

# Speaking French, German and Spanish in Swedish lower secondary school – a study on attained levels of proficiency

Jonas Granfeldt, Lund University  
Gudrun Erickson, University of Gothenburg  
Camilla Bardel, Stockholm University  
Susan Sayehli, Stockholm University  
Malin Ågren, Lund University  
Rakel Österberg, Stockholm University

*This study investigates levels of oral proficiency in French, German and Spanish attained by Swedish speaking students in lower secondary school. A total of 122 students performed two tasks: one production task and one interaction task. The oral performances were rated using scales from the Common European Framework of Reference for Languages. The rating was done in successive steps by researchers in the project and external raters. The results show that slightly less than half of the students' performances were rated at or above the expected proficiency level at the end of lower secondary school (A2.1). While there was no difference in rated levels between the two tasks, the performances by the students of German were significantly more often rated at or above the A2.1 level than the performances by students of French and Spanish. In the article, we discuss the results in relation to the few previous studies available on the topic, as well as some aspects of the learning conditions that might contribute to the interpretation of the results. In addition, certain structural phenomena regarding language education in Sweden are briefly considered in relation to equity at a general level.*

**Keywords:** Sweden, lower secondary school, oral proficiency, assessment, second foreign languages, French, German, Spanish

## 1 Introduction

This paper investigates and discusses attained levels of oral proficiency among students of the three most frequently studied second foreign languages, French, German and Spanish, at the end of lower secondary school in Sweden (age 15).

---

Corresponding author's email: [jonas.granfeldt@rom.lu.se](mailto:jonas.granfeldt@rom.lu.se)

eISSN: 1457-9863

Publisher: University of Jyväskylä, Language Campus

© 2023: The authors

<https://apples.journal.fi>

<https://doi.org/10.47862/apples.127819>

Speaking is often a central objective for students when learning a new foreign language (Fernández & Andersen, 2019). This is not surprising because an important part of everyday communication takes place in the spoken modality. Oral production and interaction are therefore central skills for students to develop both in the first foreign language (normally English) and in the second foreign language (SFL). Moreover, communicative competence is commonly defined as the main objective of foreign language education, illustrated, for example, by the action-oriented approach to foreign language learning and teaching (Council of Europe, 2001). In the current Swedish national syllabuses for SFLs - the subject labelled 'Modern languages' - there are descriptions of core contents and definitions of performance standards (criteria) to be achieved for speaking.

An important empirical question is to what extent students achieve the intended level of proficiency, not least in a perspective of quality assurance and quality enhancement of (second) foreign language education. However, even though there are two major sources of information to study students' achievements: teacher assigned subject grades and results on national tests, knowledge is currently lacking particularly referring to the attained oral proficiency in SFLs. The grades awarded by teachers are summative assessments drawing on different materials and performances over a long period of time. Therefore, grades are less informative with respect specifically to attained levels of oral proficiency. In Sweden, mandatory national tests of foreign languages are provided only for English, but there are extensive, albeit not obligatory, [national assessment materials](#) for SFLs used by most schools (Erickson, 2020). However, only some schools choose to report their test results. This means that for SFLs like French, German and Spanish there is no systematic and nation-wide collection of student performances that can answer the question about attained levels of oral proficiency. Beyond grades and national assessment materials, also external assessments projects could be a source of information. However, at the time of publication there are no large-scale international studies focusing on oral language proficiency in SFLs available. We will come back below to the findings and the debate in relation to one international survey, *The first European Survey on Language Competences* (ESLC), that included one productive modality, namely writing, but not speaking.

In sum, we know very little about students' oral proficiency by the time they have studied their SFL for a number of years in lower secondary school. Therefore, the main objective for the current article is to further that knowledge by reporting on a study of 122 15-year-old students of French, German and Spanish performing two oral tasks - one production and one interaction task. The tasks were aligned to the national syllabus and the performances were rated in accordance with the *Common European Framework of Reference for Languages* (CEFR) (Council of Europe, 2001).

## **2. Background**

### *2.1 The context of second foreign languages in Sweden*

English is the first and only obligatory foreign language in Swedish lower secondary school. Schools need to start English instruction in Year 3 (age 9) at the latest but are free to do so earlier. The total number of teaching hours in English until the end of lower secondary school is 480 hours.

While it is not mandatory to study a SFL in Swedish compulsory education, students have to make a ‘language choice’ (Språkval) (see Bardel, Ericksson & Österberg, 2019). At the latest in the year preceding Year 6 (age 12 years), they are required to choose one of the SFLs offered by the local school or an alternative to a SFL, which can be mother tongue instruction (if other than Swedish), additional Swedish or English, or Swedish sign language. Since the study of a SFL is not obligatory, students can also opt out after having started. According to data from school year 2021/22, about 90% of all students start studying a SFL, however at the beginning of Year 9 only 73% of them remain to do so (National Agency for Education (NAE), 2022c).

The issue of SFL being optional in compulsory school has brought about intense discussions since the latter part of the 20<sup>th</sup> century, not least from the perspective of equity (e.g., Erickson, Bardel, Österberg & Rosén, 2022; Krih, 2019). This aspect, implying obvious selectivity regarding students as well as differences in opportunities to learn, is not actively dealt with in the current article but needs to be borne in mind in relation to the different results reported and discussed.

Organisers are required by the school ordinance (Ministry of Education, 2011) to offer at least two of the three languages French, German and Spanish as SFLs with the majority of schools offering all three, even though there is some regional variation (Granfeldt et al., 2019, 2021). Other SFLs may be offered by schools, but in practice this is very rare. In school year 2021/22, Spanish was studied by 43% of all students in year 9 of lower secondary school, German by 16% and French by a bit fewer, 14%. The remaining 27% did not study a SFL (NAE, 2022c).

All SFLs share the same syllabus within the school subject labelled ‘Modern languages’. The total number of teaching hours for SFLs until the end of lower secondary school is 320, with a minimum of 48 h before Year 7 (normally in Year 6).

**Table 1.** English and SFLs in compulsory school

Subject	Status	Latest starting year (age)	Total number of teaching hours
English	Obligatory	Year 3 (age 9)	480
SFL	Optional	Year 6 (age 12)	320

## 2.2 Swedish language syllabuses and the CEFR

Foreign and second language syllabuses are part of the Swedish national curriculum. Since the early 1980s, they have been explicitly communicative and functional in their approach to language learning and teaching. Important influences can be traced back to Hymes (1972) and also to the Council of Europe’s work on language policies, in which Sweden took an active part (Malmberg, 2001; Trim, 2002). A gradual and general development towards language use is noticeable, most clearly manifested in the publication of the CEFR (Council of Europe, 2001), in which the expression ‘action-oriented approach’ summarizes the functional view of language conveyed. Regarding oral language proficiency (as well as writing), a distinction is made between production and interaction. This is further developed in the Companion Volume to the CEFR (Council of Europe, 2020), created to update the original 2001 document, however clarifying already on the cover page, that “the conceptual framework of which [the CEFR]

remains valid”. Further, the relationship between reception, production and interaction is expanded by and connected to mediation, a concept also included in the CEFR. Mediation is defined as activities that “make communication possible between persons who are unable, for whatever reason, to communicate with each other directly” (p. 14). Consequently, in a wide sense, this may be regarded as an advanced strategic competence, clearly connected, in particular, to oral interaction.

Language syllabuses, especially in Europe but also beyond its boundaries, are often influenced by the CEFR, both generally, through the action-oriented approach to language competence, and specifically, through the alignment to the six levels of language competence defined for different categories of users: A1-A2 basic, B1-B2 intermediate and C1-C2 proficient users (Council of Europe, 2001). The approach taken means, for example, that active language use, not least oral proficiency, is emphasized, often with an explicit focus on interaction. In addition, aspects like confidence, fluency and strategies have come to dominate over – and sometimes even replace – the traditional focus on language accuracy. Alignment between local as well as national levels of language proficiency and the CEFR may follow a meticulous scheme of procedures recommended by the Council of Europe (2009), further clarified and discussed by prominent language institutions, for example, the British Council, *Association of Language Testers in Europe* (ALTE) and *European Association for Language Testing and Assessment* (EALTA), including empirical analyses of assessment data. It may, however, also be of a more verbal and tentative character, comparing aims, functions and standards. The latter approach has been applied in Sweden, where national subject syllabuses for languages have a clearly action-oriented approach that has been textually compared to the CEFR in a number of studies (Erickson & Pakula, 2017; Oscarson, 2015), but where large-scale, empirical studies have not been conducted in any systematic way. However, examinations of a more qualitative kind, including international peer evaluation of national assessment materials, have been undertaken (Erickson, 2019). Taken together, these efforts have resulted in the following tentative comparison between a minimal Pass grade (E) for the seven Swedish levels of second and foreign language proficiency and the CEFR published by the NAE (2022b):

**Table 2.** National foreign language levels in Swedish curricula related to the CEFR (adapted from the NAE, 2022a, b)

CEFR level	Pre-A1	A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2.1	B2.2	C1	C2
Sw. levels			1	2	3	4	5	6	7		
Modern languages											
Year in Lower secondary	-Year 6 1st year long course 1 <i>Opt.</i>		Year 9 Short course <i>Opt.</i>	Year 9 Long course <i>Opt.</i>							
Course in Upper secondary			1	2	3	4	5	6	7		

<b>English</b>											
Year in Lower secondary				Year 6		Year 9					
Course in Upper secondary							5	6	7		

(opt = optional, the level in focus here is marked in bold face)

As shown, the minimal Pass level at the end of compulsory school required for the longer course of Modern languages – in the current study French, German and Spanish (‘Språkvalet’) – is considered to be A2.1, whereas the corresponding level for English is B1.1.

In the global scale of the CEFR, the A2 level (labelled Waystage) is defined as follows:

Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need. [Council of Europe, 2001, p.24]

In the illustrative descriptors used in the competence scales, a further distinction is sometimes made between the criterion level (e.g., A2 or A2.1) and ‘plus levels’ (e.g., A2+ or A2.2). This is exemplified below with the scale for overall spoken interaction from the CEFR (Council of Europe, 2001, p.74).

**Table 3.** Overall spoken interaction. Descriptors for A2.1 and A2.2 (above the horizontal line)

A2	Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations.
	Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of their own accord.

