



●
● CEFR-level writing skills
● for English, German
and French

A research study into the levels attained by Dutch students in havo (senior general secondary) and vwo (pre-university) education

SLO • Netherlands institute for curriculum development

slo



CEFR-level writing skills for English, German and French

A research study into the levels attained by Dutch students in havo (senior general secondary) and vwo (pre-university) education

August 2014

slo

Netherlands
institute for
curriculum
development

Colophon

**2015 SLO (Netherlands Institute for Curriculum Development), Enschede,
the Netherlands**

All rights reserved. Nothing of this publication may be reproduced and / or published
without prior written permission of the publisher.

Authors: Daniela Fasoglio, Anne Beeker, Kim de Jong (SLO)
Jos Keuning, Alma van Til (Cito)

Translated from Dutch: Vertaalbureau Omnitaal

Information

SLO

Department: Upper Secondary Education

PO box 2041, 7500 CA Enschede

Telephone: (053) 4840 661

Internet: www.slo.nl

E-mail: tweedefase@slo.nl

AN: 3.7343.665

Contents

Preface	5
1. Introduction	7
1.1. Context and motivation	7
1.2. Terminology	8
2. Setup of the study	13
2.1. Approach adopted	13
2.2. Determination of study group	14
2.3. Set of instruments	14
2.4. Study activities	16
3. Description of the study group	19
3.1. Age	19
3.2. Gender	20
3.3. Parents' educational level	21
3.4. Regional spread	21
4. Test content and assessment	25
4.1. Test content	25
4.2. Assessment format	28
4.3. Preliminary study for the assessment format	30
5. Results	33
5.1. English	33
5.2. German	34
5.3. French	36
5.4. Comparison of results	38
6. International standard-setting	39
6.1. Composition of expert panels	39
6.2. The procedure	41
6.3. Results	42
7. Results teacher interviews	47
7.1. Selection of schools	47
7.2. Results of the interviews	47
8. Conclusions	53
8.1. The level of writing skills <i>havo/vwo</i> in English, German and French	54
8.2. Evaluation of tests and assessment methods used	55
8.3. Implications for writing tuition in English, German and French	55
References	57
Appendices	61

Appendix 1 Balanced Incomplete Block Design	63
Appendix 2 Composition of expert panels	65
Appendix 3 Design English vwo for international standard-setting	71
Appendix 4 Examples summaries discussion 2nd round standard-setting	73
Appendix 5 Interview guidelines	83

Preface

Dutch education has a long tradition where modern foreign languages are concerned. While the emphasis during the first half of the twentieth century had been more on reading and writing skills, a shift towards communicative skills took place in the nineteen seventies and eighties. These were considered very important from a social point of view. After all, what good is the ability to read a complex story if it is impossible to simply order a cup of coffee abroad? And this has not been without effect, considering the following.

According to the *English Proficiency Index* (EF EPI, 2013), Dutch people are good English speakers; when arranging the order of English-speaking proficiencies in non-native English-speaking countries around the world, the Netherlands occupies a third position. Only Sweden and Norway demonstrate a better command of the English language. Putting things into perspective, this proficiency does have its shortcomings, as is demonstrated by the humorous book *I always get my sin*, by Maarten Rijkens, which describes the misunderstandings English-speaking Dutch people can cause. For example, when they ask for the *cock* to compliment him after having enjoyed a scrumptious dinner. Unfortunately, there is no comparable index for the command of spoken German and French, which seems to have deteriorated in the Netherlands. In table talk people often comment on the diminishing language skills of the average Dutchman. But is there, in fact, hard proof of this?

A European study (Kordes, 2012) into the reading, listening and writing skills of students upon conclusion of basic education where the two most widely taught languages are concerned (in the Netherlands: English and German), during the third year of *havo* (senior general secondary education) and *vwo* (pre-university education) and the fourth year of *vmbo* (pre-vocational secondary education), provides a balanced view. While almost 60% of the students are already at B2 level where English listening skills are concerned, the percentage of students that reach B2 for reading and writing in English is lower than in other European countries. Regarding German, the results are quite different. Whereas Dutch students reach a B level for listening and reading, particularly, some European countries predominantly show an A level. Note that no German is spoken in those countries. Therefore, compared to the reference group, Dutch is the language that has the most affinity with German. Where French is concerned, no comparative study material is available.

What happens next during the upper secondary education, in the Dutch educational system known as *tweede fase* (second phase)? In 2013, the level attained for receptive skills in reading and listening was studied by the Dutch National Institute for Educational Measurement Cito (Feskens, Keuning, van Til, & Verheyen, 2014). However, no study has, as yet, been carried out regarding the levels attained in the productive skills: writing and speaking. The study described here intends to link the evidence of one of these, i.e. writing skills, to the levels of the Common European Framework of Reference (CEFR). For a representation of the written language proficiency in English, German, and French, at the level achieved before continuing on to a Bachelor's or Master's study, the writing performances of students of 5th form *havo* and 6th form *vwo* have been studied. In this report, you will find the most important results of this study. In Chapter 1, we will describe the context that gave rise to this project.

Chapter 2 gives the setup of the study and an overview of the instruments and methods used. Chapter 3 describes the characteristics of the group studied.

In Chapter 4, we will enter into the structure of the assessment tests used for the study, the assessment directives, and the assessment procedure. The next chapter, Chapter 5, will show the scores the students achieved.

The written products have subsequently been subjected to an international standard-setting procedure to determine the CEFR level. The procedure and its results have been described in Chapter 6.

Chapter 7 presents a few interviews that have been conducted with some of the participating teachers. These teachers told us about their experiences with the teaching of writing skills in the modern foreign languages and list success factors and points of attention.

The final chapter, Chapter 8, provides a summary of the most important conclusions that may be drawn from this study. In addition, it describes the level of written language skills in English, German, and French at *havo* and *vwo* level, as these have emerged from the study.

The results of this project can be used as a basis for discussions about the curriculum for modern foreign languages in *havo* and *vwo*. As curriculum experts, we wish to stimulate such discussions and actively participate in them. We are hoping that these will lead to concrete steps to optimise the teaching of writing skills in modern foreign languages.

1. Introduction

This chapter describes the context and the considerations that have led to the setup of this study. We will also briefly explain its objective. Next, we will enter into the concepts that form the basis of the study: the Common European Framework of Reference (CEFR) and the acquisition and assessment of writing skills in modern foreign languages (FL) in *havo* and *vwo* (Dutch senior secondary education and Dutch pre-university education respectively).

1.1. Context and motivation

In 2007, the examination programmes for modern foreign languages in *vmbo*, *havo* and *vwo* were linked to the levels of language skills according to the CEFR. As a result, the globally formulated attainment level A (reading skills) assessed in the central examinations was provided with level indications in the syllabi of the Dutch Board of Examinations (*College voor Toetsen en Examens*, or CvTE). The CEFR levels of the test items in the central examinations for modern foreign languages have been based on the linkage studies of the Dutch National Institute for Educational Measurement (Cito) from 2006 (German, English, and French) and 2007 (Arabic, Russian, Turkish, and Spanish), in which the relationship between the examination programme and the CEFR on the one hand, and the test items and the CEFR on the other were examined. For these studies, a great many experts from the Netherlands were consulted. To reach international consensus on the interpretation of CEFR levels, a standard-setting procedure took place in 2013. This procedure involved test items for reading and listening skills in English, French, and German at *havo*, *vwo* and *vmbo* levels, with the help of a group of language and assessment experts from different European countries. The results of this operation have been reported by Cito (Feskens, Keuning, Van Til, & Verheyen, 2014).

Regarding the skills that are tested in the school examinations, suggestions for attainable CEFR levels have been made in the guidelines of SLO (Meijer & Fasoglio, 2007). To this purpose, an analysis of two documents took place: the level specifications of the examination programmes for the school examinations FL *havo/vwo*, as applicable in 2007 (and as developed by the Dutch committee on the renewal of final examination programmes CVEN, chaired by Prof. Van Els) and the text of the CEFR (Council of Europe, 2001) with further details contained in the document about language profiles *Taalprofielen* (Liemberg & Meijer, 2004). The levels described in the first document have been compared to the levels of the Common European Framework of Reference. However, regarding productive skills, this procedure lacked a link between the proposed target levels and the actual performances of students in the school examinations for the different languages. Such a linking is essential to provide insight into the attainability of those final levels within an educational context. Evidence-based results from educational practice should substantiate which performance requirements in CEFR terms are realistically attainable. This conclusion has been the motivation for the setup of the study, the results of which are reported in this publication.

The objective of the study and the research question formulated in it

In 2012, SLO started up a project to fulfil the need for data based on evidence from educational practice. For reasons of feasibility and taking into account the available means, the study was limited to just one of the productive skills, regarding the three most commonly taught languages, in only two of the educational sectors. Concretely, the intended result of the study was to determine the level of writing skills attained by students of English, German, and French, upon

concluding their education at *havo* and *vwo* level. Considering the fact that the target levels of *havo* and *vwo* are related to the Common European Framework of Reference, the level of writing skills was to be defined in CEFR terms.

The research question formulated in this study is as follows:

Which CEFR level is attained by students upon concluding their havo and vwo education, with regard to writing skills in English, German, and French?

1.2. Terminology

a. The Common European Framework of Reference, CEFR

CEFR refers to the Common European Framework of Reference for Languages.

The CEFR is a system of level descriptions for the modern foreign languages, which provides a common basis for the development of curriculum guidelines, educational programmes, teaching materials, and examinations throughout Europe. This way, a comparison between language qualifications becomes possible, which should facilitate mobility within Europe.

The CEFR describes which language performances correspond to a certain language proficiency level regarding both content (in terms of language actions in social contexts) and quality (in terms of grammatical correctness, vocabulary, pronunciation, spelling, etc.).

The CEFR distinguishes six stages of language performance levels. These are worked out in further detail in descriptions of what a person can do in the language concerned. These so-called descriptors are phrased in can-do statements. The six levels have been summarised in Figure 1.

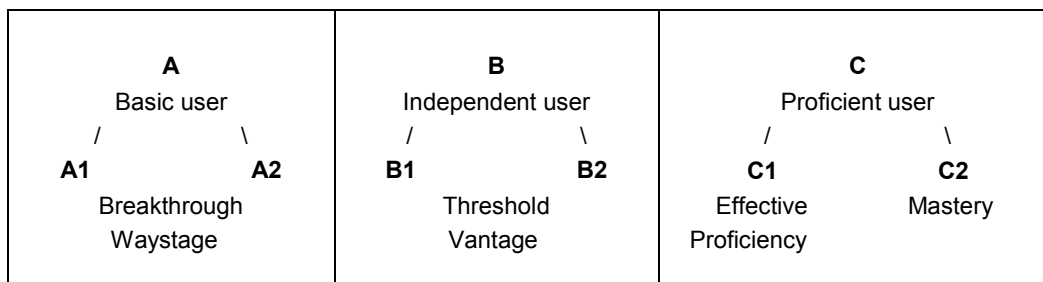


Figure 1. The six levels of the Common European Framework of Reference, CEFR

For the first level (A1), only beginner's knowledge is required, regarding very concrete situations and in familiar contexts. Subsequently, the level increases in steps up to C2, which describes effortless language proficiency, regardless of complexity, register, and implicit meanings.

The CEFR is based on five skills: listening, reading, conducting conversations (or: spoken interaction), speaking (or: spoken production), and writing. These are also the skills that are to be tested in the final examinations of the modern languages in the Netherlands.

In the CEFR, each skill is worked out in further detail, resulting in a number of sub-categories or sub-skills describing different language activities. For example, where writing skills are concerned, scales have been described for the writing of correspondence, memos, messages and forms, for the drawing up of reports and essays, for the taking of notes, for word-processing, and for creative writing.

The CEFR also distinguishes four domains: different spheres of action or areas of concern in social life, within which communicative situations are taking place. These are:

- daily life (the personal domain): situations that concern a private person, such as hobbies, contacts with relatives and friends, or pleasure reading;
- the public sector (public domain): situations in which a person operates as a member of society, for example in a restaurant, at a desk, and during contacts with companies or other organisations;

- work (professional domain): all work-related situations, including work done by students, such as weekend jobs;
- education (educational domain): all situations concerning school and training.

The Common European Framework of Reference, CEFR, provides us with a descriptive model of language-use contexts, whereby reference scales are described of *what* the language user can do with a language and *how well* he or she is able to do this. The CEFR is based on the perception of assessors/teachers about the classification of language behaviour, rather than on a language-acquisition theory (concerning the latter, refer to Fulcher, 2010; Hulstijn, 2007; Weir, 2005, and others). That is why the CEFR has been deployed as the assessment instrument for this study.

b. Writing skills in modern foreign languages

The testing of writing skills should be in agreement with the objectives of writing tuition in a modern foreign language. Therefore, it is essential to determine the objective. In literature, there is no univocal definition of writing skills in FL teaching. We have, however, come across different definitions, depending on the emphasis on, for example, the product or the process, or on communicative, linguistic, or cognitive features.

In his dissertation about the testing of writing skills, Melse (1990) has opted for the following working definition: *Writing skills refer to the skills of producing a written, cohesive, and well-rounded text in which formulations can be chosen freely.*

This definition allows a lot of leeway. On the one hand, there are the requirements for grammatical correctness and vocabulary (depending on the language concerned and the educational level), while, on the other, there is the context in which the written communication is taking place and, with it, the purpose of writing, the reader or readers, and the conventions that are appropriate to the different text types.

The explanatory scales described by the CEFR levels take both the linguistic aspects and the communication situations into account. The CEFR takes an active approach. The language learner is particularly regarded as a member of society who uses language in a socio-cultural context. Writing skills are globally described as follows, in relation to A2, B1 and B2 levels, respectively.

A2: I can write down short, simple notes and messages. I can write a very simple, personal letter, for example to thank someone for something.

B1: I can write a simple, cohesive text about topics that are familiar or of personal importance to me. I can write personal letters in which I describe my experiences and impressions.

B2: I can write a clear, detailed text about a wide range of topics that are of interest to me. I can write an essay or report, pass on information, and give reasons to support the pros or cons of a specific point of view. I can write letters in which I indicate the personal interest of events and experiences.

In other words, a basic definition of what general writing skills are understood to be is lacking. It can, however, be derived from these I-can statements.

From the above, it can be concluded that formulating a writing task at a specific CEFR level, and subsequently the assessing of it, should involve the following aspects: firstly the linguistic ones, regarding morphosyntax, range and command of vocabulary, and orthography, and secondly the contextual/communicative ones, including objective, composition, writing medium, relationship to the receiver, and socio-cultural conventions.

c. Teaching writing in the native and foreign language

- *Acquired sub-skills in the native language*

Because of the limited language range, writing in a modern foreign language has more limitations than writing in the native language - even though similar skills are involved. When writing in a foreign language, various sub-skills already present or being acquired in the mother tongue are made use of. These include cognitive skills, such as logical ordering, summarising, and a knowledge of the characteristics of text composition appropriate to the different text types. However, register choice and attunement to the objective and the target audience are determined by the socio-cultural contexts in which the foreign language is used, and the conventions to deviate from these in relation to the mother tongue.

- *Lower level of language command*

In general, students from *havo* and *vwo* will reach a lower level of command of the foreign language than the level they have acquired for their mother tongue. In addition, there are differences between *havo* and *vwo*, and between the English language on the one hand, and German and French on the other. Even though the communicative objective for writing in native and foreign languages is the same, the emphasis where FL teaching is concerned lies on aspects of language acquisition.

Still, FL teaching also concerns a skill to be acquired, which, therefore, needs to be tested. During the test, the language knowledge needs to be put into a functional perspective. In other words, the textual competence, or the capacity to construct a cohesive and well-rounded text, is to be tested. Simply put: students must demonstrate their knowledge of grammar and vocabulary as well as their ability to write a sound text.

- *Educational targets versus assessment targets*

When teaching, testing and assessing writing skills in a modern foreign language, aspects of the writing product such as lexical and morphosyntactic ones will be given more weight compared to the teaching of writing in the mother tongue. However, the ability to use a foreign language in a communicative context also plays an essential role in FL. Both the attainment levels in the final examination programme and the descriptors of the CEFR have been developed from this perspective.

