writing tasks. Models for developing students' reflective skills, e.g. self- and peer-assessment, are offered for four stages of English and one of French, German and Spanish. Moreover, there are illustrative materials focusing on partial competences (oral and/or written production and interaction in particular) for English as well as for French, German and Spanish.

A typical national test comprises four parts: an oral test, in which pairs of students talk about different subjects, a listening comprehension and a reading comprehension section with a variety of texts and tasks, usually combined to a single score for receptive skills, and a writing test, in which students are sometimes offered a choice between two different subjects. There are extensive teacher guidelines for all materials. These include test specifications, commented answers and authentic samples of benchmarked oral and written performance, cut-off scores etc.

Gudrun Erickson, 2020
gudrun.erickson@ped.gu.se

Students are informed about the national tests in different ways, by their teachers, in standardized letters, and through extensive sample materials published on the Internet. For French, German and Spanish there is an electronic test bank with different levels of accessibility, from totally confidential testing materials, via old tasks from previous national tests for teachers to include in their own tests, to completely open tasks, which serve the function of information and practice, if needed. (For further information about the different assessment materials, including samples of tasks, see the project website http://www.nafs.gu.se/english/information/. Information about the Swedish school system can be found at https://www.skolverket.se/andra-sprak-other-languages/english-engelska#h-OfficialstatisticsofSweden.

How materials are developed and standards set

The national assessment materials have partly different aims and character, from purely formative and low-stakes, to distinctly summative, compulsory and high-stakes. However, they are all based on a set of basic principles, some of which are the following:

• Making what is most important assessable, not making what is easily measurable the most important;
• Giving students the chance to show what they actually know and can do, instead of primarily trying to detect/focus on what they do not know/cannot do, e.g. by providing broad, multifaceted, varied, monolingual tests, with – to as large an extent as possible – progression of difficulty, within and between tasks;
• Enhancing validity and reliability – avoiding bias, for example by developing tests in collaboration with a wide group of stakeholders, and pre-testing all materials in large, randomly selected groups across the country;
• Detecting and presenting as much as possible of individual students' results in profiles – strengths as well as weaknesses;
• Commenting on strengths before weaknesses; when analysing weaknesses, distinguish between errors that [might] *disturb* and errors that actually *destroy* [impede] communication, i.e. between errors representing different degrees of gravity.

Considerations underlying item writing and composition of full materials concern, e.g. content, relevance and level of difficulty in relation to the syllabus in focus, and aspects of time and format as well as of gender and culture. Considerable attention is paid to opinions expressed by students and teachers in connection with piloting and pre-testing of the different tasks.

All materials are developed in close cooperation with different categories of experts, among which students, i.e. what might be considered the real stakeholders, should not be forgotten. Contacts with different national and international institutions play an important role. For each material, there is a reference group comprising different categories of teachers, teacher educators and researchers within different fields. L1 speakers contribute in various ways to the developmental work.

Tasks – items and passages – are piloted during the initial development stage, revised, and then pre-tested in randomly selected schools throughout the country, normally by around 400 students per task. Anchor items are used consistently to enable comparisons across groups, and over time. During these iterative rounds, all students and teachers are asked to comment on the different tasks. A wide range of qualitative and quantitative methods is used to analyse the results, i.e. both performance and perception data. Less well-functioning items and tasks are either removed or adjusted and then pre-tested again, until they are finally considered for inclusion in one of the materials.

Gudrun Erickson, 2020
gudrun.erickson@ped.gu.se

Standards are set in collaboration with groups of experienced teachers, employing an eclectic approach, i.e. by combining different methods for standard setting (as often recommended in the literature), with test-centred as well as student-centred points of departure: Teachers "take the test", analyse and estimate the items/tasks in relation to the syllabus, and to their experiences of student performances at the level in focus, and then suggest cut-off scores for the different grade levels. Different data from the pretesting rounds are introduced towards the end of each session and play a role in the final recommendations and decisions made. As for the selection of benchmarked samples of oral and written performances, approximately ten teachers analyse and rate, independently, a large number of authentic samples. The ratings are then analysed, with regard to inter-rater reliability, distribution etc., examples are chosen and comments produced for the teacher guidelines.