Generally speaking, the A2+ level represents a stronger proficiency at the Waystage level where learners are capable of participating more actively in conversations with some assistance and certain limitations. Additionally, learners at this level exhibit improved ability to sustain monologues, provide extended descriptions of their environment and personal experience and discuss habits, routines, plans, and arrangements.

Further information about the CEFR levels can be found on the [Council of Europe website](#); see also [Hellqvist & Thue Vold, 2022](#).

### 2.3 Oral proficiency – the construct

Oral language proficiency is a fundamental aspect of language competence (Linell, 2019), often described not only as multifaceted but also as more complex than other language competences, due to its elusive and temporary character and manifested in productive as well as interactive forms. This is especially true seen from an assessment perspective (e.g., Field, 2011; Fulcher, 2003). Oral proficiency can be conceptualized in different ways, analytically, focusing on [the combination of and relationships between] different components, or elements, such as phonology, vocabulary, and grammar (e.g., Iwashita, Brown, McNamara & O'Hagan, 2008; De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012), and holistically, based on contextually dependant, communicative functions. The latter, drawing on work by Hymes (1972) and its developments and additions by, for example, Canale and Swain (1980), Bachman and Palmer, 1996), and Byram (2021), is dominant in contemporary language syllabuses, so also in Sweden. In such documents, commonly comprising an intercultural component, the definition of oral proficiency is operational, that is, describing what learners can actually do with their language – on their own and in interaction with others – to achieve their goals. Furthermore, the communicative, or action-oriented perspective is fundamental to the development of international language documents, such as the CEFR and its more recent Companion volume (Council of Europe, 2001; 2020), which state that "the approach adopted here, generally speaking, is an action-oriented one in so far as it views users and learners of a language primarily as 'social agents', i.e., members of society who have tasks (not exclusively language-related) to accomplish in a given set of circumstances, in a specific environment and within a particular field of action" (2001, p. 9).

### 2.4 Assessment of oral language proficiency

Assessment is often used as an umbrella term for different methods to capture, describe and measure individual language proficiency, hence encompassing continuous assessments of a more formative kind as well as small- or large-scale, summative tests or exams. Furthermore, assessment is seen as closely connected to, sometimes intertwined with, learning and teaching, thus forming an educational whole; an approach reflected, for example, in the CEFR, also in its subtitle, 'learning, teaching and assessment'. When discussing assessment, some basic questions are useful for conceptual as well as operational purposes, namely why?, what?, how?, when?, who?, and..?, thus focusing on aims, construct, methods, frequency, agency, and uses, the latter also including consequences (Takala, Erickson, Figueras & Gustafsson, 2016). These questions also reflect the unified and expanded view of validity, prevalent since the latter part of the 20th century, with use and consequences in focus (Cronbach, 1971; Messick, 1989; 1996) and with a clear connection to, and emphasis on, values and ethics (cf. Kunnan, 2004; Roever & McNamara, 2006; Shohamy, 2001). It needs to be emphasised that the basic questions are applicable in relation to general language proficiency as well as to single competences. Furthermore, the different aspects affect and interact with one another in different ways; the purpose influences the choice of methods, as does the construct; methods are likely to affect the outcome/results; agency, not least of test-takers, is likely to vary in different context, and uses and consequences of results are strongly related to the purpose as well as to the construct.

As for the assessment of oral language proficiency in the study reported on here, a few aspects of the questions are in focus, namely the issues of construct, methods, and

uses/consequences. Further, the approach taken is analytic, in the sense that production and interaction are not only considered as partly different competences but also rated separately. In addition, however, a holistic score is given, which obviously comprises elements of interaction as well as production, but which is not formally regulated in relation to the two analytic ratings. For further discussion of analytic and holistic rating, see for example Iwashita & Grove (2003) and Khabbazbashi & Galaczi (2020).

### *2.5 Previous reports and research on attained proficiency levels in SFLs*

The First European Survey on Language Competence (ESLC) (European Commission, 2012) focused on reading, listening and writing skills in the first foreign language and in the most frequently studied SFL. Even though ESLC did not include speaking and the design of the study has been discussed critically (Erickson & Lodeiro, 2012), we will mention it here for comparative purposes. The participants in the ESLC were 15-year-old students at the end of lower secondary school (International Standard Classification of Education 2, UNESCO, 2011) in 14 European countries. In Sweden, students were tested in English (the first foreign language) and Spanish (the SFL). The Swedish results differed greatly between the two tested languages. In English, a majority of the students' performances exceeded the expected proficiency level in all skills that were tested (B1.1.) and results were among the highest of the participating countries. In Spanish, by contrast, performances of about 80% of the participating students did not reach the expected A2 level. The results were similar across the three tested skills, but slightly higher in reading and slightly lower in writing (European Commission, 2012, p.51).

Aronsson (2020) conducted a comparative study on oral and written proficiency in Spanish as SFL with Swedish students aged 15-16 years ( $n = 30$ ). For the collection of the oral data, Aronsson adapted a national test design for a paired speaking test with peer-to-peer interaction. The CEFR results of Aronsson's study showed that in speaking only 20% were assessed as A2.1, the minimal expected level for the age group. The remaining interactions were all rated lower than A2.1. In writing, 23% were rated at the A2.1 level or at the A2.2 level. All remaining written texts were rated as lower than A2.1. Aronsson also found that, when comparing the CEFR rating and the national grading system, oral performances rated as A2.1 were awarded the two highest grades (A and B). Oral performances which received a pass grade (E) were predominantly rated as pre-A1 or A1 level.

In a report by members of the group developing the national tests of foreign languages (the NAFS group), Axelson, Hedenbratt, Perrotte & Sebestyén (2020) studied the grades on national assessment materials from the last semester of Year 9 in French, German and Spanish. The report is based on the grading done by the students' SFL teachers, which the teachers voluntarily reported back to the NAFS group. Grades from 4,803 students in 2019 were reported (1,034 students in French, 1,623 in German and 2,146 in Spanish). The assessment materials cover all four basic skills: speaking including interaction, writing, listening and reading. For speaking, Axelson et al., (2020) found that only a small proportion of these students had received a grade indicating that they had not reached the minimal expected proficiency level at the end of lower secondary school. In French 6% of all students had received such a grade (i.e., 'F'), in German 4% and in Spanish 8%. When commenting on the results, Axelson et al., (2020, p. 51) noted that the reported grades in Spanish were the most heterogeneous. The proportion of students

that had received a grade indicating that they had not reached the minimal expected proficiency level (F) was higher in writing than in speaking in all three languages.

In Finland, Härmälä, Huhtanen, Silverström, Hildén, Rautopuro & Puukko (2014) reported on attained proficiency levels in French and German at the end of Year 9 (age 15 years) after 152 hours of teaching. The report focuses on students in Swedish-speaking schools in Finland where 85% of the participating students had Swedish as their first language. All four basic skills were assessed. Importantly, students were provided with a written “frame” for the oral interaction tasks, i.e., written dialogues in the target language (French or German). Results for French ( $n = 210$ ) and German ( $n = 174$ ) showed that the students performed best in the monologic production task. In French, about half of the students’ performances were assessed at a higher level than expected (A1. – A1.2). The results in German were higher than in French. Nearly all students (96%) in German reached the higher of the two expected proficiency levels in all tasks (A1.2). In a brief comparison with results from the Finnish schools, the authors find that the overall results in French, but in particular in German, were higher in the Swedish schools for all skills (Härmälä et al., 2014, p. 197-199).

## 2.6 Rationale and research questions

The rationale for the present study is the current lack of knowledge concerning attained proficiency levels within second foreign language education in Sweden. We focus on oral proficiency which is a particularly under-researched skill. By using the same tasks and assessment procedures and by including all three major SFLs, French, German and Spanish, we can compare results across types of tasks and across SFLs.

The research questions are:

1. What are the overall attained levels of oral proficiency in French, German and Spanish at the end of Year 9?
2. Are there differences in attained levels of proficiency a) between languages and b) between oral production and oral interaction?
3. What do the participating students think about a) their familiarity with the tasks, b) their performances and c) the degree of difficulty of the tasks?

## 3. Method

### 3.1 The TAL project

The data analysed in the present study were collected in the project *Learning, Teaching and Assessment of Second Foreign Languages – an Alignment Study on Oral Language Proficiency in the Swedish School Context* (henceforth the TAL-project), financed by the Swedish Research Council.

### 3.2 Data collection procedures in the current study

The study was approved by the Ethics Review Board of Southern Sweden (approval number 2017/745).



### 3.2.1 Participating schools and students

Data collection took place in 15 lower-secondary schools across Sweden. The schools were randomly drawn from a larger cohort of schools ( $N = 416$ ) that had been identified by Statistics Sweden (SCB) using a stratified random sampling procedure (see, e.g., Granfeldt et al., 2019 for details of the original sampling). The 15 schools in the present sample were drawn based on three parameters: language (French/German/Spanish), region (South/Mid-country/North) and (type of) organiser (municipality/independent). There were five schools per target language (French, German and Spanish).

The school leaders of the 15 schools were contacted via an email in which we explained the purpose and the organisation of the study, and they were offered the possibility to participate. The school leaders were asked to consult the teacher(s) of SFLs before replying. Two schools declined participation. They were replaced by two randomly drawn schools with the same characteristics, and both accepted to participate. Finally, using available statistics from the National Agency for Education (NAE), the sample of schools was found to reasonably reflect the variation in the larger cohort regarding demographic and socio-economic characteristics.

Once the schools had accepted to participate, one or two researchers from the project visited each school twice (except for one school where only one visit could be organised for practical reasons). During the first visit, the researcher(s) met with the SFLs teachers at the school and in particular with the teacher(s) of the target SFL at each school. On this occasion, the researcher(s) also met with all students in the targeted SFL class(es) together with their SFL teachers. The aim of this meeting was to make the project and the researchers more visible to the students and to offer them the possibility to participate. It was stressed that participation in the study was voluntary. The students who volunteered (all 15 years of age) signed an informed consent form.

In total 122 students participated in the oral part of the study (73 girls, 47 boys, 2 did not wish to state gender, see Table 4). One student (Spanish) did not complete the interaction task. Once the participating students were known, the SFL teachers were asked to form homogenous pairs for an interaction task (see below). The pairs were formed based on the teachers' perception of the similarity of students' language proficiency.