To ensure reliability of results, we have put more emphasis on the language skills component during the development of the assessment criteria regarding this study.

d. Writing skills in the final examination programme for modern foreign languages

havo/vwo

The attainment levels of the final examination programme globally describe which attainment targets should be achieved for each subdomain of a subject. These descriptions place the use of the foreign language in a social context, within which communicative goals are achieved by means of language. The attainment level for FL writing skills *havo/vwo* is formulated as follows:

Domain D – Writing skills

Sub-domain D1: language skills

5. The candidate can:

- respond adequately in written contacts with users of the target language;
- request and provide information;
- present acquired information in an adequate manner, considering objective and target audience, thereby describing matters or persons and expressing feelings and points of view;
- write a report in the language concerned (does not apply to *havo* German and French).

e. Target levels for writing skills in a modern foreign language at *havo* and *vwo* level

Attainment level D from the final examination programme *havo/vwo* comprises a global description that needs to be translated into the contexts that apply and the levels of language proficiency that are to be expected. The School Examination guidelines give suggestions for this and propose so-called 'target levels' for the different languages, sectors and skills: they are an indication of realistically attainable final levels for average students under good teaching conditions. Where *havo* and *vwo* are concerned, the following CEFR target levels have been determined for English, German, and French.

Table 1. *Target levels of writing skills in English, German and French for havo/vwo*

	<i>havo</i>	<i>vwo</i>
English	B1	B2
German	A2+	B1
French	A2+	B1

The plus level A2+ is described as a very good A2. It concerns a performance that falls within the band width of A2, but which may stand out regarding quantity and quality, for example by using a broader language range.

Examples of writing activities referred to in the CEFR include: filling in forms and questionnaires, writing articles for magazines, newspapers, newsletters, etc., making up posters and advertisements, reporting and writing reports, making summaries, taking notes for self and others, creative writing, writing personal or business letters, etc.

The testing and assessing of writing skills in the FL school examination should enable the making of a judgement regarding the command at one of the above levels.

The results of the study described in this publication give an indication of the CEFR levels at which *havo* and *vwo* students are performing.

2. Setup of the study

This chapter explains the activities that have been carried out between 2012 and 2014 to provide an answer to the research question and briefly describes the instruments developed to this purpose.

2.1 Approach adopted

In the plan of action drawn up for this study, the instruments required and the study activities have been described.

For this study, both curricular and assessment expertise were needed. That is why in the very early stages - during the preliminary phase of the project, in fact - we invited Cito to collaborate. This institute has been responsible for the development of the tests and the setting up of the assessment procedures. The tasks assigned to Cito and SLO, respectively, are shown in table 2. This table gives an overview of the activities that have been carried out within the context of this study.

The study took over two calendar years - from January 2012 up to and including February 2014.

Table 2. Overview study activities and time path (pink = SLO; grey = Cito)

2012	March	Setting up a plan of action, determination of the study group and the study instruments, attunement of tasks and competences SLO and Cito		
	April			
	May			
	June			
	July	Recruitment study group, contacts with schools	Development of six skill tests: English <i>havo</i> , English <i>vwo</i> , German <i>havo</i> , German <i>vwo</i> , French <i>havo</i> , French <i>vwo</i>	
	August		Taking of tests for selection of anchors	Information meeting with teachers
	September			
	October	First testing session	Development of assessment procedure by means of anchors	Selection of anchors for assessment (Cito/SLO)
	November			
	December	Second testing session	Setting up assessor's design and assessment procedure	Processing study group information
2013	January			
	February	Third testing session	Development of an international standard-setting procedure	Processing and analysis of scores
	March			
	April	Scoring of student work using product scale by means of anchors (SLO/Cito)	Processing study group information	Development of an international standard-setting procedure
	May			
	June			
	July	Processing and analysis of scores	Development of an international standard-setting procedure	Development of an international standard-setting procedure
	August			
	September	Recruitment of	Development of digital	Development of an international standard-setting procedure

	October	language experts for international-standard-setting	platform for international standard-setting
	November	Process of international standard-setting	Teacher interviews
	December		
2014	January		
	February	Processing and analysis of results standard-setting procedure, including setting of performance standard and setting of cutting score	
	March		
	April	Conclusions, writing of study report (i.c.w. Cito)	
	May		
	June		
	July		

2.2 Determination of study group

The study was focussed on students from 5 *havo* and 6 *vwo* during the school year of 2012-2013. A sufficiently representative study group was aimed for, so that the results could be generalised to include the full population.

Table 3 presents the numbers of examination candidates from 2013, arranged by language and sector, in relation to the size of the study group. Using this study group, the reliability level is 95%.

Table 3. *Study group*

Language and sector	Number of examination candidates 2013	Study group reached
English <i>havo</i>	55,530	371
English <i>vwo</i>	38,867	298
German <i>havo</i>	18,267	221
German <i>vwo</i>	20,039	365
French <i>havo</i>	11,499	188
French <i>vwo</i>	16,065	292

2.3 Set of instruments

The following instruments have been developed for this study:

- tests for writing skills that invite students to produce written texts at attainment levels that are considered appropriate at *havo* and *vwo* levels;
- an assessment procedure for the application of scores to the student products;
- a standard-setting procedure, so that CEFR levels can be linked to the scores, in order to determine the CEFR level of the collected written products;
- a digital platform for international expert panels to carry out the standard-setting procedure;
- interview guidelines to conduct interviews with some of the teachers who have subjected their students to the skill tests for this study, in order to obtain insight into their experiences with the teaching of writing skills in modern foreign languages.

a) Skill tests

Cito has developed a set of writing-skill tests for English, German and French at *havo* and *vwo* level. Chapter 4 describes the setup of these tests.

b) Assessment procedure

To provide the student products with a score, an assessment procedure has been developed, using a product scale by means of anchors (Kuhlemeier, 2011). In this procedure, assessors compare a passage from the student's work with two so-called anchors, i.e. two examples of written results for the same sub-assignment, which serve as reference points for a relatively good and a relatively poor result, respectively. The score of the student's work will be a sum of the individual scores for all passages in the work itself. The procedure will be discussed in further detail in Chapter 4.

c) Standard-setting procedure

To be able to determine the CEFR level of the collected written products, a standard-setting procedure was developed. By means of this procedure, a score can be determined, which serves as a cutting score or performance standard for a previously determined CEFR level. The student works with a score equal to or higher than the performance standard meet the quality requirements for that level. There are different methods for the determination of a cutting score. Test-oriented and student-oriented standard-setting methods have been distinguished (Jaeger, 1989; Kaftandjieva, 2004; Berk, 1986; Hambleton, Jaeger, & Plake, 2000). Where test-oriented methods are concerned, subject experts base the cutting score on the test content and learning content. The cutting score is independent of the test results actually achieved by the students. In case of student-oriented methods, the cutting score is determined on the basis of the results of a group of students during a test. In that case, therefore, the cutting score depends on the results of the group of students being tested. For this project, we have made use of the method of contrasting groups, which applies to the second category. Using this method, the judgements of the experts regarding which student works meet the minimum requirements of a pre-determined CEFR level and which do not are compared to the test scores that have been appointed to the works. The test score with the least classification discrepancies will be eligible for use as a cutting score for the CEFR level. This method is the only one that enables a calculation of probable decision errors (false positives and false negatives) in case of cutting-score determinations. That is why this method is generally recommended in literature where standard-setting procedures are concerned (Fulcher, 2010: 241). For a further explanation, please refer to: Sanders (ed.), 2013 (pp. 150-152). Chapter 6 describes the results this procedure has led to.

d) Online digital platform for international-standard-setting

To efficiently set standards using international panels of experts living in many different countries, a digital platform was developed to support the standard-setting procedure. Thanks to this application, the entire procedure could be gone through in just a few steps, in order to provide insight into the level estimations and the corresponding argumentations, thus also allowing the exchange of views among the experts. To facilitate readability, the student works that have been selected for the standard-setting procedure have been typed up in their entirety and uploaded to the platform. Every expert was given their own login name and password to log into the system. After logging in, every expert was given a personalised selection of the works to be assessed and the possibility to enter their own assessments. The results were automatically stored in the system, which generated overviews in Excel. For each of the three languages, the platform had been provided with separate sites. For further details, please refer to Chapter 6.

e) Interview guidelines

To obtain some insight into the experiences of teachers regarding the teaching of writing skills, interviews have been conducted with teachers from four of the participating schools. To conduct these interviews, interview guidelines were developed, based on the following two themes:

- a) factors affecting the learning results (both positively and negatively);
- b) testing and assessment methods regarding writing skills.

To draw up the interview guidelines, elements of the curricular spider web (Van den Akker, 2003) were used. This is a model in which ten different curricular components are visualised showing how they are interconnected:

- rationale,
- aims and objectives,
- learning content,
- learning activities/work forms,
- teacher role,
- materials and resources,
- learning environment
- grouping arrangements,
- time,
- assessment.

Teachers were able to place their answers under the heading of three or more of these components.

In the second part of the interview, teachers were asked about the content and form of the writing tests they usually gave their students. Subsequently, they were asked to describe their experiences with the writing test of the study and the two different assessment methods (the Cito assessment model and the scoring procedure based on anchors). Finally, teachers were asked for their golden tip to test writing skills. The guidelines used for the interviews are included in Appendix 2.

2.4 Study activities

a. Recruitment of schools

The recruitment of schools started in June 2012. To recruit schools, notices were published in subject-community sites for English, German, and French; in the publication *Levende Talen*, and in SLO's FL newsletter. In addition, all teachers FL who had participated in field-work projects by SLO (such as projects for quality assurance of school examinations and CEFR implementation) during the period 2010 – 2012 were approached, altogether over 75 schools with around 115 teachers.

The following conditions were set as a basis for participation in the study:

- the willingness to have students from the examination year of *havo* or *vwo* perform a 100-minute writing test for English, German, or French;
- the willingness to participate in an information session during one afternoon in the autumn of 2012;
- the obligation to participate in the sessions in spring of 2013, during which the student works were assessed using anchors.

At a later stage, these basic conditions were further refined with the following two conditions:

- to not allow any dictionary or grammar book – this was decided upon for reasons of international comparability of the test results;
- to *not* have the tests made on the computer but on paper – for the same reason as mentioned above.

Students had to be motivated to take their writing assignments seriously. Teachers were to experience the study as useful and valuable. That is why it was decided to offer the schools the opportunity to use the tests provided by SLO and Cito as an SE test during the year 2012 – 2013; however, this was no obligation. For the same reason, the participating schools were offered a re-examination test. Because the recruitment took place as early as June 2012, potential schools were also given sufficient time to adjust their Programme of Testing and Conclusion (*Programma van Toetsing en Afsluiting*, PTA), if necessary, to the conditions that applied to participation. Also, the opportunity to be coupled to another school from the study project was provided, to allow for a second correction. Schools were also allowed to use the test as a practice test, on condition that the participating students were already in *havo* 5 or *vwo* 6.

According to the rules and regulations that apply, the responsibility for any SE mark provided has, in all cases, been left with the teacher. To determine the marks, teachers were able to use the descriptive assessment model provided, which had been derived from the model developed by Cito for earlier writing tests *TaalstERK*. This had explicitly *not* been an obligation for the giving of marks. It was particularly intended as a service for teachers who desired to give an SE mark. As described in Chapter 1, the study was focussed on the determination of the acquired attainment levels for writing skills in terms of the Common European Framework of Reference, rather than on marks given.

The invitation to recruit participants also included: a description of the objective of the study; an explanation of the assessment method that was to be used to provide student works with scores using a product scale with anchors (see Chapter 4), for convenience's sake called 'the anchor method'; a mention of the fact that teachers were obliged to participate in the central scoring session; and the necessity, regarding the latter, to request permission by their principal to be exempted from teaching on that day.

Many schools applied. Very soon, in July 2012, a waiting list had to be drawn up for the subject of English because over 450 students had been put forward, both for *havo* and *vwo*. After the summer holidays of 2012, it turned out that there were still French and German participants to be recruited, particularly for the *havo* tests. That is why, early September 2012, another appeal was made to the teacher language communities. In the end, by October 2012, the study group was sufficiently large. During the period between application and the eventual taking of the tests in schools, the study had to deal with cancellations by schools. Reasons for cancellation included long-term illnesses of teachers concerned, as well as the conditions under which the tests had to be taken, specifically the non-allowance of dictionaries and digital resources. Although these conditions had been explicitly stated in the letter of invitation, they had, apparently, been overlooked on many occasions.

b. Testing for the purpose of anchor selection

Four schools have contributed by providing approximately 50 student works from which the anchors could be selected. These schools had participated earlier in an SLO field-work project. This preliminary testing session took place in October/November 2012 and eventually resulted in the following numbers of tests.

Table 4. *Collected student works for the preliminary testing session*

	<i>German</i>	<i>English</i>	<i>French</i>
<i>havo</i>	44	48	32
<i>vwo</i>	42	46	40

c. Information meeting

The final writing tests were delivered by Cito early October 2012. Subsequently, information regarding their content was provided to the participating schools via e-mail. The schools were asked to tell when they intended to subject students to the tests: during SE period 1, 2, or 3. In addition, the participating teachers were invited to attend an information meeting during the afternoon of 23 November 2012 in Utrecht. During this meeting, an explanation was given on the test construction used (see the tables regarding the composition of writing tests, Chapter 4), the tests were presented, and the teachers present were handed over a CEFR-related assessment model that they could use for the calculation of a mark, should the test also be used as school-examination test. Using a number of tests made by students during the preliminary study, teachers were given the opportunity to practise calculating the marks. In particular discussions arose about the newer types of assignments, such as chat sessions and

blogs, and the weighing, or weighting, of language-command mistakes in comparison to other types of errors.

Many teachers would have preferred to study the tests beforehand, so that they could have prepared their students for what was to come. This had not been possible, because the reliability of the measuring had to be assured since *teaching to the test* might have affected this.

d) The taking of the tests

The first tests were taken in December of 2012 and they were continued up to and including the first week of April 2013. To avoid anyone having prior knowledge, which would have marred the test results, all participating teachers were asked to sign a confidentiality agreement.

Testing period 1 lasted from week 44 up to and including week 51 in 2012, testing period 2 was from week 2 up to and including 6 in 2013, and the final testing period 3 was from week 7 up to and including week 14 in 2013. Ten days before the testing period indicated by the school started, the correct number of tests for each language and school type were sent to the school concerned. The receiving teacher was sent an e-mail by the administration. In a number of cases, a follow-up telephone call was placed to make sure everything had been received correctly. Immediately after the taking of the tests, the school was to make photocopies of the work. These copies were to be returned to the SLO promptly. Each work was then provided with a unique serial number, which enabled the coupling to student information without this information appearing on the work. The student data were entered into the statistical computer programme SPSS. If any copies were illegible, the teacher concerned was contacted to request new copies. In the end, 1735 student works were collected and administered in this way.

To make sure teachers would submit students to the tests in the same way, a detailed instruction sheet with test-taking conditions had been drawn up. This sheet was sent to the schools together with the tests. These once again contained the explicit instructions that the use of a dictionary was not permitted and that it was mandatory to write the tests by hand. Also, the necessity of strict confidentiality of the test content was emphasised. Other items on the instruction sheet involved regular points of attention to be observed in school tests.

e) Assessment sessions

To assess all student works using anchors (see Chapters 2 and 4), central assessment sessions were organised in Utrecht on 16 and 23 April 2013. A total of 68 assessors contributed comprising teachers, retired teachers, employees and interns of Cito, and employees of SLO. There were 25 assessors for the German tests, 21 for the English ones, and 22 for the French ones. Every assessor was presented with both *havo* and *vwo* works for assessment; each assessor was provided with around 35 student works. The (copied) student works had been distributed according to the *balanced incomplete block design* (see Appendix), which had been supplied by Cito. All participants received an assessor's booklet – see also Chapter 4 - and an optically readable form on which the scores were to be indicated by marking a dot with a black pencil. Subsequently, the teachers compared the passages written by the student to the anchors printed in the booklet, according to two aspects: communicative effectiveness and correctness.