For the French, German and Spanish materials, based on a common syllabus, a three- phased standard setting model is used. First, standards are set for each test separately, according to the model described above; after that the tests are compared, and standards suggested, by groups comprising teachers who are academically qualified and experienced in teaching two of the languages. In this phase, a list of parameters, produced in collaboration between linguists and psychometricians, is used to make sure that a wide range of relevant aspects are taken into account. Before the final decisions are made about standards and benchmarks, the results from the first two standard setting phases are compared, and data from the different pretesting rounds are further considered.

Results and reactions

Test results are continuously and routinely analysed with regard to various aspects of validity and reliability. Matters investigated obviously concern aspects of facility, distribution, internal consistency, and rater agreement. In 2008, a large-scale study of inter-rater agreement and consistency was undertaken, focusing on the final tests of English, Mathematics and Swedish at compulsory school level. 100 randomly selected, teacher rated tests were independently re-rated by three external raters. The results for English were very positive, with almost total agreement for constructed response items in Listening and Reading comprehension, and correlations between .86 and .93 for Writing; generalizability coefficient .85 (Erickson, 2009; http://www.nafs.gu.se/publikationer/). Studies of the Speaking components of the test have indicated roughly the same results as for Writing. Since identical routines are applied in the development of all FL materials, including the teachers' scoring guides provided, it can be tentatively assumed that these results are, at least to some extent, relevant to the other FL testing materials produced within the project.

As for validity, reliability and stability, continuous external and internal analyses are made to ensure that high quality is maintained and further developed. In these, validity and reliability have been found to be at a very high level (Verhelst, 2013), and continuous studies of the outcome of the Swedish national tests show that the tests of English are very stable over time (Erickson, 2018).

Test takers often give valuable suggestions for improvement of tasks, for example concerning content, clarity and perceived level of difficulty, the latter especially useful in sequencing decisions. In general, the following has been noted about students' attitudes[1]:

• Students tend to appreciate tasks that are considered authentic, pedagogical, fair and challenging;

[1] Mainly based on analyses of 15-year-old students' comments on tests of English.

Gudrun Erickson, 2020
gudrun.erickson@ped.gu.se

• Fairly regardless of students' level of proficiency, oral tasks and Writing are often the most appreciated parts of the tests.

In addition, it should be mentioned, that very similar results emerged in a survey of students' [and teachers'] views on language testing and assessment, conducted in ten European countries in 2005. A report on this (Erickson & Gustafsson, 2005) can be found on the website of the European Association for Language Testing and Assessment (EALTA), under Resources (http://www.ealta.eu.org/resources.htm). Recent studies of test-taker feedback (Finndahl & Perrotte, 2018; Hedenbratt & Axelson, 2018; Olsson, Nilsson & Lindqvist, 2018; Sebestyén & Albinsson, 2018;) further describe and comment on various aspects of students' contributions to the test development process.

Teachers' reactions to national testing and assessment are generally very positive, both to the principle as such, and to the different materials. During the past ten years, more than 90 per cent have expressed positive opinions, often concerning the breadth and variation of the tasks, the close connection between the materials and the syllabuses, the profiled presentation of results, and the support provided in the guidelines. Certain criticism obviously also occurs, mostly concerning workload – broad, qualitative assessment takes time, too much time some teachers seem to feel – but in some cases also regarding the levels of EFL proficiency required for a Pass, especially in lower secondary school, that some teachers find too low.

The outcomes of the different assessment rounds are analysed and commented on in regular reports made public on the Internet. A recent example of an additional type of publication is a compilation of articles on different aspects of the national assessment materials, authored by altogether 19 members of the test development group (Erickson, 2018, ed.). Examples of topics focused upon are aspects of test development, different types of language related issues, gender differences in test results, and practical issues of test implementation and use at local levels.

Concluding remarks

The system of national assessment in Sweden is flexible and dynamic, which means that changes, initiated by different stakeholders and based on thorough development work, are gradually introduced. However, the basic principles as well as the collaborative approach to test development remain the same, since it is felt that this contributes to validity and stability, as well as to a continued, positive assessment climate within, and hopefully also outside, the national assessment system.