**Table 4.** Participating students, gender and second foreign language

	French	German	Spanish	Total
Girls	36	17	20	73
Boys	15	17	15	47
D N W S	0	2	0	2
Total	51	36	35	122

Legend: D N W S = Did Not Wish to State

### 3.2.2 Speaking tasks

During the researchers' second visit to the respective schools, the students completed two speaking tasks, first an individual production task and then an interaction task. Both

tasks had been developed in accordance with the national syllabus for Modern language 2 (NAE, 2018)<sup>2</sup>. The students performed the tasks in a separate and quiet room in the school. Prior to performing the tasks, the students saw an instruction video in the target language explaining the two tasks. A researcher was present in the room to answer questions and, if needed, to provide additional information in Swedish. The students were given a few minutes to reflect on the task before starting. Notetaking was not allowed. The students' performances were video recorded and a separate audio recording on a laptop was made.

### 3.2.3 Production task

The production task was an individual speaking task. The scenario for the task was an upcoming school visit from a class of students in a target language country (France, Spain or Germany). A description of the task can be found in Appendix A. After performing the task, the students answered a questionnaire about the task and their experience of taking part in the study and how they perceived their performance.

### 3.2.4 Interaction task

The interaction task was a paired speaking task with an information gap feature. The scenario for the task expanded on the school visit from the production task. The interaction task was to plan a week of activities for the visiting students. A description of the interaction task can be found in Appendix A. After performing the task, the students answered a questionnaire about the task and their experience of taking part in the study and how they perceived their performance.

## 3.3 Rating procedures

The rating of the students' performances in the oral tasks was based on a set of scales taken from the CEFR and Companion Volume (Council of Europe 2001; 2020) and comprised several constellations of raters as well as a number of procedural steps.

### 3.3.1 Scales

The scales chosen for the rating of oral language proficiency were firstly the global scale provided in the CEFR, in which the basic language competences in the document are described in single holistic paragraphs. Further, a number of subscales focusing on oral functions were added. For Production this meant *Overall oral production* (CEFR, p. 58), *Sustained monologue: Describing experience* (p. 59), and *Addressing audiences* (p. 60). The corresponding illustrative scales for Interaction were given for *Overall spoken interaction* (p. 74), *Conversation* (p. 76), *Informal discussion (with friends)* (p. 77), and *Goal-oriented co-operation* (p. 79). In addition, aimed to support the overall rating required, table 3 in the Companion Volume to the CEFR (cf. Appendix C) was added, focusing on *Qualitative features of spoken language (expanded with phonology)*. In this, the following features are defined: *Range, Accuracy, Fluency, Interaction, Coherence, and Phonology*, thus with interaction described as one aspect of the overall, oral proficiency. In all cases, the neighbouring levels to A2 (sometimes labelled A2.1; cf. Table 2 above), namely A1 and B1 (sometimes B1.1) were given to strengthen and, to some extent, facilitate the rating.

### 3.3.2 Raters and ratings

As shown in Table 5, all in all, twenty-one different raters took part in the four rounds of rating conducted in the study. These were the six researchers themselves (two for each target language), two colleagues of the researchers per language, and nine external raters (three for each language), three of the latter participating in two different rounds of rating.

**Table 5.** Raters and rounds of rating

Round of rating	Rated performances (prod+inter/student) <i>French, German, Spanish</i>	Raters	Comments
1	244	6 researchers 2 <i>French</i> , 2 <i>German</i> , 2 <i>Spanish</i>	Two individual rounds of rating incl. intermediate discussion
2	60	6 colleagues 2 <i>French</i> , 2 <i>German</i> , 2 <i>Spanish</i>	Identification of student performances for the external round (3)
3	60	9 External raters 3 <i>French</i> , 3 <i>German</i> , 3 <i>Spanish</i>	CEFR-experienced Raters
4	60	3 External raters (from round 3) 1 <i>French</i> , 1 <i>German</i> , 1 <i>Spanish</i>	Raters with ample large-scale assessment experience. Additional performances (from round 1)

In the following, brief descriptions are first given of the raters and then of the different rounds of rating that were undertaken.

The six **researchers** in the *TAL* project, five women and one man, had somewhat different academic profiles: two with a background in French linguistics, one in Italian, one in Spanish, and another in psycholinguistics with some emphasis on German. Further, two represented the field of language education, one with a special focus on issues of language assessment. Four of the six researchers had a teacher education degree and experiences of longer or shorter periods of teaching in schools. Researchers worked on the languages in pairs, based on the researchers' language profiles.

The **colleagues** of the researchers were all experienced language educators with different specializations, not least in teaching and assessment, including solid acquaintance with the CEFR.

Originally, only one **external rating** session was planned, for which three raters per language were recruited. Raters were chosen based on the researchers' knowledge and experience of the CEFR in different educational contexts in Sweden, in which assessment is a regular activity. The final candidates were asked to confirm their familiarity with the CEFR and with assessment and rating. The group consisted of five women and four men

working in different educational contexts and with different focal points regarding learning, teaching and assessment, however all of them with ample and relevant experience of the CEFR.

Towards the end of the project, it was decided to conduct a final round of rating, on a **new set** of student performances from the original sample (see further below). In this round, **three of the external raters** took part (two women and one man), one per language, and all of them professionally active in large-scale assessment, including oral proficiency.

All raters in all four rounds of rating rated the students' performances in the two tasks separately, starting with the production task followed by the interaction task. Finally, they provided an overall assessment of the students' oral proficiency based on both the production task and the interaction. All raters indicated for each of the three ratings (production, interaction and overall) if a particular student performance was:

- Below the A2.1 level
- On the A2.1 level
- Above the A2.1 level

In the first round of rating, the six researchers, working in three pairs, one for each language, initially rated all student performances individually for the three aspects mentioned. They then met to discuss, compare and document their ratings, first in pairs, later also in the whole group. This resulted in a second individual rating, used for analyses in the current study. In addition, each group discussed and decided on a number of student performances to be used in the external rating phase. The final inter-rater agreement among the researchers was moderate (Spanish  $\alpha = .656$ ) to high (French  $\alpha = .894$  and German  $\alpha = .900$ ). For more details, see Appendix B.

Secondly, the examples tentatively chosen by the researchers were analysed and commented on by two experienced colleagues of the researchers in each target language. The result of this procedure was a set of 20 performances per language, representing three levels of proficiency, namely the targeted level A2.1, and examples below and above this level. For these examples, there was total agreement regarding levels between the two researchers and their colleagues.

The external rating of 20 student performances in each language by three raters per language (round three of the rating process) was conducted during a full day and was preceded by individual preparation based on materials distributed prior to the meeting. In addition, during the first hours of the day, information was given, scales analysed, and some concrete rating conducted and discussed (cf. familiarization as recommended by the Council of Europe, 2009). The actual rating was done on individual computers with earphones. The day was concluded with joint sessions in the groups, in which a number of questions regarding processes and products were discussed. The inter-rater agreement among the external raters was high (French  $\alpha = .834$  and German  $\alpha = .834$ , Spanish  $\alpha = .867$ ). For more details, see Appendix B. Shortly after the rating took place, the external raters were sent an electronic survey and asked to give their feedback on various aspects of the procedure. Their general impressions of the session and of the procedure were positive.

A fourth and final round was conducted of 20 performances per language chosen from the first round of rating where in some of them, there was a certain disagreement between the two researchers, in others total agreement. One external rater per language participated. This round was undertaken to further validate the results and to enhance the understanding and possible interpretation of the data.

### 3.3.3 Overview of ratings for analysis

As explained above, the students' performances were rated in successive rounds and by different raters. Since the number of raters varied, we use *predominant rating* in order to define a performance level for each student performance. When all raters agreed, the predominant rating is the rating about which there was agreement between raters. In cases where raters did not agree, the rating that most raters chose is the predominant rating. If, for example, two raters rated a student's performance in production as reflecting the A2.1 level and a third rater of the same performance rated it as below A2.1, the predominant rating for this performance is A2.1. We use the term 'predominant rating' as a reminder that assessment of language proficiency is normally associated with variability between raters. Table 6 below presents the 344 performances where a predominant rating could be identified. These are the ratings that were used in the analyses that will be presented in the following sections. The table also presents the few performances ( $n = 20$ ) where no predominant rating could be identified due to an even number of raters. These cases will not be discussed further.

**Table 6.** Number of ratings per performance and per language

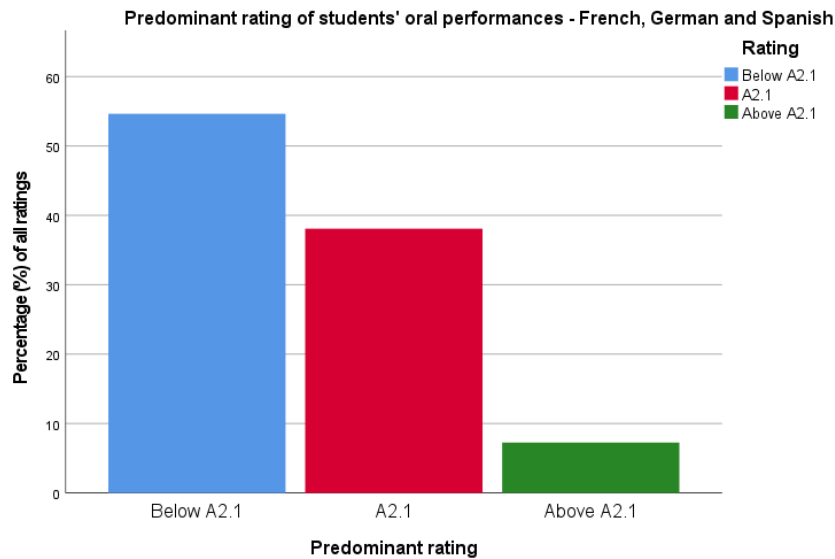
	Production	Interaction	Overall proficiency	Missing ratings	No predominant rating identified	Total
French	49	51	51	-	2	154
German	32	36	34	-	6	104
Spanish	29	32	32	2	12	105
Total	110	119	117	2	20	366