As a result, each passage was assessed twice. This way of scoring took around 15 minutes per student work. However, the differences regarding language, student and teacher were considerable: for *vwo* English, for example, the average often amounted to 25 minutes. As it turned out, the available time had not been sufficient to receive back all the works distributed during the assessment sessions together with their scores. After conclusion of the two assessment sessions, approx. 450 unscored student works remained. This remainder of tests were scored at a later stage and returned, which is why the *block design* could no longer be precisely followed. In July of 2013, all scores had been realised.

3. Description of the study group

For the study described in this report, students have been used from the school year 2012-2013 who were taking their final examinations *havo* or *vwo* in the subjects English, German or French. When selecting the students, we aimed for sufficient representativeness, although we were limited by the dependence on teachers' willingness to submit their students to the study. The teachers were recruited via national networks.

In this chapter, a description is given of the study group regarding a number of characteristics: age, gender, educational level of the parents, and region. Wherever possible, these were set against the characteristics of the population as a whole addressed by this study, i.e. final examination candidates *havo* and *vwo* in the Netherlands.

3.1 Age

The majority of students from the study group *havo* are aged between 16 and 18. Their ages barely deviate from the ages of the nation-wide student population. The percentages in the figure below have been calculated on the basis of data from the Dutch Central Statistical Office CBS regarding the ages of *havo* students during school year 2012-2013 (CBS, 2014). The percentage of young students in the study group is somewhat smaller; on the other hand, the percentage of older students is somewhat higher; however, both differences are not significant. Apart from this point, the study group can be regarded as representative for the student population *havo* in the Netherlands. Figure 2 shows the average ages of students for the different languages.

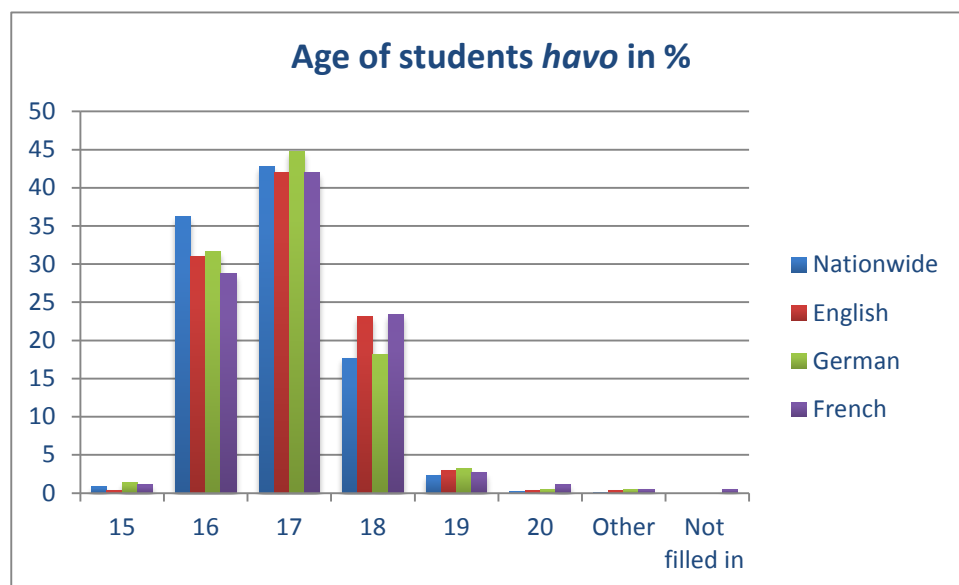


Figure 2. Age of students *havo* (in %)

Where the study group *vwo* is concerned, students were predominantly aged between 17 and 18. The national figures (CBS, 2014) show a similar picture, with minimal differences in the ratio between 17 and 18-year olds. As a result, the study group *vwo* can also be regarded as representative for the student population *vwo* in the Netherlands. The data for *vwo* are visualised in Figure 3.

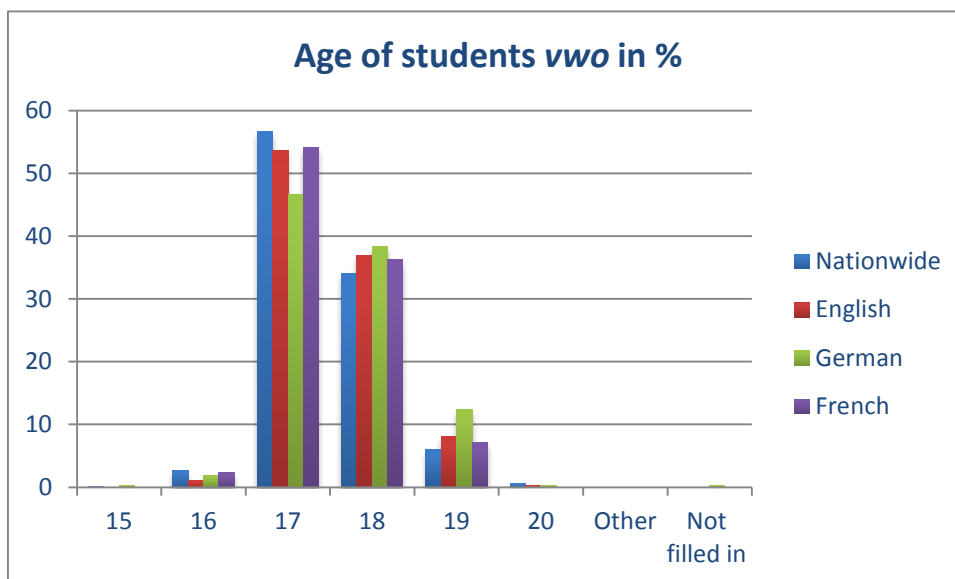


Figure 3. Age of students vwo (in %)

3.2 Gender

Both in *havo* and *vwo*, the study group comprises more girls than boys. There are differences in ratio between the numbers of girls and boys where the three languages are concerned. For English *havo*, 46.9% of the study group comprised boys and 53.1% girls. This ratio agrees with the national data regarding the school year 2012-2013 - the cohort that has been studied: 47.2% of this population comprised boys and 52.8% girls. For English *vwo*, 49.7% of the study group comprised boys and 50.3% girls. On a national scale, the number of girls also exceeded the boys, with a somewhat larger difference between boys and girls (46.6% boys and 53.4% girls). Where the other languages are concerned (with the exception of German *vwo*) the difference between the number of girls and the number of boys is greater than for English, with 77.1% (French *havo*), 65.4% (French *vwo*) and 62.9% (German *havo*) of girls. These ratios have been compared to the national ratios and split up according to the choice of profile. From this, it was established that many more girls than boys choose the C&M (culture & society) profile, whereby *havo* requires a second modern foreign language to be chosen as a profile subject. 14.6% of the girls choose this profile against 3.1% of the boys.

Based on the above comparisons, it was assumed that the study group was sufficiently representative of the entire population regarding this characteristic. An overview of the percentages of the study group per subject is presented in Table 5. All national data used for comparisons originate from the Dutch Central Statistical Office CBS (CBS Statline, 2014).

Table 5. Percentage of boys and girls in *havo* and *vwo*

	Nationwide	English	German	French
Boys <i>havo</i>	47.2%	46.9%	37.1%	22.9%
Girls <i>havo</i>	52.8%	53.1%	62.9%	77.1%
Boys <i>vwo</i>	46.7%	49.7%	52.9%	34.6%
Girls <i>vwo</i>	53.3%	50.3%	47.1%	65.4%

3.3 Parents' educational level

For all languages, the educational level of the parents of the students participating in the study group *vwo* was higher than the educational level of parents of *havo* students. Parents of *havo* students typically completed an education at intermediate vocational level (25.2%) or higher vocational level (Bachelor's degree) (23.4%), while a little over half of the parents of *vwo* students have a Master's (21.5%) or Bachelor's (29.6%) degree. The percentage of parents with only secondary education or less is higher in the case of *havo* (36.3%) than in the case of *vwo* (26.7%). These percentages have been compared to the national situation, as described in the study by Hiteq (Groeneveld, Benschop, Olvers, & Van Steensel, 2010), from which it has become apparent that children attending *vwo* more often have parents with higher education than children attending *havo*. Therefore, the study group can be considered representative of the entire population. See Figure 4 for a visualisation of the different percentages for this variable.

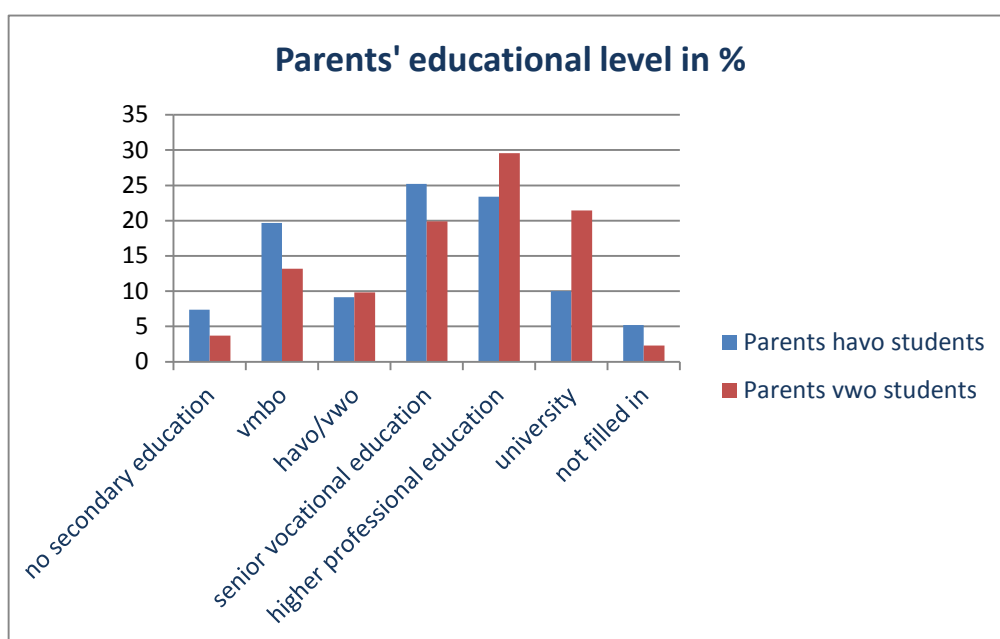


Figure 4. Parents' educational level in %

3.4 Regional spread

When selecting the study group, the primary objective was to recruit schools able to provide final examination classes *havo* and *vwo* for the taking of tests for English, French, or German, under test conditions set by the originators of the study. During the selection, it was endeavoured to select schools from every region of the Netherlands. Figure 5 represents the regional spread of the study group *havo* and *vwo*, across the northern region (including the provinces of Groningen, Friesland and Drenthe), the western region (Utrecht, Noord-Holland, Zuid-Holland and Zeeland), the eastern region (Overijssel, Gelderland and Flevoland), and the southern region (comprising the provinces of Noord-Brabant and Limburg). It can be concluded that all regions are represented, although not equally.

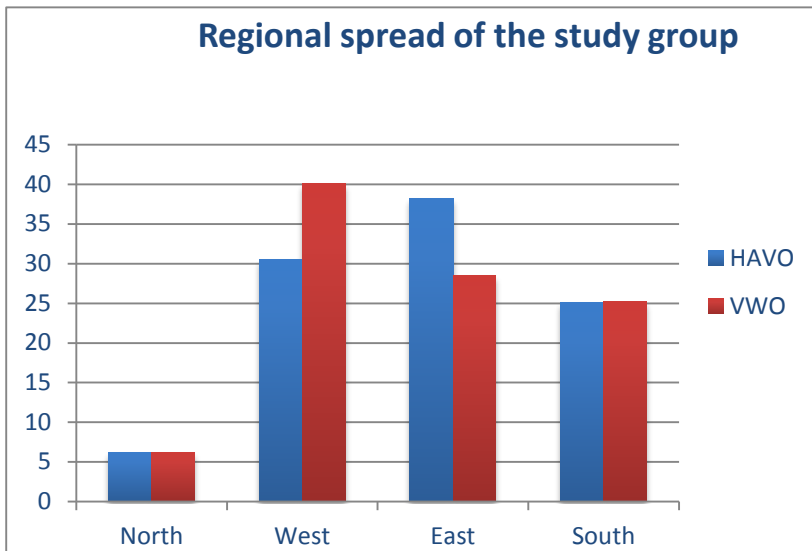


Figure 5. Regional spread of the study group

Within each region, the different languages are not represented in a balanced way. This is particularly striking for the eastern region, where *havo* is concerned. The representativeness of the study group is guaranteed, however, by comparing the characteristics described earlier. Figures 6 and 7 represent the distribution of the study group across the regions, split up by language.

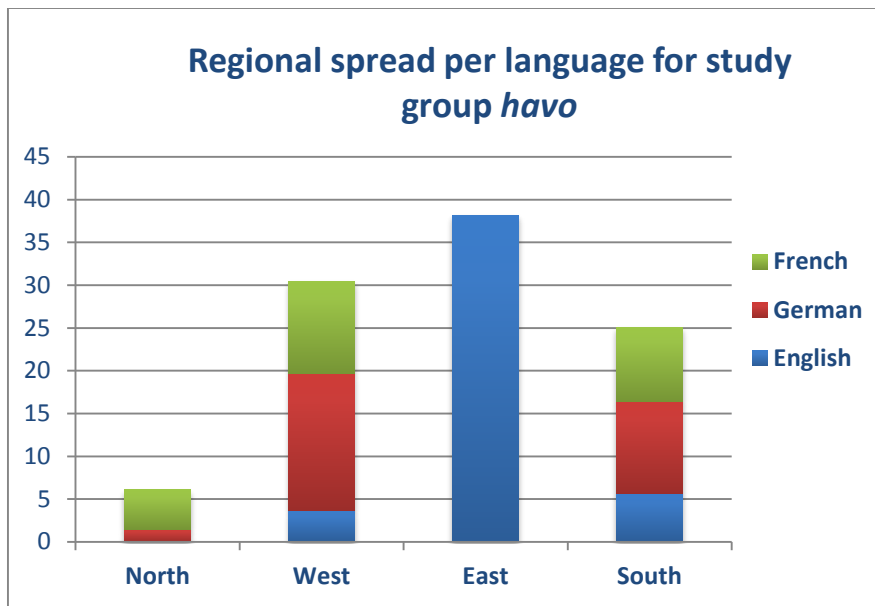


Figure 6. Regional spread per language for study group *havo*

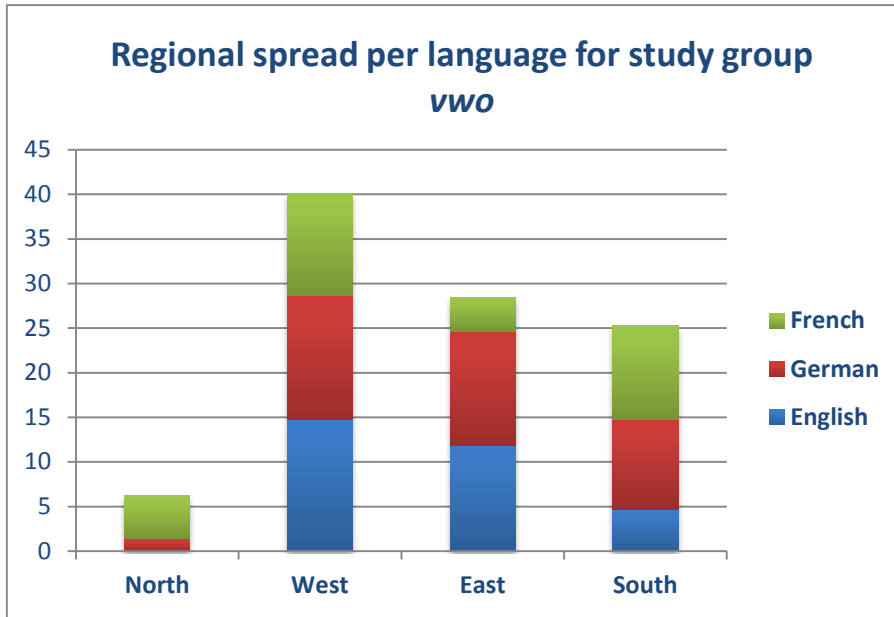


Figure 7. Regional spread per language for study group *vwo*

4. Test content and assessment

This chapter describes how the tests used in this study have been constructed. In addition, the creation of the assessment format and the procedure around anchor selection is explained.