= = = = = = = = =

POST SCRIPTUM 2020

During the last decade, several changes to the Swedish school system have been undertaken, for example the introduction of earlier grading, an increased number of grade levels, more national tests, and tests in wider range of subjects. Furthermore, the national tests have been criticized, following studies by the Swedish Schools Inspectorate, indicating problems concerning inter-rater consistency [in the case of English, not corresponding to the results obtained in 2008, as reported on p. 4]. In addition, there are demands of clarification of the role and weight of the national tests in relation to teachers' grading. Following this, a politically initiated, independent inquiry of the system at large was undertaken by a special investigator and reported in March 2016 (SOU 2016:25). After extensive consideration by a wide range of stakeholders, gradual changes and modifications based on the inquiry have been politically suggested and/or decided. Probably the most noticeable of these is the decision about digitalization of the assessment system, which was originally intended to be

completed in 2022 but which has recently been postponed by one year. Another concerns the weight of the aggregated national test result in teachers' grading, which is to be increased; teachers shall now take the results into 'special consideration', however not quantified, but still combine the results with their continuous observations. Furthermore, the government proposes mandatory marking of national tests by someone else than the students' own teachers (methods still to be suggested). Finally, the number of mandatory national tests in upper secondary school has decreased; as from 2018, only tests in final courses for the different study programs are obligatory, whereas schools are free to decide whether the preceding tests are to be used.

Further reading

Erickson, G. (2019). Holistic Peer Analyses of National Tests in Relation to the CEFR. In Huhta, A., Erickson, G. & Figueras, N (eds.), *Developments in language education – a memorial volume in honour of Sauli Takala.* http://www.ealta.eu.org/resources.htm

Erickson, G. (ed.). (2018). *Att bedöma språklig kompetens. Rapporter från projektet Nationella prov i främmande språk [Assessing language competence. Reports from the National assessment project for foreign languages].* University of Gothenburg, Reports from the Department of Education and Special education, issue 16. http://hdl.handle.net/2077/57683

Erickson, G. (2018). Extern bedömning av kunskap *eller* Om värdet av att se med andras ögon [External assessment of knowing, *or* On the value of seeing through the eyes of others]. Skolverket, *Artiklar till webbkurs om bedömning och betygssättning* [Articles for a web based course on assessment and grading]. *https://www.skolverket.se/skolutveckling/kurser-och-utbildningar/artiklar-till-webbkurser-om-bedomning-och-betygssattning*

Erickson, G. (2010). Good Practice in Language Testing and Assessment – A Matter of Responsibility and Respect. In Kao, T. and Lin Y. (Eds.), *A New Look at Teaching and Testing: English as Subject and Vehicle, pp. 237-258. Taipei, Taiwan: Bookman Books Ltd.*

Erickson, G. & Åberg-Bengtsson, L. (2012). A Collaborative Approach to National Test Development. In Tsagari, D. & Czepes, I. (Eds.), *Collaboration in Language Testing and Assessment*, pp. 93-108. Frankfurt: Peter Lang Verlag.

Finndahl, I. & Perrotte, F. (2018). *'Det var faktiskt kul att prata om vardagliga saker som man kan prata om i Frankrike typ' [It was actually fun to talk about everyday things that you can talk about in France].* In Erickson, G. (ed.); see reference above.

Hedenbratt, L. & Axelson, P. (2018). *Motiverad för uppgiften? Test-taker feedback ur ett genusperspektiv. [Motivated for the task? Test-taker feedback from a gender perspective].* In Erickson, G. (ed.); see reference above.

Olsson, E., Nilsson, S., & Lindqvist, AK. (2018). Test-taker feedback in the development process of National tests of English. *Acta Didactica Norge, 12*(4).

Sebestyén, Å. & Albinsson, H. (2018). *Was piensan les élèves? Test-taker feedback i samband med utprövningar i tyska, spanska och franska [Was piensan les élèves? Test-taker feedback in connection with pre-testing in German, Spanish and French].* In Erickson, G. (ed.); see reference above.

Gudrun Erickson, 2020
gudrun.erickson@ped.gu.se

Statens offentliga utredning, SOU 2016:25: *Likvärdigt, rättssäkert och effektivt – ett nytt nationellt system för kunskapsbedömning* [Equivalent, fair and efficient – a new, national system for knowledge assessment.] Stockholm: Utbildningsdepartementet. http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2016/03/sou-201625/

Verhelst, N. (2013). Comparison of national tests across time: some psychometric and statistical issues. Intern rapport till Skolverket [Internal report for the Swedish National Agency for Education].

Gudrun Erickson, 2020
gudrun.erickson@ped.gu.se