## 4. Results

### 4.1 Ratings of students' performances

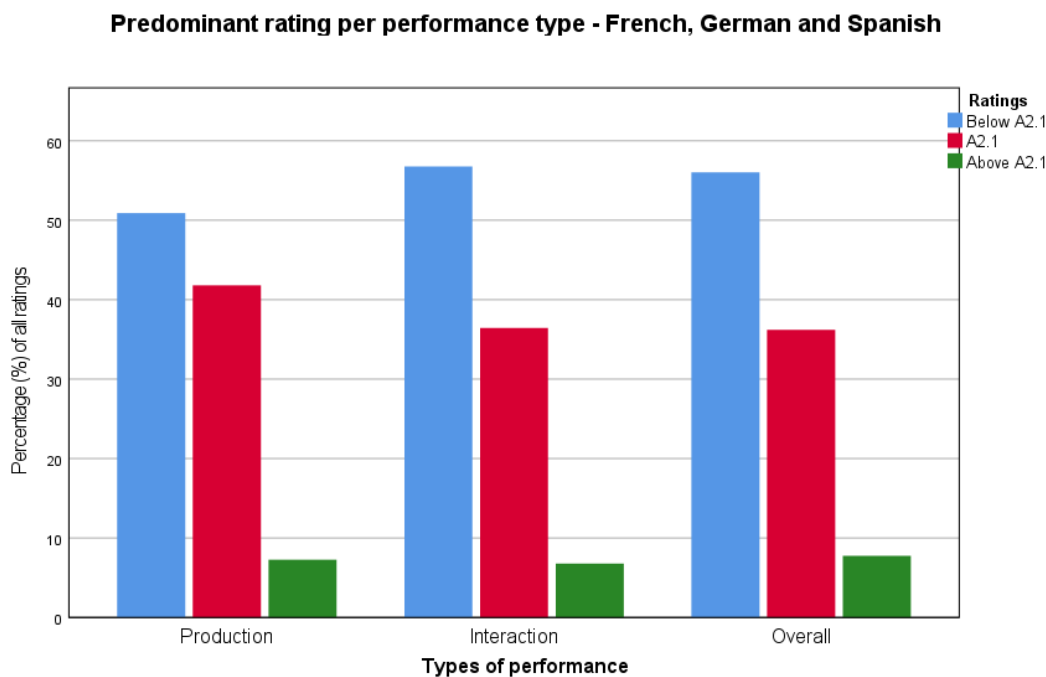
In the following, we will first present the quantitative results from the rating of the students' oral performances and then the students' feedback. The results are based on the predominant rating when all raters are included, as explained in the preceding section.

Figure 1 below shows the predominant rating of students' oral performances in the three SFLs (French, German and Spanish) taken together. In this figure, also all three performance types (production, interaction and overall) are collapsed ( $n = 344$ ). Cases where a predominant rating could not be established were excluded ( $n = 20$ , see Table 6 above). Two ratings were missing from the data set. Figure 1 shows that the rating Below A2.1 was the most frequent rating, representing 55% of the total number of ratings in the data set, followed by A2.1 (38%) and Above A2.1 (7%).



**Figure 1.** Predominant rating of students' performances (all SFLs, all performances)

Next, we will consider the same data according to different performances. In this analysis, the predominant ratings of the three performances (production, interaction and overall) are presented separately.



**Figure 2.** Predominant rating per type of performance (all SFLs)

Figure 2 shows that the distribution of ratings is similar across the three performances with the rating Below A2.1 being slightly more frequent in interaction (57%) and overall (56%) as compared to production (51%). Conversely, the rating A2.1 is slightly more

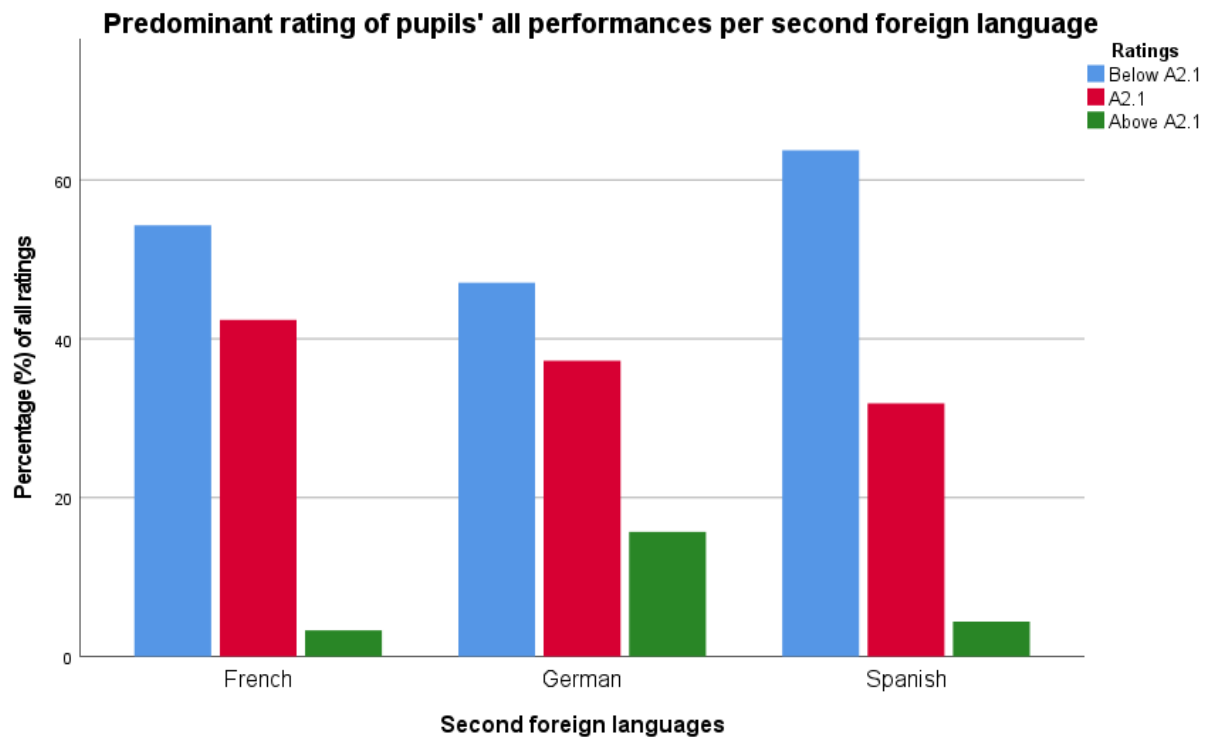
frequent in production (42%) compared to interaction (36%) and overall proficiency (36%). However, according to a Kruskal-Wallis H test there are no significant differences between the ratings of the three performances  $\chi^2(2) = .776, p = .679$ . The similarity of ratings across the three performances (production, interaction and overall) was confirmed by a series of Spearman rank order correlations between the ratings. The results are presented in Table 7.

**Table 7.** Correlation matrix of ratings across students' performances (all SFLs)

		Production	Interaction	Overall proficiency
Spearman's rho	Production		.749**	.800**
	Interaction	.749**		.944**
	Overall	.800	.944**	

\*\*  $p < .000$

Next, we will consider the rating of each of the three SFLs separately. Figure 3 below shows the distribution of ratings of students' performances in French, German and Spanish respectively.

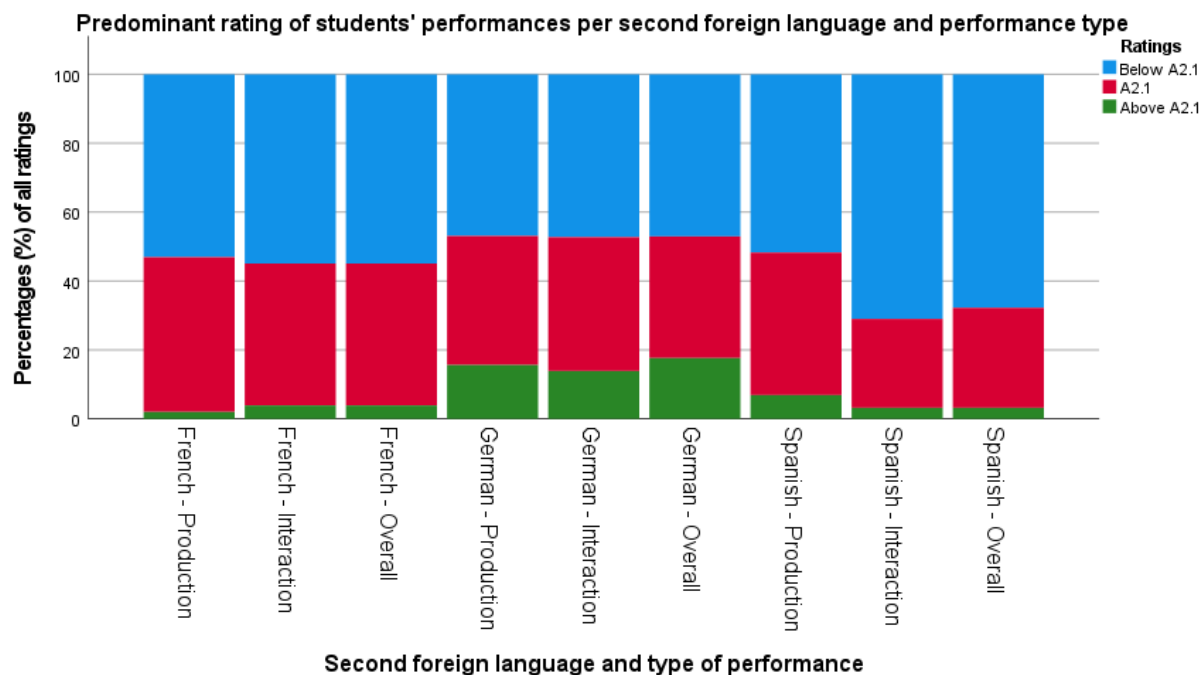


**Figure 3.** Predominant rating of students' performances per second foreign language (all performances)

The highest proportion of Below A2.1 ratings was found in Spanish (64%), followed by French (54%) and German (47%). The highest proportion of A2.1 ratings were found in French (42%) followed by German (37%) and Spanish (32%). With respect to ratings Above A2.1, German stands out with a proportion of 16%. When all ratings of students' performances in each of the three target languages were compared, a Kruskal-Wallis H

test showed that there was a statistically significant difference between the ratings,  $\chi^2(2) = 8.028$ ,  $p = 0.018$ . A series of Dunn's pairwise tests showed a significant difference between the ratings of students' performances in German and Spanish ( $p = .016$ ) whereas the difference between ratings of performances in German and French was approaching significance ( $p = .060$ ). The difference between the ratings of performances in French and Spanish was not significant ( $p = .224$ ).