4.1 Test content

Cito has developed writing tests for German, English, and French on two levels: viz. one for *havo* and one for *vwo*. Each test comprises several writing tasks. This has been done so that a wide range of writing activities could be included, resulting in a clear picture of what students can do when writing in English, French, or German. While constructing the writing tasks and composing the tests, the aim was to pursue as wide a variation as possible. This variation concerned the following text characteristics:

- type of text (informative, instructive, demonstrative, and narrative texts)
- text genre (letters, contributions to internet forums, reviews, e-mails, contributions to chat sessions, etc.)
- length of text (number of words to be produced)
- topic (as closely matching the social environment of the examination candidates as possible)
- CEFR domain (personal, public, educational and professional domain)
- CEFR descriptor (1. correspondence, 2. notes, messages, forms, 3. reports, 4. free writing)

In the tables 6-11, it has been indicated to what extent each writing task matches the text characteristics listed above. From these tables it also emerges that all writing tests have been composed in a varied way and that they offer a broad range of the different characteristics. The missing category is narrative texts, i.e. free writing. This has been a conscious choice, because narrative texts greatly tax the creative ability of the writer, which determines the quality of the written product.

The levels of the writing tests match the level indications that have been defined as attainment levels in the guidelines for school examinations modern foreign languages *havo/vwo* (see Chapter 1.3). According to the target levels formulated in these guidelines, the writing tests for German and French should have identical CEFR requirements. During the construction, the test constructors have somewhat deviated from these, because it was assumed that students would demonstrate better writing skills in German than in French, because of the relationship between German and Dutch. In the tables below, the CEFR level envisaged for each writing task during the construction is indicated.

Table 6. *Composition of writing tests havo English*

	writing task 1	writing task 2	writing task 3
type of text	informative	informative	demonstrative
text genre	contribution chat session	contribution musical magazine	e-mail
minimum number of words	60	60	150
topic	School matters	Pop festival	Hilarious research
CEFR domain	personal and educational domain	public domain	personal domain
CEFR descriptor	correspondence	reports	correspondence
envisaged CEFR level	B1	B1	B1

Table 7. *Composition of writing tests vwo English*

	writing task 1	writing task 2	writing task 3
type of text	informative/demonstrative	informative/demonstrative	demonstrative
text genre	letter to the editor	contribution internet forum	contribution internet forum
minimum number of words	75	75	180
topic	Trash TV	Comic strips	Deforestation
CEFR domain	public domain	public domain	public domain
CEFR descriptor	correspondence	notes, messages, forms	reports
envisaged CEFR level	B2	B2	B2

Table 8. *Composition of writing tests havo German*

	writing task 1	writing task 2	writing task 3
type of text	informative/demonstrative	informative/persuasive	informative/demonstrative
text genre	contribution chat session	promotional broadcasting text	blog
minimum number of words	50	50	130
topic	Matters of interest to young people (side job, clothing money, etc.)	Holiday job	School trip to Austria
CEFR domain	personal domain	personal and professional domain	personal and educational domain
CEFR descriptor	correspondence	notes, messages, forms	reports
envisaged CEFR level	A2+	A2+	B1

Table 9. Composition of writing tests vwo German

	writing task 1	writing task 2	writing task 3
type of text	informative	informative	demonstrative
text genre	contribution chat session	letter	contribution internet forum
minimum number of words	60	60	150
topic	Preparations visit to concert	Reaction to article in German magazine	Organ donation
CEFR domain	personal domain	public domain	public domain
CEFR descriptor	correspondence	correspondence	reports
envisaged CEFR level	B1	B1	B2

Table 10. Composition of writing tests havo French

	writing task 1	writing task 2	writing task 3	writing task 4
type of text	informative	informative/demonstrative	informative	demonstrative
text genre	notes	form	e-mail	contribution internet forum
minimum number of words	40	40	50	50
topic	Personnel seeking work	Sports	Holiday job	On strike
CEFR domain	personal domain	personal domain	professional domain	public domain
CEFR descriptor	notes, messages, forms	notes, messages, forms	correspondence	notes, messages, forms
envisaged CEFR level	A2+	A2+	A2+	A2+

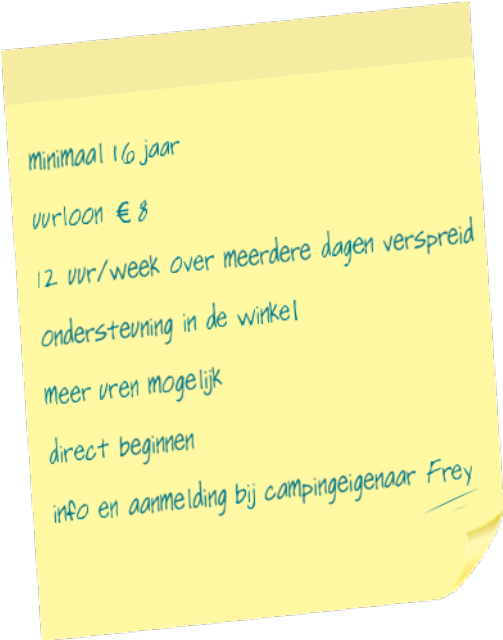
Table 11. Composition of writing tests vwo French

	writing task 1	writing task 2	writing task 3
type of text	informative/demonstrative	informative/demonstrative	informative/demonstrative
text genre	form	contribution internet forum	review
minimum number of words	60	60	120
topic	Pets	Parking problems	Holiday job
CEFR domain	personal domain	public domain	professional domain
CEFR descriptor	notes, messages, forms	notes, messages, forms	reports
envisaged CEFR level	B1	B1+	B1

The writing tasks were formulated in Dutch, sometimes expanded upon with an article or comic strip in the target language. Each task comprised a description of a situation and an assignment. The situation description was used to clarify the objective of the text and its target audience to the students. The assignment itself comprised a number of instructions regarding the content of the text. Where this study was concerned, the duration of a test equalled a period of two lessons. As an illustration, a similar assignment has been included below. The original assignments will not be published because a follow-up study may be carried out.

Description of the situation

You have accepted a holiday job for four weeks on a campsite in Spain. Pretty soon, it becomes apparent that there is more work than you can handle. Your boss wants to pin up a notice to recruit an extra employee. Because there are rather a lot of French guests, he wants to draw up the text in French. And because you have a good command of different languages, the campsite manager wants you to write this text. He has written the information to be included on a piece of paper.



Assignment

Use these instructions in a well-written and promotional piece of text. Use at least 50 words.

Figure 8. Example of a writing assignment

4.2 Assessment format

To assure reliability, the student works have been split up in separate passages, each of which corresponds to a certain sub-assignment (in the above example, the inclusion of the hourly wage is one of the sub-assignments). In the table below, the number of passages for each writing task is given.

Table 12. Number of passages for each writing task, in each language

	English	German	French
writing test havo			
writing task 1	7	6	4
writing task 2	7	7	3
writing task 3	10	9	5
writing task 4			4
total	24	22	16
writing test vwo			
writing task 1	7	7	5
writing task 2	7	5	5
writing task 3	9	9	10
total	23	21	20

Each passage is scored separately. This approach makes it possible to identify quality fluctuations within a student's work. These quality fluctuations are useful because a poor student may occasionally produce a well-written passage and a skilled student may come up with a few sentences that are not as well-rounded as may be expected.

During the assessment, two dimensions were focussed upon: communicative effectiveness and correctness. This splitting up is in line with the tradition of foreign-language tuition, in which different sub-skills are distinguished. The objective is to keep the assessment as impartial as possible, and to avoid bias of assessors when determining which sub-skill should be more severely judged. In the tables below, we have indicated which aspects should be categorised as communicative effectiveness aspects, and which as correctness aspects.

Table 13. *Aspects of communicative effectiveness and correctness*

communicative effectiveness	
content	Has the requested content element been adequately represented?
public and target orientedness	Is the linguistic usage of the text aimed at the reader and is it appropriate to the situation in which the language is used? Has the language use been adapted to the function of the text?
organisation	Has the text fragment a logical build-up?
use of words (both command and range)	Have the correct words been used? Has the student used an adequate variation of words, if applicable?
correctness	
grammar	Word and sentence structure
spelling	Is the spelling of the text correct? Note that declensions fall under grammar.
interpunction	Has correct use been made of punctuation marks and capitals?
(letter) conventions	Have, where applicable, correct conventions been used regarding lay-out, addressing, conclusion, etc.?

The assessment has taken place using a product scale by means of anchors (Kuhlemeier, 2011). It has been demonstrated that this way of assessing will lead to reliable results (Van den Bergh & Rijlaarsdam (1986), Schoonen & De Glopper (1992), Melse & Kuhlemeier (2000), Van Gelderen, Oostdam, & Van Schooten, (2011), Feenstra (2012), Bouwer a.o., in press). In this study, using a product scale meant that assessors were to compare a passage from the student's work to two other passages, taken from works resulting from the same sub-assignment, which are considered reference points – so-called anchors. For every passage, a high anchor (a relatively good work) and a low anchor (a relatively poor work) have been selected. As a result, the assessment becomes a mere sorting task. Each passage from the student's work is compared to a low and a high anchor. If the quality of the passage is lower than the low anchor, 1 score point is given. If the quality of the passage is between the low and the high anchor, 2 score points are given. If the quality of the passage is higher than the high anchor, 3 score points are given. In the analyses, 1, 2 and 3 score points are subsequently coded to 0, 1 and 2.

Each test is accompanied by a special assessment booklet. This booklet contains a low and a high anchor for each of the passages to be assessed. To illustrate this, we are presenting three sets of anchors from the assessment booklets for German, English, and French below.

Table 14. *Anchors German, English and French*

German <i>havo</i>			
Sub-assignment	Anchor	Com	Cor
Indicate why you think that working is important to you, e.g. <work experience / independence / responsibility >.	Low: Ich finde arbeite sehr gut. Es gibt erfahrung in arbeiten. Und es ist sozial, kontakt mit Menschen, Menschen helfen. Dass ist sehr schön.	1/2/3	1/2/3
	High: Ich finde arbeit sehr richtig, weil du später in dein Leben auch arbeiten musst.		

English <i>vwo</i>			
Sub-assignment	Anchor	Com	Cor
First, write an introduction to the topic.	Low: Deforestation is a main problem in the whole world. For the environment it's a disaster, for some it's their livelihood.	1/2/3	1/2/3
	High: I am writing with reference to an article that I have read from WWF and Greenpeace about deforestation. It happens more often and it destroys our world.		

French <i>vwo</i>			
Sub-assignment	Anchor	Com	Cor
Explain how you discovered the parking places.	Low: Quand j'irai au supermarché, j'ai vu des parkings pour femmes.	1/2/3	1/2/3
	High: J'ai découvert des parkings spéciaux pour femmes au supermarché à Montréal où j'ai acheté quelque chose.		

4.3 Preliminary study for the assessment format

The anchors that have been used for the assessment of the writing tests were developed on the basis of a preliminary study whereby four assessors and some 50 student works for *havo* and 50 for *vwo* were involved for each language. The preliminary study comprised two rounds. For the first round, the assessors selected a basic rendering of the passage for each sub-assignment, from a stack of some 50 student works. The starting point for this selection was the mediocre student. In other words, a rendering was selected that was obviously produced by a student who would have passed the examination by the narrowest margin. Next, around 50 student works were assessed - first for the dimension communicative effectiveness, and subsequently for the dimension correctness. The score of the basic rendering was set to 100. Next, the assessors indicated for each passage how this related to the basic rendering. The score scale could be used freely. To make the scores of the different assessors comparable, these were transformed to a standard-normal distribution after conclusion of the assessment.

Round two of the preliminary study concerned the selection of anchors for the passages to be assessed. The starting point was the basic rendering from the first round. Around a standardised score of the basic rendering, an interval was created corresponding to 33 percent of the students. Next, anchors were selected from the stack of student works that had a

standardised score (almost) identical to the lower and upper limits of the interval. The assessors' agreement was taken into account during this selection. The result of this procedure was that the assessment booklet contained anchors based on the works of dozens of students.

To illustrate the procedure above, Figures 9 and 10 present the standard-normal distribution and the interval (red/pink) for two sub-assignments (items) from the French *havo* test. As indicated earlier, it is expected that 33 percent of the students will achieve a score within the interval. The centre of the shaded red part of the interval corresponds to the score of the basic rendering. From the first graph (Figure 9) it was established that a low anchor (anchor 1) corresponding to a standardised score of -0.582 was to be looked for among the student works. The high anchor (anchor 2) should be as close as possible to score 0.289. Another thing that was established from the graph, is that it is expected that 28 percent of the students in the population will produce a passage that will be given a lower score than -0.582 (the value of the low anchor). It is also expected that 38.6 percent of the students will write a passage that will be given a score higher than the standardised one for the high anchor (0.289). The figures in Figure 10 can be interpreted in a similar way. We can observe that the anchors for the second assignment have been selected in such a way that it is expected that 47.1 percent of the students will perform lower than the first anchor, that 33.0 percent will perform between the first and second anchor, and that 19.5 percent will perform better than the second anchor. These percentages are different from those for the first assignment. In practice, this means that it will be easier for students to gain points for sub-assignment 1 than for sub-assignment 2. This variation in the level of difficulty is desirable, because the easier sub-assignments will enable students that perform relatively poorly to obtain a few score points, while the more difficult ones will provide better students with the challenge to obtain a maximum number of score points.

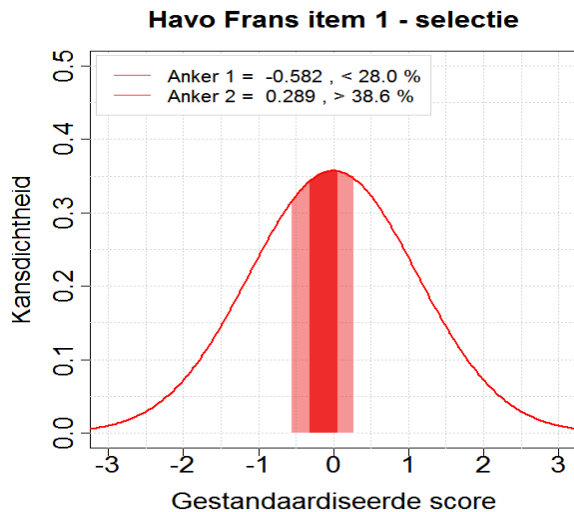


Figure 9. Probability-density function of the preliminary study for the *havo* French test, sub-assignment 1

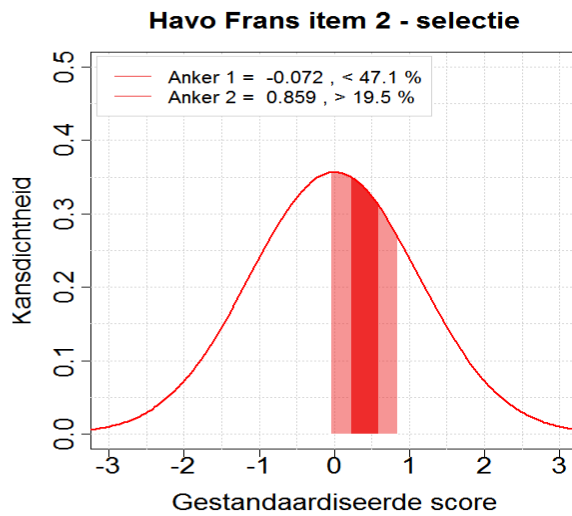


Figure 10. Probability-density function of the preliminary study for the *havo* French test, sub-assignment 2

At first, separate anchor sets (low and high anchor) were selected for both assessment dimensions - communicative effectiveness and correctness. However, this turned out to involve too large a reading load for the assessors. That is why it was decided to select a single, combined anchor set for the two dimensions. To assess a passage on both communicative effectiveness and correctness, it had to be compared to one and the same anchor set, which substantially reduced the reading load for the assessors.

5. Results

In this chapter, we present overviews of the scores that have been given to the writing products. For each test, the maximum score that could be obtained is included, as well as the minimum and maximum scores actually obtained, and the average score of the entire study group.

5.1 English

havo

Table 15 presents a brief summary of the content of the writing test for English *havo*. The test results for writing skills English *havo* have been included in Table 16. Altogether, 371 students performed this test. The maximum score that could be obtained was 96: 28 for the first and second assignments, respectively, 40 for the third assignment; fifty percent of the score concerned communicative effectiveness (content, public and target orientedness, organisation, and use of words), and fifty percent concerned correctness (grammar, spelling, interpunction, and textual conventions).

The student scores were between 8 and 96 points; the average score was 49.52 points, denoting 51.6% of the maximum score. The second and third assignments were not performed as well as the first one (assignment 1: average 55.7% of the maximum score, assignment 2: 51%, assignment 3: 49.1%).

One of the five participating schools scored considerably lower than the others (36.44 points on average).

There were no significant differences between the results achieved by boys and those by girls (49.65 vs. 49.41%).

There were no significant differences between the score for communicative effectiveness (average 25.4, or 52.9% of the maximum score) and the one for correctness (average 23.3; 48.6%).

Table 15. *Test content English havo*

	Assumed CEFR level of the writing product	Type of text	Topics
Assignment 1	B1	Chat session	Contribution chat session
Assignment 2	B1	Contribution to musical magazine	Contribution musical magazine
Assignment 3	B1	E-mail	E-mail

Table 16. *Results writing skills English havo*

	N	Min. score obtained	Max. score obtained	Max. possible score	Ave. score	% compared to max. possible score	Standard deviation
Assignment 1	371	2	28	28	15.60	55.7%	4.864
Assignment 2	371	0	28	28	14.27	50.0%	5.446
Assignment 3	371	0	40	40	19.65	49.1%	8.014
Total score	371	8	96	96	49.52	51.6%	16.306

Vwo

Table 17 presents a brief summary of the content of the writing test for English *vwo*. The test results for writing skills English *vwo* have been included in Table 18. The test has been made by 298 students. The maximum score that could be obtained was 92: 28 for the first and second assignments, respectively and 36 for the third assignment; fifty percent of the score concerned communicative effectiveness (content, public and target orientedness, organisation, and use of words), and fifty percent concerned correctness (grammar, spelling, interpunction, and textual conventions).

The student scores for English *vwo* were between 2 and 89 points; the average score was 43.34 points, denoting 48.7% of the maximum score. The third assignment was not performed as well as the first two (assignment 1: average 47.6% of the maximum score, assignment 2: 48.4%, assignment 3: 45.7%).

Two out of the six participating schools scored substantially lower than the others (27.80 and 39.43 points) and one school scored higher (49.26 points) than the average.

There were no significant differences between the results achieved by boys and those by girls. There were no significant differences between the score for communicative effectiveness (average 22.8, or 49.6% of the maximum score) and the one for correctness (average 20.7; 45%). There was a slightly better score for the first one than for the second one.

Table 17. *Test content English vwo*

	Assumed CEFR level of the writing product	Type of text	Topics
Assignment 1	B2	Letter to the editor	Trash TV
Assignment 2	B2	Contribution to internet forum	Comic strips
Assignment 3	B2	Plea	Deforestation

Table 18. *Results writing skills English vwo*

	N	Min. score obtained	Max. score obtained	Max. possible score	Ave. score	% compared to max. possible score	Standard deviation
Assignment 1	298	0	28	28	13.33	47.6%	5.592
Assignment 2	298	0	27	28	13.56	50.2%	5.495
Assignment 3	298	0	36	36	16.45	45.7%	7.379
Total score	298	2	89	92	43.34	48.7%	16.507

5.2 German

Havo

Table 19 presents a brief summary of the content of the writing test for German *havo*. The test results for writing skills German *havo* have been included in Table 20. 221 students completed this test. The maximum score that could be obtained was 88: 24 for the first assignment, 28 for the second one, and 36 for the third one; fifty percent of the score concerned communicative effectiveness (content, public and target orientedness, organisation, and use of words), and fifty percent concerned correctness (grammar, spelling, interpunction, and textual conventions).

The student scores were between 0 and 52 points; the average score was 28.44 points, denoting 32.3% of the maximum score. There were great differences in quality regarding performances: the first assignment was performed much better than the third, which scored much lower than the other two (assignment 1: average 43.6% of the maximum score, assignment 2: 35.3%, assignment 3: 22.4%).

Looking at the averages of the results for the 12 participating schools, there were two that particularly stood out, these scored 19.2 and 43.2 points, respectively; the other schools varies between 24 and 32.9 points.

With 29.6 points, girls scored significantly higher than boys (26.5).

The scores for communicative effectiveness and correctness did not display any differences between them (16.4 versus 16.5 and 35.7% versus 35.9%, respectively, of the maximum score).

Table 19. *Test content German havo*

	Assumed CEFR level of the writing product	Type of text	Topics
Assignment 1	A2+	Chat session	Matters of interest to young people (side job, clothing money, etc.)
Assignment 2	A2+	Promotional broadcasting message for personnel advertisement	Holiday job
Assignment 3	B1	Blog	School trip to Austria

Table 20. *Results writing skills German havo*

	N	Min. score obtained	Max. score obtained	Max. possible score	Ave. score	% compared to max. possible score	Standard deviation
Assignment 1	221	0	21	28	10.47	43.6%	4.3913
Assignment 2	221	0	20	28	9.88	35.3%	4.2483
Assignment 3	221	0	22	36	8.09	22.5%	4.5156
Total score	221	0	52	88	28.44	32.3%	10.5525

Vwo

Table 21 presents a brief summary of the content of the writing test for German vwo. The test results for writing skills German vwo have been included in Table 22. 365 students completed this test. The maximum score that could be obtained was 84: 28 for the first assignment, 20 for the second one, and 36 for the third one; fifty percent of the score concerned communicative effectiveness (content, public and target orientedness, organisation, and use of words), and fifty percent concerned correctness (grammar, spelling, interpunction, and textual conventions). The student scores were between 0 and 72 points; the average score was 31.28 points, denoting 37.2% of the maximum score. The first assignment was performed better than the other two; the second one was the worst of the three (assignment 1: average 43.2% of the maximum score, assignment 2: 31.8%, assignment 3: 35.6%).

There were significant differences in the averages of the 15 participating schools; these varied between 16.83 and 40.3 points.

With 34.87 points, girls scored significantly higher than boys (28.07).

There were no significant differences between the score for communicative effectiveness (average 16.64 points, or 39.6%) and the one for correctness (average 15.12; 36%).

Table 21. *Test content German vwo*

	Assumed CEFR level of the writing product	Type of text	Topics
Assignment 1	B1	Chat session	Preparations visit to concert
Assignment 2	B1	Letter to the editor of a magazine	Reaction to article in German magazine
Assignment 3	B2	Contribution to internet forum	Organ donation

Table 22. *Results writing skills German vwo*

	N	Min. score obtained	Max. score obtained	Max. possible score	Ave. score	% compared to max. possible score	Standard deviation
Assignment 1	365	0	28	28	12.10	43.2%	5.4665
Assignment 2	365	0	20	20	6.36	31.8%	4.1414
Assignment 3	365	0	33	36	12.82	35.6%	7.1660
Total score	365	0	72	84	31.28	37.2%	14.4975

5.3 French

Havo

Table 23 presents a brief summary of the content of the writing test for French *havo*. The test results for writing skills French *havo* have been included in Table 24. 188 students completed this test. The maximum score that could be obtained was 64: 16 points for the first assignment, 12 for the second one, 20 for the third one, and 16 for the fourth one; fifty percent of the score concerned communicative effectiveness (content, public and target orientedness, organisation, and use of words), and fifty percent concerned correctness (grammar, spelling, interpunction, and textual conventions).

The student scores were between 0 and 58 points; the average score was 23.02 points, denoting 36% of the maximum score. The fourth assignment was performed worse than the other three (assignment 1: average 38.4% of the maximum score, assignment 2: 40.8%, assignment 3: 38.2%, assignment 4: 27.1%).

There were large differences in the averages of the 14 participating schools; these varied between 14.63 and 34.50.

There were no significant differences between the results achieved by boys and those by girls. There were no significant differences between the score for communicative effectiveness (average 12.33, or 38.6% of the maximum score) and the one for correctness (average 11.27; 35.2%).

Table 23. Test content French havo

	Assumed CEFR level of the writing product	Type of text	Topics
Assignment 1	A2+	Notes	Personnel seeking work
Assignment 2	A2+	Form	Sports
Assignment 3	A2+	E-mail	Holiday job
Assignment 4	A2+	Contribution to internet forum	On strike

Table 24. Results writing skills French havo

	N	Min. score obtained	Max. score obtained	Max. possible score	Ave. score	% compared to max. possible score	Standard deviation
Assignment 1	188	0	16	16	6.15	38.4%	3.604
Assignment 2	188	0	12	12	4.90	40.8%	2.933
Assignment 3	188	0	20	20	7.63	38.2%	4.471
Assignment 4	188	0	15	16	4.34	27.1%	3.443
Total score	188	0	58	64	23.02	36.0%	12.382

Vwo

Table 25 presents a brief summary of the content of the writing test for French vwo. The test results for writing skills French vwo have been included in Table 26. 292 students completed this test. The maximum score that could be obtained was 80: 20 for both the first and the second assignment, and 40 for the third assignment; fifty percent of the score concerned communicative effectiveness (content, public and target orientedness, organisation, and use of words), and fifty percent concerned correctness (grammar, spelling, interpunction, and textual conventions).

The student scores were between 1 and 78 points; the average score was 38.36 points, denoting 48% of the maximum score. The first assignment was performed better than the other two (assignment 1: average 57.4% of the maximum score, assignment 2: 44.7%, assignment 3: 44.9%).

The averages of the 18 participating schools varied between 25.78 and 51.94.

With 39.70 points, girls scored a little higher than boys (35.81). This difference is not significant, however.

Students scored somewhat higher for communicative effectiveness than for correctness (average 20.20, or 50.5% of the maximum score, compared to 17.72; 44.3%).

Table 25. Test content French vwo

	Assumed CEFR level of the writing product	Type of text	Topics
Assignment 1	B1	Motivational letter	Pets
Assignment 2	B1+	Contribution to internet forum	Parking problems
Assignment 3	B1	Review on website	Holiday job

Table 26. Results writing skills French vwo

	N	Min. score obtained	Max. score obtained	Max. possible score	Ave. score	% compared to max. possible score	Standard deviation
Assignment 1	292	0	20	20	11.47	57.4%	4.536
Assignment 2	292	0	20	20	8.93	44.7%	4.666
Assignment 3	292	0	40	40	17.95	44.9%	9.321
Total score	292	1	78	80	38.36	48.0%	16.143

5.4 Comparison of results

Table 27 shows a comparison between the average scores for the different languages and sectors.

Table 27. Percentages of scores for all languages

	Ave. score	Max. possible score	% compared to max. possible score
English havo	49.52	96	51.6%
German havo	28.44	88	32.3%
French havo	23.02	64	36.0%
English vwo	43.34	92	48.7%
German vwo	31.28	84	37.2%
French vwo	38.36	80	48.0%

From the overview, it becomes apparent that the performance for the *havo* English test was considerably better than the German and French ones. While, on the other hand, *vwo* students scored a lot lower for German than they did for English and French. Only for English *havo*, there have been students who obtain maximum scores.

Where all three languages are concerned, students did not score significantly higher for communicative effectiveness or correctness, although the scores regarding the first criterion were somewhat higher in general (except for German *havo*).

The final and longer assignment was not performed as well as the other ones. This was true for all three languages and in both sectors. This assignment assumed the production of a longer text.

Only where German *havo* and *vwo* were concerned did girls score significantly higher than boys.

6. International standard-setting

To be able to determine the CEFR level of the collected written products, a standard-setting procedure was a necessary step. In this chapter, we will describe the process that led to the setting of cut-off scores.

6.1 Composition of expert panels

To be able to apply the selected method correctly, it was of great importance to make sure all members of the expert panel had a lot of knowledge and experience of the CEFR to ensure well-founded decisions could be reached. In addition, it was important to include native speakers of the target language as well as persons who work in the countries where the target language is spoken in the group of subject experts. This would increase international support for the results of the study. Thirdly, the professional background of the members of the expert panel was to be taken into account. To prevent biased interpretation of the CEFR, it was recommended to recruit panel members from different occupational groups. Finally, to participate in an expert panel for standard-setting, it was not an absolute requirement to have experience with this work. However, it was an advantage, because less time would then have to be spent on explaining the procedure. For the above-mentioned reasons, the recruitment of subject experts for the expert panels took place on the basis of the following criteria:

- a thorough knowledge of the CEFR, on account of one's own professional background;
- distribution over occupational groups, each of which is involved with the CEFR in its own way;
- as many native speakers of the target language as possible;
- preferably, and wherever possible, members of the standard setting procedure for reading and listening skills, performed by the Dutch National Institute for Educational Measurement Cito, in September of 2013 (see Chapter 1);
- representatives from the Netherlands, as well as the countries where the target languages are spoken and other European countries.

Where the latter criterion is concerned, it has to be kept in mind that English should be distinguished from German and French, because of the differences in their position within the curricula of secondary education in the European countries. Whereas English is taught as a compulsory subject in most countries across Europe, German and French are not taught in all European countries; their position within the curriculum varies from one country to the next. It was important, therefore, to make sure that all European regions were represented where English was concerned, while regarding German and French, we particularly recruited in countries where the target languages were spoken, and which also play an active role in the teaching of these languages in other parts of Europe.

The composition of the expert panels for writing skills was 17 experts English, 17 experts German, and 21 experts French. Initially, a request for participation was directed at participants in the standard setting procedure for reading and listening skills (Feskens a.o., 2014); this number was later expanded by way of a second recruitment via those experts and their own networks.

Figure 11 illustrates the origins of the three expert panels.

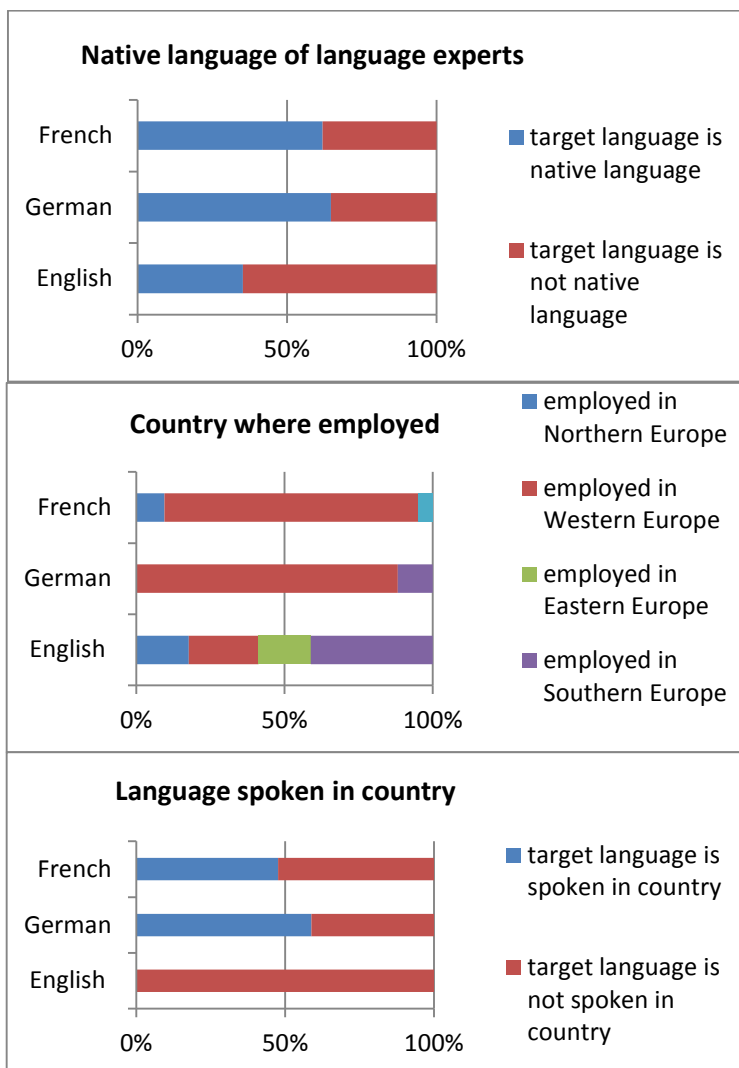


Figure 11. Origin of the members of the expert panels.