Last, we will consider the ratings of the respective performances for the three SFLs. The results are shown in Figure 4.



**Figure 4.** Predominant rating of students' performances per second foreign language and performance type

Figure 4 shows that for both French and German, the rating of the three performances is largely similar. In French proportions of Below A2.1 ratings vary between 53% (production) and 55% (overall and interaction). The same figures for German are 47% (production) and 47% (interaction). However, in Spanish the differences are more pronounced, and proportions of Below A2.1 ratings vary from 52% (production) to 71% (interaction). In other words, there seems to be more variability in the rating of Spanish as compared to French and German. While descriptively there are some differences between the ratings of the different performance types across SFLs, a series of Kruskal-Wallis H test showed, however, that there was no statistically significant differences between the SFLs with respect to the ratings of production, ( $\chi^2(2) = 1.168$ ,  $p = .569$ ), interaction ( $\chi^2(2) = 4.622$ ,  $p = .099$ ) or overall ( $\chi^2(2) = 2.863$ ,  $p = .133$ ). Since significant differences between the SFLs were found when the ratings of all three performances were collapsed (cf. above), the absence of statistically significant differences here may be assumed to be due to the relatively small sample size.



#### 4.2 A selection of results on students' feedback

In this section we will present some of the feedback that students provided after completing each of the two tasks. We will focus here on results from items concerning the familiarity with the types of tasks, the students' self-perceived satisfaction with their performances and the perceived comparative difficulty of the two tasks.

Starting with the familiarity with the types of tasks, two items targeted this question. Results are presented in Figures 5 and 6 per second foreign language.

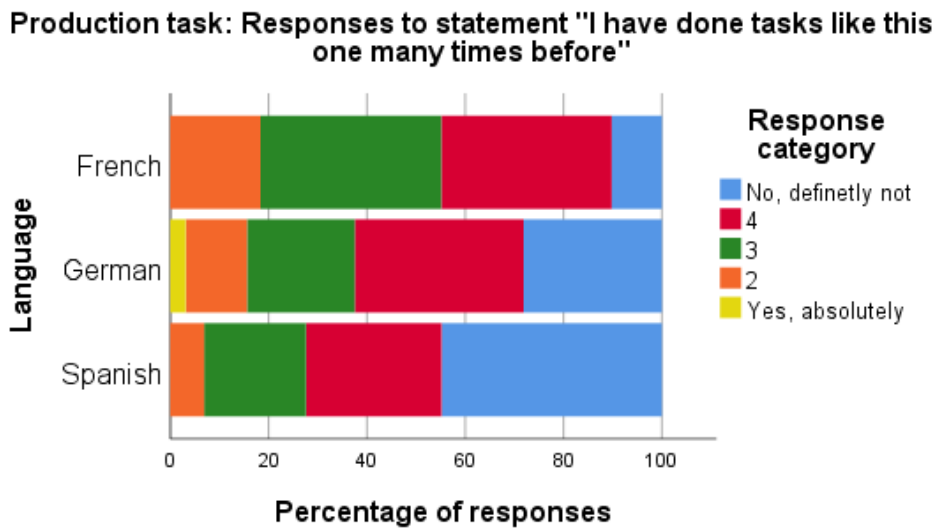


Figure 5. Students' responses to familiarity with production task

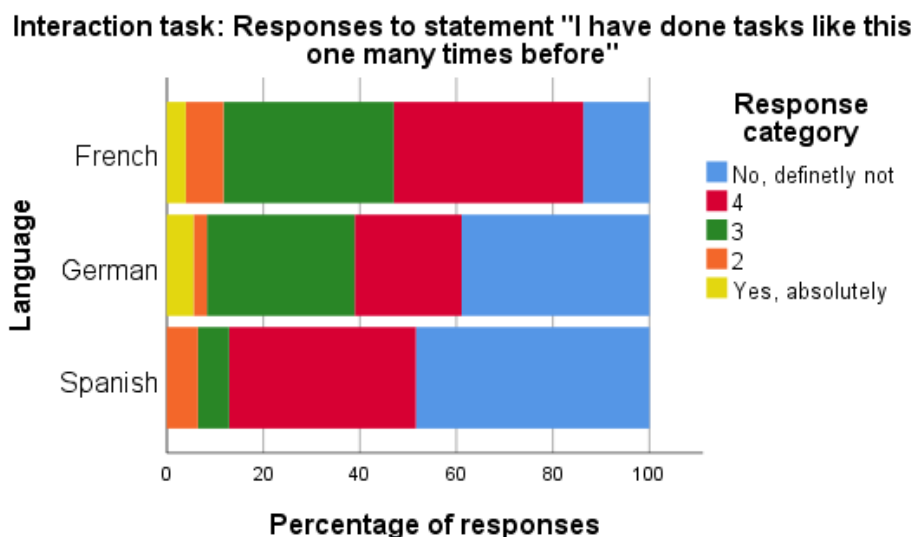


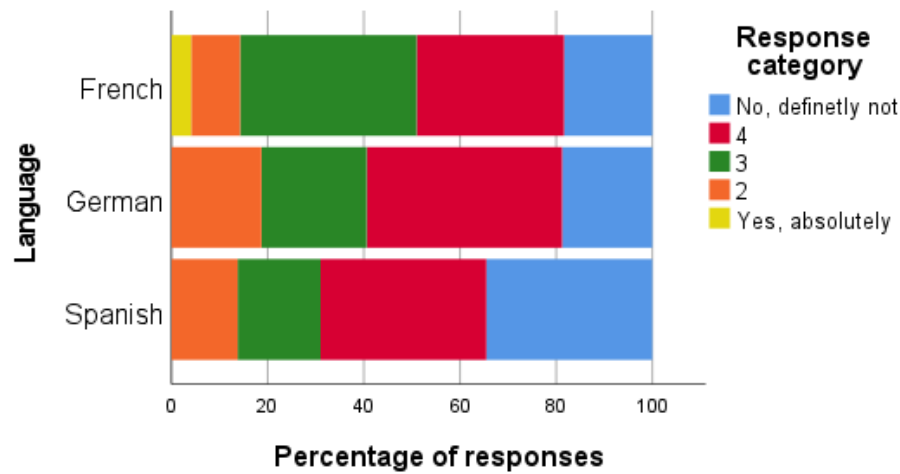
Figure 6. Students' responses to familiarity with interaction task

To begin with, a majority of students expressed that they had not often done tasks like the production and the interaction task before. Response categories 4 and 5 ('definitely not') are dominant in all languages and for both tasks. Second, there is a difference

between on the one hand Spanish, and French and German on the other hand. A larger proportion of students in Spanish express that they have definitely not performed tasks like these ones before.

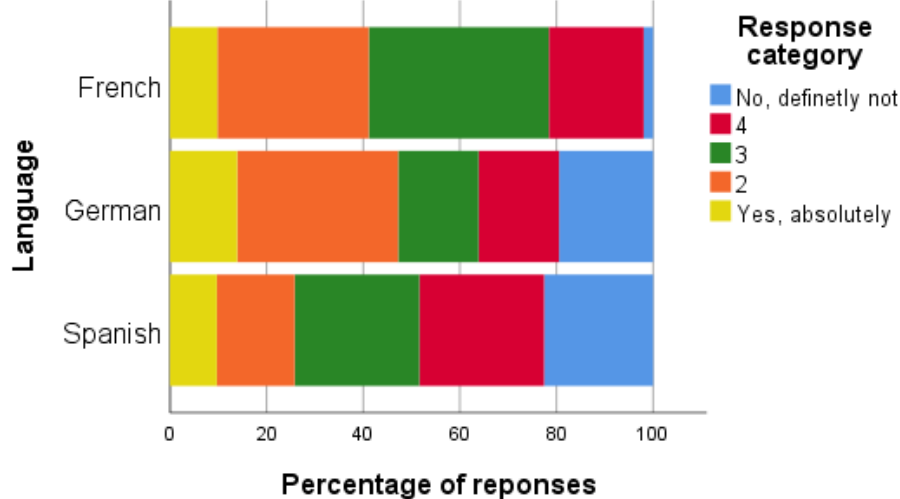
Moving to the self-perceived satisfaction with their performance (Figures 7 and 8), students responded slightly differently with respect to the two tasks. A larger proportion of students in all three languages agreed more that they performed well on the interaction task when compared to the production task. Students in French were most pleased with their performances, followed by the students in German. Consequently, the students in Spanish were the least pleased with their performances in the two tasks.

**Production task: Responses to statement "I think I performed well on this task"**



**Figure 7.** Students' responses to self-perceived satisfaction with performance on production task

**Interaction task: Responses to statement "I think I performed well on this task"**



**Figure 8.** Students' responses to self-perceived satisfaction with performance on interaction task

When it comes to the perceived difficulty of the two tasks, results point in the same direction (Figures 9 and 10). All students found that the production task (speaking alone) was much more difficult than the interaction task (speaking with a friend). Responses are very similar in the three languages.