Figure 12 shows a distribution across the different occupational groups for all three languages. Many experts of French fall within the category 'other': these concern curriculum developers, teacher trainers, publishers, or public servants. Where English and German are concerned, the experts particularly concern test developers and researchers.

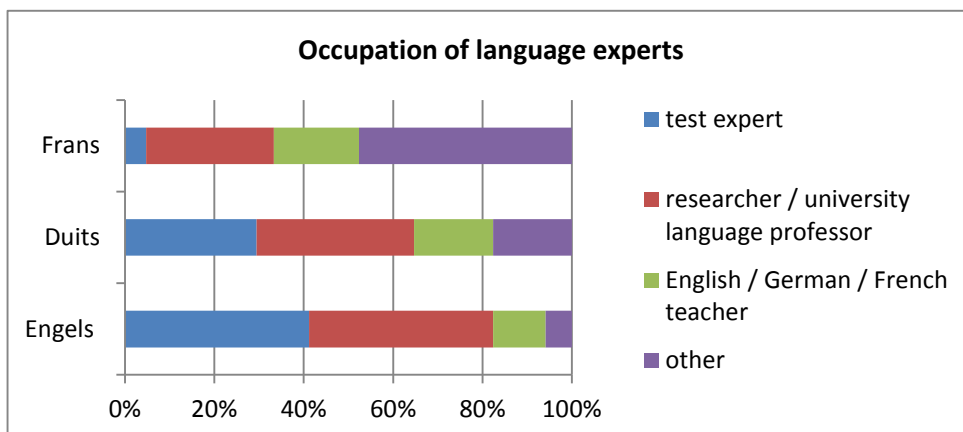


Figure 12. Occupation of language experts.

From the graphs below (Figure 13) it emerges that a vast majority had participated in similar standard-setting procedures before, such as the study regarding the standard-setting for reading and listening skills of Dutch final examination students, mentioned earlier.

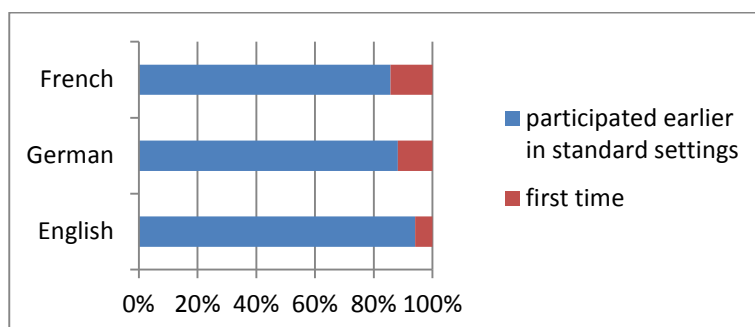


Figure 13. Earlier experiences with procedures for standard-setting.

Appendix 2 includes lists of the members of the three expert panels.

6.2 The procedure

For each language and each school type, an identical procedure was developed; i.e. for English *havo*, English *vwo*, German *havo*, German *vwo*, French *havo*, and French *vwo*.

For each procedure, 52 student works were randomly selected from the study group, with a total score between the 10th and the 90th percentile. During the selection, we made sure that each total score was represented optimally, and that as many works as possible were selected that had been assessed twice.

The writing products were distributed across the subject experts on the basis of a *balanced incomplete block design* with 13 experts and 16 works for each subject expert. Each writing product was assessed four times. In Appendix 3, the expert design for English *vwo* has been included as an example.

The standard-setting procedure that was developed for this study, comprised four phases: During the preliminary phase, the subject experts familiarised themselves with the test assignments; next, they determined which CEFR level was concerned, based on five writing products with a score of around the 50th percentile. The experts were all provided with the same works. Based on their answers, the CEFR level for which a performance standard was to be benchmarked was determined for each test.

During phase 1, the experts kept in mind the minimum requirements of the particular CEFR level, and subsequently indicated whether or not these were met where 16 writing products were concerned. These writing products were assigned to them according to a Balanced Incomplete Block Design.

During phase 2, the subject experts entered into discussion with each other. Based on the results from the first round, six student works were selected that would score around the preliminary CEFR performance standard. The degree of consensus among experts was taken into account: works were selected about which the subject experts disagreed most. These student works were then submitted to all subject experts, who were then invited to indicate for each of the six works whether or not the particular CEFR level had been achieved. Next, they came up with arguments for their decision. These arguments related to the two dimensions on which the assessment with anchors had taken place: communicative effectiveness and correctness. After this round, all arguments were summarised in an overview. The objective of this operation was to help subject experts become more aware of the way in which they had set out to work when applying their scores, thereby increasing the consensus on which student works meet the requirements of a given CEFR level. This objective was achieved. To illustrate

this, Appendix 4 includes an overview of the arguments given for one of the writing products per language.

During phase 3, the subject experts once again assessed the 16 writing products of phase 1 and either confirmed or modified their first assessment. This phase also indicated how each writing product had been assessed by each of the four experts during the first phase.

6.3 Results

Below, the results of the standard-setting procedure are described, as well as the conclusions that can be drawn from these.

Preliminary assessment

The experts were requested to determine which CEFR level was concerned, based on the five writing products with a score of around the 50th percentile. Based on their answers, it was decided which CEFR performance standard was to be determined. This could deviate from the target levels determined in 2007, which, after all, were not known to the expert panels. The results of the preliminary assessment are displayed in table 28.

Table 28. Results preliminary assessment for each language and for each school type

English <i>havo</i>	B1
English <i>vwo</i>	B2
German <i>havo</i>	B1
German <i>vwo</i>	B2
French <i>havo</i>	A2
French <i>vwo</i>	B1

Assessment and determination of cut-off scores

In Table 29, the results of the first and third assessment rounds of the standard-setting procedure have been summarised.

Table 29. Results of the standard-setting procedure for writing skills

Round	Subject	School type	Cut-off	FPR	TPR	P0
1	German	<i>havo</i>	33	0.21	0.65	0.68
1	German	<i>vwo</i>	38	0.23	0.53	0.69
1	English	<i>havo</i>	43	0.25	0.61	0.76
1	English	<i>vwo</i>	43	0.24	0.66	0.66
1	French	<i>havo</i>	21	0.21	0.65	0.72
1	French	<i>vwo</i>	38	0.21	0.55	0.61
3	German	<i>havo</i>	32	0.19	0.73	0.89
3	German	<i>vwo</i>	37	0.24	0.63	0.82
3	English	<i>havo</i>	38	0.22	0.80	0.89
3	English	<i>vwo</i>	43	0.24	0.65	0.82
3	French	<i>havo</i>	21	0.19	0.66	0.80
3	French	<i>vwo</i>	37	0.23	0.60	0.73

Column P0 gives the extent of absolute expert consensus; in other words, the extent to which the experts assigned student works to the same CEFR level. P0 has been defined as the ratio of the number of concordant assessments over the total number of assessments. Expert consensus had increased considerably during the second assessment. This applied to all languages and both sectors and can be defined as high to very high: all values were > 0.70. The columns FPR and TPR give the *False positive rate* and the *True positive rate* for the relevant cut-off score. The *False positive rate* is the percentage of students to whom the CEFR level was erroneously assigned. In other words, these are the students with a score lower than the cutting score, but who, according to the experts, did meet the requirements of the benchmarked CEFR level. The *True positive rate* is the percentage of students to whom the CEFR level was correctly assigned. These are the students with a score higher than the cut-off score, and who, according to the experts, in fact did meet the requirements of the benchmarked CEFR level. When choosing the cut-off score, a maximally high *True positive rate* and a minimally low *False positive rate* (maximum 0.25) were pursued.

Figure 14 illustrates the distribution within the population in relation to the chosen cut-off scores.

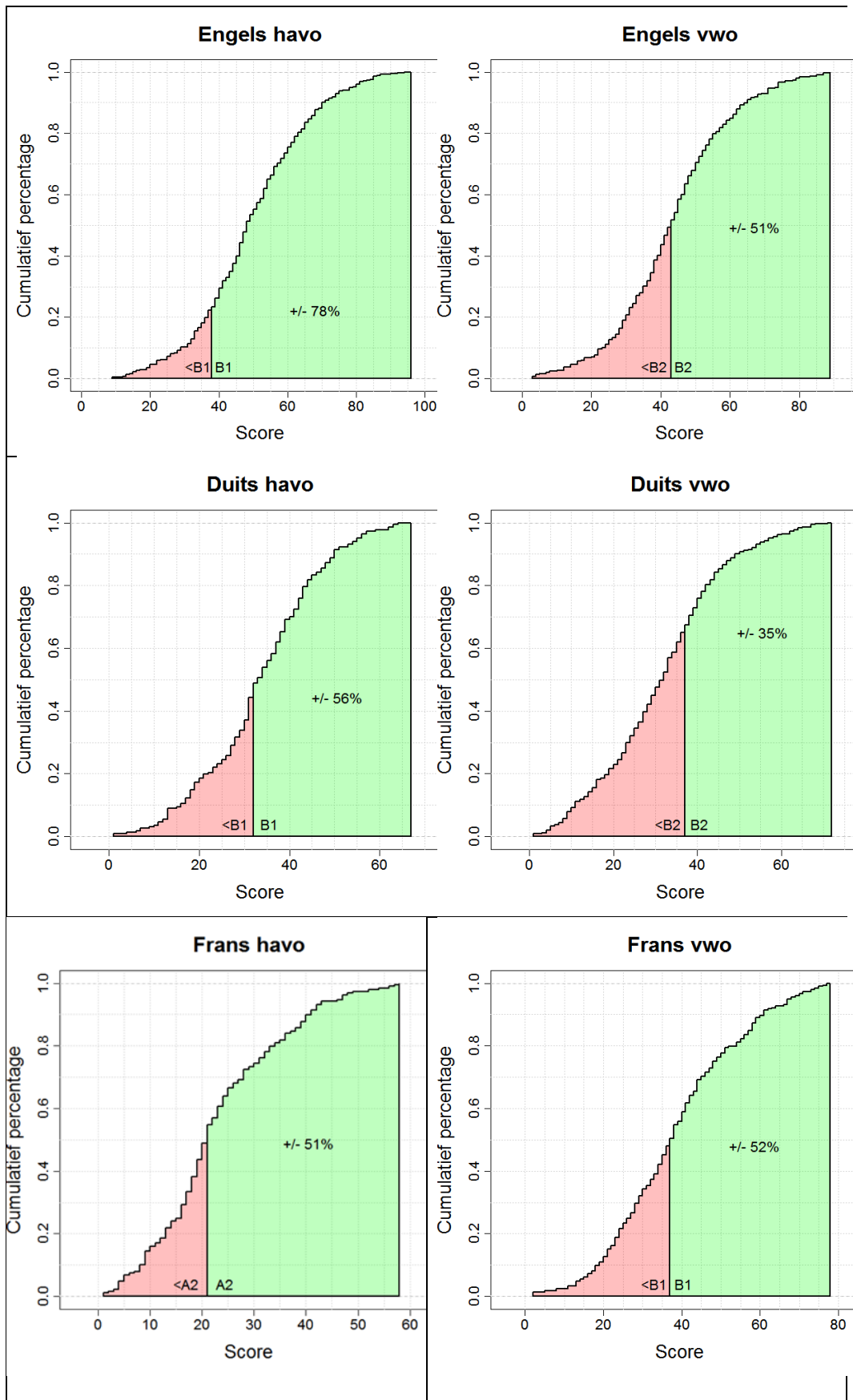


Figure 14. Laid-down cut-off scores.

The standard-setting procedure has revealed that around 78% of the study group English used for *havo* meets the performance requirements of the B1 level. Where English *vwo* was concerned, around 51% of the study group met the performance standard for the B2 level. B1 and B2 are also the levels that were used as a starting point for the construction of the tests for English, for *havo* and *vwo*, respectively.

Where German *havo* was concerned, the CEFR performance standard studied was B1. This was achieved by approximately 56% of the study group. For German *vwo*, the standard-setting procedure resulted in the performance standard B2. This level was achieved by only 35% of the study group. The writing assignments for German *vwo* were partly based on B1 (the first two assignments) and partly on B2 (third assignment).

The chosen performance standard for French *havo* was A2; 51% of the study group achieved this. A similar percentage of *vwo* students (52%) achieved the B1 standard.

It could also be that a part of the number of students who achieve a standard, also meet the requirements of the level above it. It may be possible, for example, that some of the total number of students who achieved the B1 level for English *havo* - almost 78% - also meet the requirements of B2 where language command is concerned. However, the tests and procedures that were used do not make it possible to determine this. The same applies to the levels below. It is not possible to determine which percentage of students who fail the B2 standard for German *vwo*, would, in fact, have achieved the B1 level.

Table 30 summarises the results of the standard-setting procedure and compares these to the target levels that have been laid down in the guidelines for school examinations *havo/vwo*.

Table 30. *CEFR performance standards and percentages of students who meet these*

	Target level CEFR according to School Examination guidelines	Envisaged CEFR level of the writing assignments used in this study	CEFR performance standard determined in this study	Percentage of students achieving the CEFR performance standard
English <i>havo</i>	B1	B1	B1	77.63
English <i>vwo</i>	B2	B2	B2	50.67
German <i>havo</i>	A2+	A2+ and B1	B1	55.66
German <i>vwo</i>	B1	B1 and B2	B2	35.07
French <i>havo</i>	A2+	A2+	A2	51.06
French <i>vwo</i>	B1	B1 and B1+	B1	52.05

In Chapter 8, we will go further into the conclusions that may be drawn from these results.

7. Results teacher interviews

To obtain some insight into the experiences of teachers with the teaching of writing skills, interviews were conducted in December 2013 with teachers from four of the participating schools. The questions concerned the factors affecting the learning results (both positively and negatively) and the testing and assessment methods used for writing skills. In this chapter, the results of the interviews with the participating schools are presented for each individual question.

7.1 Selection of schools

To conduct the teacher interviews, four schools have been approached because of their striking scores –these deviated from the average either positively or negatively. This particularly concerned the languages German and French. One of the schools had scored higher than the average of the entire study group, where English, German and French were concerned, and regarding both *havo* and *vwo*. The other three schools showed more differentiation in the results: one of the schools had participated only with French *havo/vwo* and German *havo/vwo*, whereby the scores for French *havo* were way above average, while the others remained just below average. A second school showed higher than average scores for German *vwo*, while remaining way below average where the other languages were concerned. The latter had scored way above average for both French *vwo* and French *havo*.

7.2 Results of the interviews

Teachers of two of the above-mentioned schools attended a focus-group meeting in Utrecht. Five teachers of the one school and one of the other were present. One school was visited before conducting the interview. Five teachers were present during this visit. Of one school, three teachers replied to the questionnaire in writing.

Below, we will discuss the answers of the schools, by theme. Group A comprises teachers whose students scored above average (twelve teachers) and Group B comprises teachers whose students scored below average (two teachers). Unfortunately, Group B only concerns a small number of teachers. Apparently, teachers from schools where scores deviated negatively from the average were less willing to participate in the interviews.

Factors affecting the learning results

Could you indicate three factors that have particularly affected the results for your subject where your school is concerned?

The following positive factors were mentioned:

- Rationale: this was given by one teacher from Group A. This teacher indicated having developed a vision document in which CEFR plays an important role.
- Teacher role: all teachers in Group A agreed that a good teacher will affect learning in a positive way. According to them, a good teacher is an enthusiastic one, will continually develop in his discipline, speaks the target language during the lesson, collaborates well with colleagues, will provide his students with high-quality materials, and coaches his students.
- Resources and materials: the teachers in Group A mentioned:

- a. a good teaching method in which skills form the backbone;
 - b. in addition, extra materials (if necessary self-developed) focussed on the various skills and grammar.
- Teaching content: three teachers from Group A mentioned offering specific examination idiom. According to these teachers, focussing on the learning of important words would support the different skills tested in the examination.
- Learning activities: teachers from Group A listed:
 - a. internationalisation activities;
 - b. activating work forms;
 - c. having students correct the writing products together;
 - d. giving feedback during the lesson and allowing students to immediately apply this in practice. One of the teachers from Group B indicated a desire to give feedback regarding student works during the lesson, but regretted not having sufficient time for this. He/she is considering the deployment of *peerScholar*, which is an online tool to enable students to work, learn, and reflect together.
- Learning environment: one teacher from Group A works in a computerised classroom. This allows the giving of feedback during the lesson. Another teacher from Group A gave a very positive reaction to the fact that their school had been equipped with ultramodern classrooms a few years ago. Discussing the works with the whole class using a multimedia projector is considered highly efficient and effective.
- Learning targets: one teacher from Group A mentioned this point. He/she considers it positive to use structured frameworks to prepare students for the formulation of an adequate response when making writing assignments.
- Time: this was given by one teacher from Group B. For the past two years, *vwo* upper secondary students had 3 lessons per week.
- Grouping arrangements: one teacher from Group A mentioned this point. According to him/her working in smaller classes (with up to 24 students) will raise effectiveness.
- Assessment: strikingly, this point was not mentioned by any of the teachers.

What do you feel are the bottlenecks regarding the teaching of your subject?

The following bottlenecks were mentioned:

- Time: this category was unanimously mentioned as a bottleneck threatening good teaching. The factor time was explained in different ways:
 - a. One teacher in Group A mentioned that two lessons per week are insufficient for *vwo*. On the other hand, the four hours they had in *havo* has a positive influence. Also, a teacher from Group B indicated a lack of contact time to allow him/her to enter into the subject matter more deeply.
 - b. Teachers from Group A indicated that schedule problems may result in less effective lessons:
 - double periods at the end of the day,
 - lessons at inconvenient moments during the day,
 - scheduling (all) languages consecutively.
 - c. One teacher from Group A indicated having insufficient time to develop additional material.
- Assessment: Here, Group A mentioned the necessity to attune the Programmes of Testing and Conclusion (*Programma's van Toetsing en Afsluiting*, PTA) to the different languages to obtain a balanced distribution of skills across the year (not the same skill at the same moment for the different languages). In addition, one teacher from Group A mentioned the rather too free conditions according to which school examinations can be developed and organised, which

he/she considered a major pitfall for all schools in the Netherlands. This produces great differences among schools, resulting in students with highly different levels of knowledge and skills entering the next phase of their educational career.

- Resources and materials: the teachers in Group B mentioned these. The writing assignments from the methods are too closed and their level is not high enough. They are insufficiently geared to the assignments for international certificates.

The following questions were asked:

- a. how to move from closed to open assignments?
- b. how to teach students a good vocabulary for their writing skills?
- c. why should/should not a dictionary be allowed during a writing-skills test?

- Learning activities: teachers of Group B desire more opportunities to put what has been learned into practice. This includes internationalisation activities.
- Grouping arrangements: teachers from both Group A and Group B mentioned this. Working with large groups, particularly where English is concerned, interferes with the checking of the student works during the lesson together. As a result, it is not possible for the students to practise with feedback.
Teachers from Group A also indicated their wish to work on a more individual basis more often, in order to make use of the student interests and also to attract more science students.
- Rationale: teachers from Group B want a rationale that is not just outcome-oriented.

In addition to the bottlenecks listed above, teachers from Group A have also indicated the following needs:

- languages should be more visible in the school;
- languages should be higher on the agenda of the Ministry of Education (because this is presently not the case, students are less attracted to languages);
- the CEFR should, like the reference frameworks for Dutch language and arithmetic, be a discussion item in the Netherlands;
- university level command and scientific orientation - to help students make the most of their abilities.

A young teacher English/German/French is hired in your school. What would be your most important tip for him/her?

Below, all tips mentioned are listed. All tips, except for the last one, were given by teachers from both Group A and Group B:

- introduce the school rule *target language is language of tuition* – known as the *doeltaal-voertaal* system in Dutch (a native speaker as a teacher will make things easier or, in any case, more natural);
- stimulate students to practise speaking the target language (for example, ask them to record their own voice for an hour each week and require students to answer questions in the target language during a lesson);
- use single-language methods to facilitate the use of the target language as the language of tuition;
- continue to develop yourself, keep up with the times, renew your assignments, invest a lot of time in the preparation of lessons;
- learn to develop your own materials, including good writing-skills assignments;
- dare to let go of existing methods, add extra material for literature, writing skills, and speaking skills;
- ask more experienced teachers a lot of questions and use their feedback in your daily practice;
- make sure you remain the owner of the materials used;

- ensure communicative language tuition, pay more attention to skills than to grammar;
- stimulate students to practise speaking as well as reading. Students should be allowed to make their own mistakes a number of times; this way, they will learn better;
- provide students with feedback on their work and allow them to correct their own mistakes;
- organise exchange contacts for students (e.g. a correspondence project with a school abroad);
- stimulate students to make a portfolio (is convenient for follow-up education, will stimulate independence, and it fits in with career-orientation and career-guidance efforts);
- do not aim too high. Be aware of what is realistic. Allow students to celebrate successes;
- Stimulate independent learning and help students to critically view their own work;
- do not use a dictionary. Teach students to trust their own knowledge. Discuss this with all departments and all languages (this tip originated from teachers from Group A).

Assessment

What should be tested regarding writing skills?

The schools from Group A feel that grammar and vocabulary should be tested integrally. A student should be able to make open writing assignments that are linked to can-do statements. One teacher indicated that knowledge is no longer tested during the first stage of secondary education and the teachers of one school test so-called *chunks*, which are memorised pieces of sentences, at lower levels.

The teachers from Group B do still test grammar and vocabulary separately in intermediate tests (30% of the final mark). These teachers feel that they should stimulate correct use of language. Depending on the stage of the learning process, they work up to the free assignments that offer students a lot of autonomy.

How do you test writing skills?

- Content
- Form

Teachers from Group A make use of :

- a. Cito SE tests,
- b. Cito CEFR writing tests (these are also used for the training of writing skills),
- c. self-developed tests, building up from closed assignments towards more open ones, and which include formal and informal letters, e-mails, and, where English is concerned, essays.

When developing open assignments, teachers have indicated that they find it important to gear the content to the students' perception of the world. The assignments should be practical, realistic and fun. They prefer to use authentic material.

Teachers from Group B develop their own tests. These concern short texts for the first stage of secondary education and build up to longer ones (>300 words) for the upper secondary education.

Text types used include letters (formal and informal ones), a response to a topic on the internet or in a magazine, diary entries, a description of hobbies, and argumentations.

Where the training for writing skills is concerned, the assignments from the method are used and during one period a reader is used.

- Frequency

Teachers from Group A emphasise the importance of formative testing of writing skills. The idea is that practice makes perfect. Students should know where they stand within the learning process. In many cases, free writing assignments, Cito CEFR writing tests, and old SE tests are given to prepare the students for the SE, so that they know what type of test to expect. In two schools, a (digital) portfolio is used.

Teachers have a different way of handling the frequency of the writing-skills tests:

- one teacher will test writing skills in each course year,
- another teacher will test once in forms 4 and 5, and twice in form 6,
- teachers from another school prefer to subject their students to multiple smaller writing-skills tests.

One teacher from Group B will set a test every period using the writing assignments from the method, while dedicating one period to give extra attention to writing skills by means of a reader developed by the teacher him/herself.

- Assessment models

The teachers use both self-developed models and assessment models that have been supplied by Cito to accompany their tests.

All assessment models will assess according to the following criteria:

- content;
- form (grammatical correctness, spelling, interpunction and layout, use of words);
- coherence and cohesion.

The Cito assessment models lack a criterion to reward creative students for their efforts. The self-developed assessment models do, however, present the opportunity to award extra points to originality and texts that flow smoothly.

According to the teachers from Group A, the latter will work fast and reliably. This reliability is further increased by frequent consultation with colleagues regarding content. They are also used to correcting a sample of each other's student works, and thoroughly discuss the differences in assessment.

One teacher from Group B does not feel that the assessment method according to the criteria mentioned above is sufficiently objective. It is not known whether he/she is involved in frequent discussions with colleagues regarding the tests made, like the teachers of the successful schools.

Why do you choose for this method?

The schools that use the Cito CEFR writing tests have indicated two reasons for doing so:

- These tests enable the setting up of a learning continuity pathway for writing skills from A1 up to and including the level of the final examination.
- It takes less time to use a Cito test than to develop a good test oneself.

One teacher from Group A indicated that the writing of letters only is a limiting factor, because the structure is rather pre-programmed. This does not really teach writing skills as such. He/she mentioned another disadvantage of the Cito tests, which involved the risk of these becoming more of a translation assignment, which has little to do with writing skills.

The teachers from Group B have indicated that the tests used keep in mind the different aspects of writing skills and that their build-up runs parallel to the language acquisition in terms of the CEFR. According to them, an assessment according to the criteria mentioned earlier is possible within the available time span.

This year, will you be using a similar writing test to the one developed by Cito for this study project?

Teachers from two schools from Group A have indicated their preference to work with more loosely set-up, open tests and will continue to use the self-developed tests. Another teacher from Group A has indicated his/her preference to work with Cito tests, because time is short. Teachers from the third school from Group A have indicated their intention to continue to work with this type of test, because of the communicative nature of the test, with open, free assignments.

Teachers from Group B have indicated their intention to continue using the same tests they were submitting their students to before they got involved in this study.

In this project, two assessment methods have been discussed: the Cito assessment model and the 'anchor method'). What is your opinion of these two methods?

The Cito assessment model has already been judged in an earlier paragraph.

The opinions about assessment on the basis of anchors are divided:

- One teacher from Group A has indicated that he/she considered it an interesting method and proposes that students select their own anchors, based on the assumed level.
- Another teacher from that group did not think it would be a suitable method for school, but does believe it will be useful in studies and standardisation procedures.
- A third teacher from the same group did not consider this method practicable and feels that the assessment would be rather too subjective.
- The teachers from Group B do not consider assessment by means of anchors suitable for use in schools. One of them mentioned the alleged subjectiveness of the method.

What is your golden tip for a good writing-skills test?

Below are the golden tips offered by the interviewees:

- Creativity.
- Variation.
- More humour in the test.
- Interesting topics.
- Topics that are close to the students' perception of the world/contemporary.
- Have the test checked by a native speaker.
- Mixture of close and open assignments (to be able to assess students' abilities during the first stage of secondary education and to determine whether to give them *havo* or *vwo* advice).
- A test should completely fit in with the language-acquirement level of the student.
- In the examination classes, the writing tests should be as open as possible.

8. Conclusions

The results of the standard setting made it possible to formulate an answer to the research question. A quick reminder, the research question was formulated thus:

Which CEFR level is attained by students upon concluding their havo and vwo education, with regard to writing skills in English, German, and French?

In the table below, we have summarised the most important information from this study in an overview.

Table 31: *Cut-off scores of the standard-setting procedure for writing skills and the percentage of student works that meet the performance standard of the chosen CEFR level.*

Subject	School type	CEFR level for which the performance standard has been determined	Size of study group	Max. possible score	Number of score points corresponding to the CEFR performance standard	Percentage of students achieving the CEFR performance standard
English	<i>havo</i>	B1	371	96	38	77.63
English	<i>vwo</i>	B2	298	92	43	50.67
German	<i>havo</i>	B1	221	88	32	55.66
German	<i>vwo</i>	B2	365	84	37	35.07
French	<i>havo</i>	A2	188	64	21	51.06
French	<i>vwo</i>	B1	292	80	37	52.05

Based on these results, we can observe the following.

English

In case of *havo*, B1 is easily achieved. It was the highest result of the six groups. Further study into a B2 performance standard should indicate the performance of *havo* students at a higher level.

In *vwo*, around half of the study group achieved B2. This relatively low percentage partly explains the writing problems in English encountered by Bachelor and Master students during their university education (see SLO, 2008 and Wierda-Boer, 2008).

German

The chosen performance standard B1 for *havo* was achieved by approximately 56% of the study group. This agrees with the assumption during the construction of the test assignments that students would demonstrate better writing skills in German than in French (Chapter 4). However, the performance standard B2 for *vwo* was achieved by only a minority. Further study into a B1 standard should demonstrate how students of German *vwo* perform in relation to the underlying level.

French

In case of French *havo*, the A2 performance standard was achieved by just over half of the study group.

The same applied to the B1 performance standard in case of French *vwo*.

From the results above, it appears that the differences in level between *havo* and *vwo* are quite small where German is concerned. The same is true, although to a lesser extent, regarding English. However, the available information does not allow any conclusions about the distance in performance between *havo* and *vwo*. To determine the difference in level between the two sectors, they should both be subjected to the same measurement. The CEFR performance standards that have been determined correspond to the minimum performance requirements for the CEFR level concerned. The higher the level, the more acquisition time is needed to go through the full band width of it and reach the next level (a.o. Tschirner, 2009).

8.1 The level of writing skills *havo/vwo* in English, German and French

How are we to interpret the results of this study? Are the learning outcomes of FL education satisfactory? Or are improvements desirable? When interpreting the study results, we assumed that 75% of the students should be able to achieve the target level. We have derived this standard from the one used by the Expert Group Learning Continuity Pathways for Language (2008) as a standard indication for the fundamental quality of education. At the same time, we assumed that the target levels that have been formulated for FL can be compared with the fundamental levels formulated for Dutch language by the Expert Group. Target levels, in case of the examination programme for modern foreign languages, indicate realistically attainable final levels for average students under good teaching conditions. This definition differs from the one used for the reference levels for Dutch language (Expert Group Learning Continuity Pathways for Language and Arithmetic, 2008). There the term target level - commonly called the T-level, in Dutch S-level - applies to a student who already operates on a level higher than the foundation or basic one - commonly called the F-level that applies to him or her. However, target levels for modern foreign languages should, in principle, be attainable to all students and therefore correspond to the F-level for Dutch language.

If we apply the 75% standard to the results of our study, the following may be concluded. Only for *havo* English did over 75% of the students attain the CEFR performance standard. For the other languages and sectors, these percentages are between 50% and 56%. With the exception of German *vwo*, where the percentage was even lower: only a minority (35%) achieved the performance standard. This means that, except for *havo* English, the 75% standard was not achieved for any language and in any sector.

The percentages resulting from the standard-setting procedure, therefore, indicate that learning outcomes of FL writing skills allow room for improvement.

As opposed to English and French, the CEFR performance standards for German deviate from the target levels determined for *havo* and *vwo* in 2007 (see Table 32). A similar deviation occurred in the test construction, as well. As has been explained in Chapter 4.1, it was assumed that Dutch students more easily acquire the German language, because it is more similar to Dutch than French. This has been the reason for a level differentiation between the German and French tests.

Table 32. Target levels *havo/vwo* and CEFR performance standards

Subject	School type	CEFR target level according to School Examinations guidelines	CEFR performance standard
English	<i>havo</i>	B1	B1
English	<i>vwo</i>	B2	B2
German	<i>havo</i>	A2+	B1
German	<i>vwo</i>	B1	B2
French	<i>havo</i>	A2+	A2
French	<i>vwo</i>	B1	B1

8.2 Evaluation of tests and assessment methods used

The tests that have been developed for this study, turned out to deviate from the daily practice in some schools. The differences related to the chosen text types as well as to the fact that dictionaries were not allowed. The teachers interviewed were positive about both.

The greater variation in text types, including contributions to blogs and chat sessions in addition to essays and formal letters, was more in line with the situations students may come across in real life and in which they need to communicate in a foreign language. Moreover, the command of writing skills is demonstrated more clearly, while also allowing a CEFR level to be linked to the results. The writing assignments used in this study could have contained a few more open tasks, allowing students to demonstrate their command of structural as well as content aspects corresponding to the particular type of text.