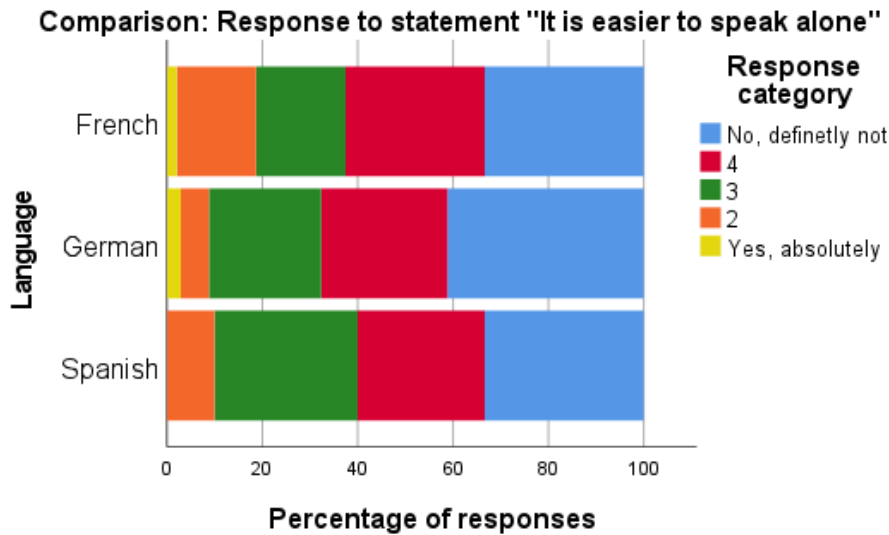


Figure 9. Students' responses to perceived difficulty of production task

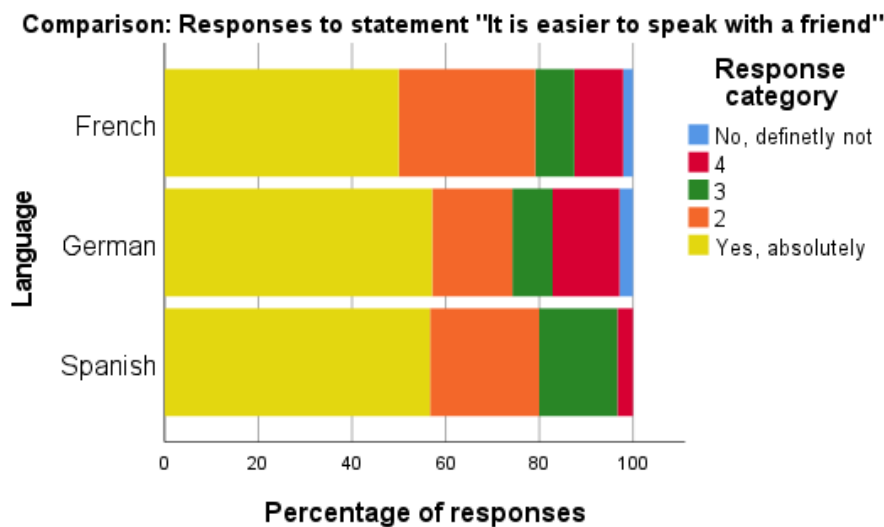


Figure 10. Students' responses to perceived difficulty of interaction task

## 5. Summary and discussion

We will start by summarising the results in relation to the three research questions. With respect to the first question, we found that slightly less than half of the student performances were rated as A2.1 or above. In other words, just under half of the performances were predominantly rated as being on or above the expected minimal proficiency level for SFLs at the end of Year 9. However, in relation to the second

research question we found that there were differences between the languages. A higher proportion of performances in German were rated as A2.1 or above than in French and Spanish. The students performed two tasks, one production task and one interaction tasks, and raters rated three performances (production, interaction and overall). However, the results showed that there were no significant differences in the ratings of the different performances. Whether this indicates a reasonably unidimensional construct or difficulties in assessing the two modalities is something that needs to be studied further. The last research question concerned the students' reported familiarity with the tasks, their satisfaction with their performances and the perceived difficulty of the tasks. The students reported that they were not very familiar with any of the two tasks, in particular the students of Spanish. Moreover, the students were more satisfied with their performances in the interaction task which they also generally found to be easier than the production task. The students in Spanish reported being the least pleased with their performances.

When turning to a discussion of these results, it is important to stress that the results need to be interpreted with caution. First, the sample consisted of small number of self-selected SFL students who are not necessarily representative of the group of SFL students in Swedish lower-secondary school. Second, student reports and researchers' observations on site indicate that a number of students felt uneasy about the data collection situation. Many students were not comfortable being recorded speaking their SFL in the presence of a researcher that they did not know very well. This situation could have affected their performances negatively. Third, it needs to be kept in mind that the recorded performances represent a one-off snapshot of the students' oral proficiency on a certain day at a given time. The recordings are not necessarily representative of the participating students' overall oral proficiency. Fourth, the rating of the performances showed that while the inter-rater reliability was quite high in French and German among all groups of raters, the analyses showed more variable results for Spanish in the rating done by the researchers.

While keeping the caveats above in mind, the result indicating that less than half of the students' performances reached the expected proficiency level A2.1 is noteworthy and needs further attention and research. When discussing these results, we start by recalling the results reported in the ESLC (European Commission, 2012) and in Aronsson (2020). In both these studies only Spanish was included, and they both pointed towards around 80% of the performances in writing (ESLC) and in speaking (Aronsson, 2020) not reaching the expected A2.1 level in Spanish. In the present study, the percentage of performances not reaching the A2.1 level ranges from 47% in German to 64% in Spanish. However, the studies are difficult to compare because a) the investigated competence is different (ESLC did not include speaking) and b) the rating procedures differ. Hence, no firm conclusions can or should be drawn when comparing the different studies.

Moreover, the results here can only partially be compared to those in the report by Axelson et al. (2020). These authors found that in the teacher reported grades from the voluntary national tests for SFLs, only very small proportions of students (4-8%) did not receive a pass grade in Year 9. This result would imply that the vast majority would have reached the A2.1 level in stark contrast to the findings in the present study. However, the differences with respect to types of data need to be underlined. Axelson et al.'s (2020) study is based on data from schools who voluntarily reported their test results. Also, it needs to be remembered that in this case the Swedish national syllabus was used, not

the CEFR. True, a certain degree of alignment between these two documents has been undertaken, albeit only partially and focusing of semantics rather than empirical data. Nevertheless, the differences in outcomes are striking. Without going further into details here, the discrepancy between existing studies together with the paucity of studies in general underline again the need for further research to understand and, if possible, ultimately affect the attained levels of proficiency among students of SFLs in Sweden.

Rather than discussing the attained levels of proficiency *per se*, it seems to be more fruitful to reflect upon some of the learning conditions that might be leading to these results. We note that a large proportion of students report not being used to performing tasks like the production and interaction tasks used in the present study. While it needs to be kept in mind that these results are based on students' reports rather than documentation of teaching activities, they coincide with some other observations of learning conditions in the SFL classroom. A recent report from The Swedish Schools Inspectorate (2022) based on the observation of Modern language teaching in 34 lower-secondary schools also points to the low degree of target language use and finds that in two out of three schools there is a need to develop the use of the target language in the classroom. According to the report, speaking is primarily used for a restricted number of routine-based pragmatic functions (e.g., greetings and leave-taking). Another observation is that students often prepare their speaking by answering textbook questions in writing beforehand (p.19). This picture is confirmed in a study by Aronsson (2023) of Spanish. In Aronsson's study speaking activities represented only 9% of the classroom activities, while writing (40%) and reading (37%) were found to be the two most frequent activities.

In the *TAL* project, we have previously studied the reported use of target language in the classroom. Erickson et al. (2022) found that the teachers responding to the extensive questionnaire used in the *TAL* project ( $n = 315$ ), reported that Swedish was used to a large extent as the medium of instruction in their classrooms, more so with younger students but also towards the end of compulsory school. Estimations of the proportion of target language use in the classroom over the whole time of instruction from year 6 to 9 showed that less than 30 percent stated more than half of the time, with only six percent estimating the amount of target language use to 75 percent or more. Several reasons were given, focusing in particular on students' limited comprehension of the target language and their first language as a prerequisite for understanding and learning. In this, no comments of the fact that more than 20 percent of students in compulsory school have another L1 than Swedish were made.

The issue of target language use can be discussed from a number of angles, ranging from tradition to beliefs about language learning and attitudes to inclusion. However, if the amount of time the teachers use the target language in the classroom is (very) low, it is difficult to expect students to feel at ease when performing production and interaction tasks in this language, or indeed, in real-life situations.

Continuing on this point, students in all three SFLs reported finding interaction easier than production and students were clearly happier with their performances in interaction than in production. However, these perceptions were not reflected in higher ratings of interaction compared to production. In fact, descriptively, production was rated higher than interaction even though the difference was not statistically significant. Similar results were found in Härmälä et al. (2014) on Finnish students in Swedish

speaking schools. The students' perception of the production task in the current study could possibly reflect a feeling of unease when speaking by themselves in the SFL, again pointing to the suggestion that they are not used to speaking the SFL. In such a situation, it is easy to understand that support from a peer is considered helpful. In addition, when focusing on the two oral performances, it needs to be pointed out that interaction, being a co-constructed activity, is complex and far from easy to assess, with performances intertwined and partly interdependent. Moreover, scales, standards and criteria do not always provide enough support for raters to distinguish between the two and rate them as two separate, but still connected, skills (Borger, 2018; May, 2011).

Another learning condition which seems relevant here is the typological distance between the first language of most of the students, namely Swedish, and the respective SFLs. The results suggest that a higher proportion of students of German reach the expected proficiency level A2.1 at the end of lower secondary school than students of French and Spanish. The same result was reported by Härmälä et al. (2014) from the Swedish speaking schools in Finland. German and Swedish are Germanic languages and have a high degree of lexical overlap (Bohnacker et al., 2016). Similarity between previously acquired languages and the foreign or second language being acquired has been argued to be an important factor (see e.g., Ringbom, 2007). Moreover, the first foreign language of the students in the present study is English which might additionally positively influence the acquisition of aspects of German. Indirect evidence for a positive typological proximity effect can also be found in the study by Härmälä et al. (2014) where the comparison between the results from the Swedish speaking schools and the Finnish speaking schools suggests an advantage for the former. The students with Finnish as their first language performed worse in both French and German. In Sweden, all SFLs (with the exception of Chinese) share the same syllabus and students are expected to reach the same levels of proficiency.

Yet another type of learning conditions is relevant when discussing the results in Spanish. Compared to French and German, the performances in Spanish were rated lowest. The same finding was reported by Axelson et al. (2020) who focused on reported grades from national assessment materials, even though the differences between French, German and Spanish were smaller in their study. Several tentative explanations for this result can be advanced. First, as Spanish is by far the most studied SFL in compulsory school with more students than French and German together, there will be possibly more variability among Spanish students' performances than in the other two SFLs. Also, groups are usually much larger as compared to the other two languages, which is an additional factor to be considered, for pedagogical as well as equity-related reasons (Finndahl, 2023). Second, Spanish is the SFL with the highest proportion of uncertified teachers (NAE, 2020) and a previous report has pointed to certified teachers as a success factor for SFL teaching (NAE, 2013; Riis & Francia, 2013). It is also the most recently introduced SFL among the three SFLs in this study, even though this was in 1994, and the problem of teacher recruitment remains (Österberg, 2021). Third, the students in French receive, on an average, higher grades than the students in Spanish and German (Bardel et al., 2019) and French has been considered to be the most prestigious of the three SFLs (Tholin & Lindqvist, 2009, Krigh, 2019). The number of factors illustrates just how complex the situation is for teaching, learning and assessment of SFLs and underlines again the need for more large-scale research in this domain.

## 6. Conclusion

In Sweden, there is a paucity of research focusing on students' attained oral proficiency with the consequence that we know very little about what learners actually can do when speaking their second foreign language after a certain number of years of study in school. Therefore, the present study, which is part of a larger project focusing on second foreign languages in lower secondary school, studied attained levels of proficiency in 122 students of French, German and Spanish on the basis of one productive and one interactive task. The results showed that just under half of the students' oral performances were assessed at or above the intended CEFR-level at the end of year 9 (A2.1). The remaining performances were all assessed as lower than the expected level. While the patterns of low performances were generally the same across languages, we found that significantly more student performances in German were rated at or above the expected level of proficiency than in French and Spanish. With respect to the two different tasks (production and interaction), we did not find any significant differences in rated proficiency levels, but a clear majority of the students reported that interaction was the easier task and the task where they were more pleased with their own performance.

Even though there are differences in study designs, these results align with the few previous empirical investigations of attained levels of proficiency in Spanish. More importantly they underline the need to further discuss the learning conditions for students to develop oral proficiency through the teaching of second foreign languages. One important aspect is the use of the target language in the classroom by teachers which has been found to be low in previous studies. Another aspect is the unease that the students in our study reported feeling when they were asked to produce or interact in French, German or Spanish without the support of any written material. This unease can be indicative of the fact that students have few opportunities to practice their oral proficiency more freely in class which is also reflected in the students' feedback on the tasks used in the present study.

Finally, it needs to be remembered that far from all students in Swedish compulsory school choose to study a second foreign language. Thus, the group that actually does is clearly self-selected in various ways, for example regarding gender and socio-economic background. There is obviously no way of knowing what the results reported on would have been like, had SFL been a compulsory subject, albeit it seems unlikely that they would have been affected in a more positive way. On the whole, this emphasizes the need for further research of teachers' and policymakers' beliefs regarding the role of and need for additional languages, teachers' approaches to learning and teaching a new language, and different students' rational for choosing the way they do. These are issues of clear interest to language education but also connected to aspects of equity and equality of schooling at large.

### Funding

We gratefully acknowledge funding from the Swedish Research Council, grant number 2015-01088 to J. Granfeldt.

## Endnotes

<sup>1</sup> There are currently two courses of Modern languages in compulsory school: a long course (320 h) running from year 6 to 9, and a short course of an additional modern language (177 h) offered for years 8-9. The NAE has suggested that the latter, referred to as the student's choice and currently taken by < 2 % of all students, be abolished.

<sup>2</sup> References in the text are made to the valid national syllabuses at the time of the study. The latest revision (2022) has not changed the action-oriented approach expressed in earlier regulatory documents.

## Disclosure statement

The author declared no conflict of interest.

## Author contribution statement

**Granfeldt:** Writing - Original Draft, Formal Analysis, **Erickson:** Writing - Original Draft, Methodology, **Bardel:** Writing - Review & Editing, Task Development **Sayehli:** Writing - Review & Editing, **Ågren:** Writing - Review & Editing, Task Development **Österberg:** Writing - Review & Editing, Task Development **All authors:** Conceptualization, Data Collection, Rating

## References

- Aronsson, B. (2020). A study of learner profiles in Spanish as a second language in a Swedish instructional setting: Writing versus speaking. *Journal of Linguistics and Language teaching*, 11(1): 69-90. <https://doi.org/10.1177/13621688221144282>
- Aronsson, B. (2023). Language learning activities in the Spanish L2 classroom related to a task-based framework: What types are the most commonly occurring according to Swedish learners? *Language Teaching Research*. Online first view. <https://doi.org/10.1177/13621688221144282>
- Axelson, P., Hedenbratt, L., Perrotte, F. & Sebestyén, Å. (2020). *Resultatrapport – Bedömningsstöd från 2019. Franska, spanska och tyska, språkval åk 9/moderna språk 2*. [Report - The assessment study in 2019. French, Spanish and German as a language choice in year 9/Modern language level 2] Institutionen för pedagogik och specialpedagogik Göteborgs universitet. Retrieved on 2022-02-21 from [https://www.gu.se/sites/default/files/2020-12/Resultatrapport%20E2%80%9320Bed%C3%B6mningsst%C3%B6d%20fr%C3%A5n%202019\\_0.pdf](https://www.gu.se/sites/default/files/2020-12/Resultatrapport%20E2%80%9320Bed%C3%B6mningsst%C3%B6d%20fr%C3%A5n%202019_0.pdf)
- Bachman, L. & Palmer, A. (1996). *Language Testing in Practice*. Oxford University Press.
- Bardel, C., Erickson, G., & Österberg, R. (2019). Learning, teaching and assessment of second foreign languages in Swedish lower secondary school-dilemmas and prospects. *Apples: Journal of Applied Language Studies*, 13(1): 7-26. <https://apples.journal.fi/article/view/97936/55949>
- Bohnacker, U., Lindgren, J. & Öztekin, B. (2016). Turkish- and German-speaking bilingual 4-to 6-year-olds living in Sweden: Effects of age, SES and home language input on vocabulary production. *Journal of Home Language Research* 1, 17-40. <https://doi.org/10.16993/jhrlr.26>
- Borger, L. (2018). *Investigating and validating spoken interactional competence: Rater perspectives on a Swedish national test of English*. [Doctoral thesis, University of Gothenburg]. Gothenburg University Publications Electronic Archive. <https://gupea.ub.gu.se/handle/2077/57946>.



- Byram, M. (2021). *Teaching and Assessing Intercultural Communicative Competence. Revisited* (2<sup>nd</sup> ed.). Multilingual Matters.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47. <https://doi.org/10.1093/applin/1.1.1>
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press. <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. A Manual*. <https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion Volume*. <https://www.coe.int/en/web/education/-/common-european-framework-of-reference-for-languages-learning-teaching-assessment-companion-volume>
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (Second edition, pp. 443-507). American Council on Education.
- De Jong, N., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). Facets of Speaking Proficiency. *Studies in Second Language Acquisition*, 34(1): 5-34. <https://doi.org/10.1017/S0272263111000489>
- 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. (2020). National assessment of foreign languages in Sweden. University of Gothenburg, Gothenburg, Sweden. [https://www.gu.se/sites/default/files/2020-04/Nat Assesment of Foreign Lang in Swe2020.pdf](https://www.gu.se/sites/default/files/2020-04/Nat_Assesment_of_Foreign_Lang_in_Swe2020.pdf)
- Erickson, G., & Lodeiro, J. (2012). Bedömning av språklig kompetens – En studie av samstämmigheten mellan Internationella språkstudien 2011 och svenska styrdokument [Assessment of language competence: A study of the relationship between The European Survey on Language Competences 2011 and Swedish regulatory documents for languages] Skolverket. <https://www.skolverket.se/publikationer?id=2831>
- Erickson, G., & Pakula, H.M. (2017). Den gemensamma europeiska referensramen för språk: Lärande, undervisning, bedömning – ett nordiskt perspektiv [The Common European Framework of Reference for Languages: Learning, teaching, assessment – a Nordic perspective]. *Acta Didactica Norge*, 11(3). <http://dx.doi.org/10.5617/adno.4789>
- Erickson, G., Bardel, C., Österberg, R., & Rosén, M. (2022). Attitudes and ambiguities—teachers’ views on second foreign language education in Swedish compulsory school. In Bardel, C., Hedman, C., Rejman, K. & Zetterholm, E. (Eds.), *Exploring Language Education: Global and Local Perspectives* (pp.157-223). Stockholm University Press. <https://doi.org/10.16993/bbz>
- European Commission. (2012). *First European survey on language competencies*. Final report. Luxembourg: Publications Office of the European Union. <https://op.europa.eu/sv/publication-detail/-/publication/42ea89dc-373a-4d4f-aa27-9903852cd2e4>
- Fernández, S. S., & Andersen, H. L. (2019). Oral proficiency in second and third foreign languages in the Danish education system. *Apples: Journal of Applied Language Studies*, 13(1). <https://doi.org/10.17011/apples/urn.201903011689>
- Field, J. (2011). Cognitive validity. In M. Milanovic & C. Weir (Eds.), *Examining Speaking. Research and practice in assessing second language speaking* (pp. 65-111). Cambridge University Press.
- Finndahl, I. (2023). *Young students’ Language Choice in Swedish compulsory school – expectations, learning and assessment*. [Doctoral thesis, University of Gothenburg].

- Gothenburg University Publications Electronic Archive.  
<http://hdl.handle.net/2077/74850>
- Fulcher, G. (2003). *Testing second language speaking*. Longman.
- Granfeldt, J., Sayehli, S., & Ågren, M. (2019). The context of second foreign languages in Swedish secondary schools: Results of a questionnaire to school leaders. *Apples: Journal of Applied Language Studies*, 13(1). <https://doi.org/10.17011/apples/urn.201903011688>
- Granfeldt, J., Sayehli, S., & Ågren, M. (2021). Trends in the study of Modern languages in Swedish lower secondary school (2000–2018) and the impact of grade point average enhancement credits. *Education Inquiry*, 12(2), 127-146. <https://apples.journal.fi/article/view/97937/55950>
- Hellqvist, B. & Thue Vold, E. (2022). *GERS i korta drag. En lärarguide för Gemensam europeisk referensram för språk [The CEFR in short. A teachers' guide to the Common European Framework for Languages]*. Fortbildningsavdelningen för skolans internationalisering, Uppsala universitet. <https://www.edu.uu.se/fba/resurser/>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Penguin.
- Härmälä, M., Huhtanen, M., Silverström, C., Hildén, R., Rautopuro, J., & Puukko, M. (2014). *Inlärningsresultaten i främmande språk i de svenskspråkiga skolorna 2013: A-läro kursen i engelska samt B-läro kurserna i franska, tyska och ryska*. [The learning results in foreign languages in the Swedish speaking schools 2013: The A-course in English and the B-courses in French, German and Russian]. Nationella centret för utbildningsutvärdering. Publikationer 2014:6. [https://karvi.fi/app/uploads/2014/10/KARVI\\_0614.pdf](https://karvi.fi/app/uploads/2014/10/KARVI_0614.pdf)
- Iwashita, N., & Grove, E. (2003). A comparison of analytic and holistic scales in the context of a specific-purpose speaking test. *Prospect*, 18(3), 25-35. [http://www.ameprc.mq.edu.au/\\_data/assets/pdf\\_file/0013/230044/18\\_3\\_2\\_Iwashita.pdf](http://www.ameprc.mq.edu.au/_data/assets/pdf_file/0013/230044/18_3_2_Iwashita.pdf)
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed Levels of Second Language Speaking Proficiency: How Distinct? *Applied Linguistics*, 29(1), 24-49. <https://doi.org/10.1093/applin/amm017>
- Khabbazbashi, N., & Galaczi, E. (2020). A comparison of holistic, analytic, and part marking models in speaking assessment. *Language Testing*, 37(3), 333-360. <https://doi.org/10.1177/0265532219898635>
- Krigh, J. (2019). *Språkstudier som utbildningsstrategi hos grundskoleelever och deras familjer [Language studies as educational strategy among lower secondary students and their families]*. [Doctoral dissertation, Uppsala University]. <https://www.diva-portal.org/smash/get/diva2:1360033/FULLTEXT01.pdf>
- Kunnan, A. (2004). Test fairness. In M. Milanovic & C. Weir. (Eds.), *European Language Testing in a Global Context. Proceedings of the ALTE Barcelona conference, July 2001*. Cambridge University Press.
- Linell, P. (2019). The Written Language Bias (WLB) in linguistics 40 years after. *Language Science*, 76, 101230. <https://doi.org/10.1016/j.langsci.2019.05.003>
- Malmberg, P. (2001). Språksynen i dagens kursplaner [Languages in today's course plans]. In R. Ferm, & P. Malmberg (Eds.), *Språkboken: En antologi om språkundervisning och språkinläring [The language book: An anthology on language teaching and language learning]*, (pp. 16-25). Skolverket/ Liber Distribution.
- May, L. (2011). Interactional Competence in a Paired Speaking Test: Features Salient to Raters. *Language Assessment Quarterly*, 8(2), 127-145. <https://doi.org/10.1080/15434303.2011.565845>.
- Messick, S. A. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Third edition, pp. 13-103). American Council on Education/ Macmillan.
- Messick, S. A. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256. <https://doi.org/10.1177%2F026553229601300302>
- National Agency for Education. (2013). *Att tala eller inte tala spanska: en fördjupning av resultaten i spanska från Internationella språkstudien 2011*. [To speak or not to speak

- Spanish: an analysis of the results in Spanish from the European Survey on Language Competences] Stockholm. <https://www.skolverket.se/publikationsserier/ovrigt-material/2013/att-tala-eller-inte-tala-spanska?id=3186>
- National Agency for Education. (2018). *Curriculum for the compulsory school, preschool class and school-age educare*. <https://www.skolverket.se/download/18.31c292d516e7445866a218f/1576654682907/pdf3984.pdf>
- National Agency for Education. (2020). *Pedagogisk personal i skola och vuxenutbildning läsåret 2019/20*. Report. <https://www.skolverket.se/publikationsserier/beskrivande-statistik/2020/pm---pedagogisk-personal-i-skola-och-vuxenutbildning-lasaret-2019-20>
- National Agency for Education. (2022a). *Kommentarmaterial till kursplanen i engelska. Grundskolan*. [Commentary on the English curriculum. Compulsory school]. Retrieved on 2022-12-05 from <https://www.skolverket.se/publikationer?id=9786>
- National Agency for Education (2022b). *Kommentarmaterial till kursplanen i moderna språk. Grundskolan*. [Commentary on the modern languages curriculum. Compulsory school] Retrieved on 2022-12-05 from <https://www.skolverket.se/publikationer?id=9791>
- National Agency for Education. (2022c). *Statistik [Statistics]* <https://www.skolverket.se/skolutveckling/statistik/>
- Oscarson, M. (2015). Bedömning på systemnivå – En komparativ studie av stegsystemet i språk i den svenska skolan och språknivåer i Europarådets *Common European Framework of Reference for Languages (CEFR)*. [Assessment at the system level – A comparative study of the ‘step system’ for languages in the Swedish school system and language levels in the Council of Europe *Common European Framework of Reference for Languages (CEFR)*.] *Educare* 2015(2), 128-153.
- Riis, U., & Francia, G. (2013). *Lärare, elever och spanska som modernt språk: Styrkor och svagheter – möjligheter och hot* [Teachers, students and Spanish as modern language: Strengths and weaknesses – opportunities and threats]. Uppsala University: Fortbildningsavdelningen för skolans internationalisering [Centre for Professional Development and Internationalisation in Schools].
- Ringbom, H. (2007). *Cross-linguistic Similarity in Foreign Language Learning*. Multilingual Matters.
- Roever, C., & McNamara, T. (2006). Language testing: the social dimension. *International Journal of Applied Linguistics*, 16(2), 242-258. <https://doi.org/10.1111/j.1473-4192.2006.00117.x>
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391. <https://doi.org/10.1177%2F026553220101800404>
- Swedish Schools Inspectorate. (2022). *Skolans arbete för kvalitet i undervisningen moderna språk – Undervisning för kommunikativ förmåga [Work in schools for quality in modern language instruction. Education for communicative competence]*. <https://www.skolinspektionen.se/globalassets/02-beslut-rapporter-stat/granskningsrapporter/tkg/2022/moderna-sprak/skolans-arbete-for-kvalitet-i-undervisningen-i-moderna-sprak.pdf>
- Takala, S., Erickson, G., Figueras, N., & Gustafsson, J-E. (2016). Future Prospects and Challenges in Language Assessments. In D. Tsagari & J. Banerjee (Eds.), *Contemporary Second Language Assessment, Contemporary Applied Linguistics* (pp. 299-315). Bloomsbury Academic.
- Trim, J. (2002). *Modern languages in the Council of Europe 1954-1997*. Council of Europe; Language Policy Division, Strasbourg. <https://rm.coe.int/modern-languages-in-the-council-of-europe-1954-1997-international-co-o/1680886eae>
- Österberg, R. (2021). Perfil laboral y formativo del profesorado de español como lengua extranjera o segunda (ELE/EL2) en Suecia. [Employment and training profile of teachers of Spanish as a foreign or second language in Sweden.] *Journal of Spanish Language Teaching*, 8(1), 1-15. <https://doi.org/10.1080/23247797.2021.1952712>

## Appendices

### Appendix A

Here we provide descriptions of the two speaking tasks. We illustrate with the Spanish version of the tasks.

#### Production task

The students were asked to imagine a Skype-conversation where they would present themselves and their school in the target language to a student in a visiting class. To assist them, the students had a worksheet with photos representing school and leisure activities/objects together with a couple of keywords written in the respective target language. Students were asked to speak as much as possible and invited to go beyond the photos of the sheet if they wished to.



Figure. Worksheet, production task (Spanish)

#### Interaction task

Each pupil had a worksheet in front of them representing an agenda of a fictive week. The two worksheets were different, and students were instructed not to show the worksheet to the other pupil. On each of the worksheet several activities were already planned and written in the target language, but the activities were different and on different days/time on each of the worksheet. In addition, the worksheet contained pictures of sample activities with associated keywords written in the target language.

The students were asked to suggest at least five activities to their peer and to discuss if and when each activity could be scheduled during the week. Pupil A would typically make a suggestion to pupil B including the name of the activity and when it could be scheduled. Pupil B needed to look at his/hers schedule to make sure nothing else was planned at this time and respond. If the suggested activity clashed with something which was already planned, pupil B needed to say no to the suggestion, explain why, and, perhaps, suggest an alternative day of the week.

LA PLANIFICACIÓN DE UNA SEMANA CON UNA CLASE ESPAÑOLA



Figure. Worksheet, interaction task, Student A (Spanish)

LA PLANIFICACIÓN DE UNA SEMANA CON UNA CLASE ESPAÑOLA



Figure. Worksheet, interaction task, Student B (Spanish)

## Appendix B

Here we present details on inter-rater agreement among researchers as raters (Table 1) and among the external raters 1 (Table 2)

Table 1. *Inter-rater agreement among researchers as raters*

	Cronbach's Alpha	Spearman rank order  (rho)	Number of raters	Number of cases
French	.894	.808**	2	153
German	.900	.822**	2	108
Spanish	.656	.505**	2	103

Table 2. *Inter-rater agreement among external raters (first group)*

	Cronbach's Alpha	Spearman rank order  (rho)	Number of raters	Number of cases
French	.834	.608**	3	30
German	.834	.631**	3	30
Spanish	.867	.663**	3	30

## Appendix C

Table 3: Qualitative features of spoken language (expanded with phonology), Companion Volume, Appendix 3

Level	RANGE	ACCURACY	FLUENCY	INTERACTION	COHERENCE	PHONOLOGY
B1	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.	Pronunciation is generally intelligible; can approximate intonation and stress at both utterance and word levels. However, accent is usually influenced by other language(s) he/she speaks.
A2	Uses basic sentence patterns with memorised	Uses some simple structures correctly, but still	Can make him/herself understood in very	Can ask and answer questions and respond to simple	Can link groups of words with simple connectors like	Pronunciation is generally clear

	phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	systematically makes basic mistakes.	short utterances, even though pauses, false starts and reformulation are very evident.	statements and indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	"and, "but" and "because".	enough to be understood, but conversational partners will need to ask for repetition from time to time. A strong influence from other language(s) he/she speaks on stress, rhythm and intonation may affect intelligibility, requiring collaboration from interlocutors. Nevertheless, pronunciation of familiar words is clear.
A1	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorised repertoire.	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.	Can ask and answer questions about personal details. Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.	Can link words or groups of words with very basic linear connectors like "and" or "then".	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by interlocutors used to dealing with speakers of the language group concerned. Can reproduce correctly a limited range of sounds as well as the stress on simple, familiar words and phrases.

Received March 15, 2023  
 Revision received August 21, 2023  
 Accepted September 21, 2023