Teachers assumed that not having been able to use dictionaries did not have any negative effect on the student performances. In addition, many teachers were convinced that if no tools were allowed, this would make students more motivated to work on their vocabulary. It makes sense to go into this more deeply, to find out if this is, in fact, the case.

In the opinion of many of the participating teachers, assessing student works using a product scale with anchors takes up a lot of time (too much, even). Although highly suitable for research activities, they believed these are less suitable for educational purposes. However, it is not clear whether this method is actually more time-consuming than other assessment methods. It should be studied how things may be optimised for the anchor method to be made suitable for an educational situation. After all, in case of a *high-stake* test - SE tests are a part of the final examination - a reliable assessment is extremely desirable. During the interviews, the opinions about this method were also divided; if it were used for educational practice, modifications would certainly be required. It has also become clear that intensive practice is a prerequisite if these new types of assessment are to be implemented.

8.3 Implications for writing tuition in English, German and French

The teachers who were interviewed have indicated a number of bottlenecks that limit optimum learning outcome where writing tuition in modern foreign languages is concerned. These relate to different elements of the curriculum, particularly to time, assessment, grouping arrangements, and teaching materials.

In addition, the teachers indicated several success factors and gave suggestions to reinforce the learning effect. These concern virtually all elements of the curriculum; in particular emphasising the teacher's role in stimulating and monitoring the learning process.

Evaluating activities on a broader scale - including field consultations, lesson observations, and pilots - should provide evidence of the effects of these aspects on the learning outcome, in order to arrive at effective suggestions for optimisation of writing tuition in the modern foreign languages. Examples include: feedback activities (*assessment for learning*), improving the

assessment skills of teachers, and setting up forms of monitoring regarding the quality of testing and assessment.

At the same time, both linguistic and contextual/communicative aspects of writing skills should be given equal attention when developing the curriculum for writing tuition in FL, as mentioned earlier in Chapter 1. Attainment targets, learning activities, and test forms should be in line with this.

As emphasised by the Dutch Advisory Council for Education (2006), there is a growing need for evidence of the effectiveness of newly implemented learning activities and methods in the context of the growing emphasis on quality and accountability in education. The results of this study illustrate the present situation and may be used as a benchmark for interventions that are aimed at improvement of performance and measurement of their effects.

References

- Akker, J. van den (2003). Curriculum perspectives: An introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (eds.), *Curriculum landscapes and trends* (pp. 1-10). Dordrecht: Kluwer Academic Publishers.
- Association of Language Testers in Europe (2011). *Manual for language test development and examining for use with the CEFR*. Strasbourg: Council of Europe.
- Bergh, H. van den, & Rijlaarsdam, G. (1986). Problemen met opstelbeoordeling? Een recept. [Issues with essay evaluation? A recipe.] *Levende Talen*, 413, 448–454.
- Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56, 137-172.
- Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (in press). Effect of genre on the generalizability of writing scores, *Language Testing*, 1-28.
- CBS Statline (2014, 2 mei): VO; leerlingen, onderwijssoort, leerjaar, leeftijd. Obtained 8 July 2014, from <http://statline.cbs.nl/statweb/dome/?TH=3430&LA=nl>
- Council of Europe (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2009). *Manual for relating language examinations to the CEFR*. Strasbourg: Language Policy Division.
- English Proficiency Index (2013). Consulted on 30 June 2014: www.ef.com/epi
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (Expert Group Continuous Curricular Pathways for Language and Arithmetic) (2008). *Over de drempels met taal en rekenen: hoofdrapport van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen [About barriers in language and arithmetic; main report by the Expert Group Continuous Curricular Pathways for Language and Arithmetic]*. Enschede: SLO.
- Feenstra, H. (2012). *De betrouwbaarheid van een beoordelingsschaal voor schrijfproducten [The reliability of an assessment scale for writing products]*. Poster presentation, Onderwijs Research Dagen 2012, Wageningen.
- Feskens, R., Keuning, J., Til, A. van, & Verheyen, R. (2014). *Prestatiestandaarden voor het ERK in het eindexamenjaar. Een internationaal ijkingsonderzoek (rapportage) [Performance standards for the CEFR in the final examination year. An international calibration study (report)]*. Arnhem: Cito.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.

- Gelderen, A. van, Oostdam, R., & Schooten, E. van (2011). Does foreign language writing benefit from increased lexical fluency? Evidence from a classroom experiment. *Language Learning*, 61, 281-321.
- Groeneveld, M.J., Benschop, M., Olvers, D., & Steensel, K. van, (2010). *Ouders@havo/vwo*. Hilversum: Hiteq.
- Hambleton, R.K., Jaeger, R.M., Plake, B.S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24, 355-366.
- Hulstijn, J.H. (2007). The shaky ground beneath the CEFR: quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal* 91-4, 663-667.
- Jaeger, R.M. (1989). Certification of student competence. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 485-514). New York: American Council on Education and Macmillan.
- Kaftandjieva, F. (2004). Standard-setting. In: *Reference Supplement to the preliminary pilot version of the Manual for Relating Language examinations to the CEF* (Section B). Strasbourg: Language Policy Division.
- Liemberg, E., & Meijer, D. (2004). *Taalprofielen [Language profiles]*. Enschede: NaB-MVT.
- Kordes, J., & Gille, E. (2012). *Vaardigheden Engels en Duits van Nederlandse leerlingen in Europees perspectief [Skills English and German regarding Dutch students in a European perspective]*. Arnhem: Cito.
- Kuhlemeijer, H. (2011). Het construeren van praktijktoetsen [The construction of practical tests]. In P. Sanders, *Toetsen op School* (125 – 141). Arnhem: Cito.
- Meijer, D., & Fasoglio, D. (2007). *Handreiking schoolexamens moderne vreemde talen havo/vwo - Engels, Frans, Duits [Guidelines school examinations for modern foreign languages in havo/vwo - English, French, German]*. Enschede: SLO.
- Melse, L. (1990). *Schrijftoetsen voor de moderne vreemde talen [Writing tests for the modern foreign languages]*. Apeldoorn: Van Walraven.
- Melse, L. & Kuhlemeier, H. (2000). *Beoordeling van schrijfproducten met en zonder ankers. Is er verschil? [Assessment of writing products with and without anchors. Is there a difference?]*. Arnhem: Cito.
- Onderwijsraad [Dutch Advisory Council for Education] (2006). *Naar meer evidence based onderwijs [Towards more evidence-based education]*. The Hague: Onderwijsraad (Dutch Advisory Council for Education).
- Sanders, P. (red.) (2013). *Toetsen op school [Testing in school]*. Arnhem: Cito.
- Schoonen, R., & De Glopper, C. M. (1992). Toetsing van schrijfvaardigheid: problemen en mogelijkheden [Testing of writing skills: problems and opportunities]. *Levende Talen*, 470, 187-195.

SLO (2008). *Filling the Gap. Voorbereidend wetenschappelijk Engels in de tweede fase [Pre-university English during upper secondary education]*. Obtained 30 June 2014:
<http://fillingthegap.slo.nl>

Til, A. van, Beeker, A., Fasoglio, D., & Trimbos, B. (2011). *Toetsen en beoordelen met het ERK [Testing and assessing using the CEFR]*. Arnhem/Enschede: Cito/SLO.

Tschirner, E. (2009). Das ACTFL OPI und der Europäische Referenzrahmen. In: *Babylonia 2/05* (pp. 50-55). Comano (CH): Stiftung Sprachen und Kulturen.

Weir, C.J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. In *Language Testing* 22, 281-300.

Wierda-Boer, H. (2008). *Aansluiting Engels. Een onderzoek naar de aansluitingsproblematiek van het vwo-vak Engels met de universiteit [Attunement of English. A study of the attunement problems of the vwo subject of English to university level]*. Nijmegen: IOWO.

Appendices

Appendix 1 Balanced Incomplete Block Design

student work	assessor 1	assessor 2	assessor 3	assessor 4	assessor 5	assessor 6	N
1	0	1	0	1	0	0	2
2	0	1	0	1	0	0	2
3	0	0	0	1	1	0	2
4	0	0	0	1	1	0	2
5	1	1	0	0	0	0	2
6	1	1	0	0	0	0	2
7	1	0	1	0	0	0	2
8	1	0	1	0	0	0	2
9	0	0	0	0	1	1	2
10	0	0	0	0	1	1	2
11	0	0	1	0	1	0	2
12	0	0	1	0	1	0	2
13	1	0	0	0	0	1	2
14	1	0	0	0	0	1	2
15	0	1	1	0	0	0	2
16	0	1	1	0	0	0	2
17	0	0	0	1	0	1	2
18	0	0	0	1	0	1	2
19	0	0	1	1	0	0	2
20	0	0	1	1	0	0	2
21	0	1	0	0	0	1	2
22	0	1	0	0	0	1	2
23	1	0	0	0	1	0	2
24	1	0	0	0	1	0	2
25	0	1	0	0	1	0	2
26	0	1	0	0	1	0	2
27	1	0	0	1	0	0	2
28	1	0	0	1	0	0	2
29	0	0	1	0	0	1	2
30	0	0	1	0	0	1	2
31	0	1	0	1	0	0	2
32	0	1	0	1	0	0	2
33	0	0	0	1	1	0	2
34	0	0	0	1	1	0	2
35	1	1	0	0	0	0	2
36	1	1	0	0	0	0	2
37	1	0	1	0	0	0	2
38	1	0	1	0	0	0	2
39	0	0	0	0	1	1	2
40	0	0	0	0	1	1	2
41	0	0	1	0	1	0	2
42	0	0	1	0	1	0	2

43	1	0	0	0	0	1	2
44	1	0	0	0	0	1	2
45	0	1	1	0	0	0	2
46	0	1	1	0	0	0	2
47	0	0	0	1	0	1	2
48	0	0	0	1	0	1	2
49	0	0	1	1	0	0	2
50	0	0	1	1	0	0	2
51	0	1	0	0	0	1	2
52	0	1	0	0	0	1	2
53	1	0	0	0	1	0	2
54	1	0	0	0	1	0	2
55	0	1	0	0	1	0	2
56	0	1	0	0	1	0	2
57	1	0	0	1	0	0	2
58	1	0	0	1	0	0	2
59	0	0	1	0	0	1	2
60	0	0	1	0	0	1	2
N	20	20	20	20	20	20	120

Appendix 2 Composition of expert panels

a. Expert Panel English

Name	Institution	City	Country
Danilo Rini	Università per Stranieri	Perugia	Italy
Peter Holt	Sabancı Universiy, School of Languages	Istanbul	Turkey
Irena Budreikiene	Adolfas Šapoka Gimnasium	Utena	Lithuania
Dina Tsagari	University of Cyprus, Dept. Of English Studies	Nicosia	Cyprus
Tiina Lammervo	University of Jyväskylä, Centre for Applied Language Studies	Jyväskylä	Finland
Carol Spoettl	Universität Innsbruck	Innsbruck	Austria
José Pascoal	Alameda Universidade, Faculdade Letras	Lisbon	Portugal
Robert Kuzka	Alameda Universidade, Faculdade Letras	Lisbon	Portugal
Veronika Froehlich	Pädagogische Hochschule Heidelberg	Heidelberg	Germany
Robert Marcz	University of Pécs	Pécs	Hungary
Charalambos Kollias	Hellenic American University	Athens	Greece
John Etxeandia	Department for Education, Linguistic Policy and Culture of the Basque Government	Gernica	Spain
Olga Kvasova	Taras Shevchenko National University of Kyiv	Kiev	Ukraine
Tamra Kavytska	Taras Shevchenko National University of Kyiv	Kiev	Ukraine

Name	Institution	City	Country
Josu Romera	Official Language Schools, Department for Education, Linguistic Policy and Culture of the Basque Government	San Sebastian	Spain
Hildegunn Lahlum Helness	University of Bergen, Institutt for Fremmedspråk	Bergen	Norway
Margreet van Aken	Cito	Arnhem	The Netherlands

b. Expert Panel German

Name	Institution	City	Country
René Beunk	Candea College	Duiven	The Netherlands
Simone Bormann-Knoll	Hamburger Volkshochschule	Hamburg	Germany
Kathrin Kunkel-Razum	Duden Verlag	Berlin	Germany
Brigitte Mitteregger	ÖSD-Zentrale (Österreichisches Sprachdiplom Deutsch)	Wien	Austria
Manuela Glabionat	Alpen-Adria Universität ÖSD-Zentrale	Klagenfurt Wien	Austria
Rob Verheyen	Cito	Arnhem	The Netherlands
Eva Wiedenkeller	Universität Freiburg	Freiburg	Switzerland
Lukas Wertenschlag	Universität Freiburg	Freiburg	Switzerland
Beate Zeidler	TELC GmbH (The European Language Certificates)	Frankfurt	Germany
Dessislava Todorova	Ludwig-Maximilians-Universität München	München	Germany
Katerina Touraki	Goethe Institut	Athens	Greece
Gé Stoks	Alta Scuola Pedagogica Canton Ticino	Locarno	Switzerland
Henk Claassen	Stedelijk Gymnasium	Nijmegen	The Netherlands
Stefanie Dengler	Goethe Institut	München	Germany
Gabriele Gippner	Senefelder-Schule Humboldt Universität	Treutlingen, Berlin	Germany

Name	Institution	City	Country
Gisela Merker	Goethe Institut	Amsterdam	The Netherlands
Sonja Kuri	Università degli Studi	Udine	Italy

c. Expert Panel French

Name	Institution	City	Country
Bart Lamote	AKOV, Exam committee Secondary Education	Brussels	Belgium
Florence Sudre	Private university	Tokyo	Japan
Anthippi Potolia	Inalco - Institut national des langues et civilisations orientales, filière FLE; Mairie de Paris (Cours Municipaux d'Adultes)	Paris	France
Suzanna Dejonghe	Flemish Government - Department for Education - Commission for Exams (cfr. Standards and testing agency)	Brussels	Belgium
Laurence Beltran	Université d'Avignon et des Pays de Vaucluse	Avignon	France
Jürgen Mertens	Université des Sciences de l'Éducation Ludwigsburg	Ludwigsburg	Germany
Geneviève Baraona	Institut National des Langues et Civilisations orientales	Paris	France
Faezeh Amar	IRFFLE de l'Université de Nantes	Nantes	France
Dominique Thomaes-Jauréguiberry	Montessoricollege Eindhoven/Fontys Teachers' Training College French Tilburg	Eindhoven/Tilburg	The Netherlands
Christine O'Leary	Sheffield Hallam University	Sheffield	UK
Liliane Koecher	Institut International d'Études Françaises, Université de Strasbourg	Strasbourg	France
Marguerite Bickel	Editions Didier	Malakoff	France

Name	Institution	City	Country
Christelle Hoppe	IRFFLE de l'Université de Nantes	Nantes	France
Catherine Kancellary Delage	Université de Bordeaux	Bordeaux	France
Maria Brems	AKOV, Quality assurance agency in Education and Training	Brussels	Belgium
Marina Segeat Mistretta	Schola Mediterranea	Argelès sur mer	France
Marita Härmälä	The Finnish National Board of Education	Helsinki	Finland
Michel Wauthion	Institut français des Pays-Bas, ambassade de France aux Pays-Bas	The Hague	The Netherlands
Sylvie Senges	Université de Bordeaux	Bordeaux	France
Theresa Weiler	BIFIE (Federal Institute of Education Research)	Vienna	Austria
Trees Aler	Ass. Professeurs de langues vivantes, section de français	Amsterdam	The Netherlands

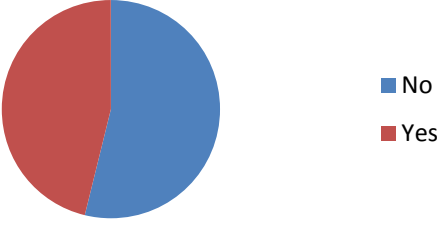
Appendix 3 Design English vwo for international standard-setting

Student	Score points	Assessor													Nbe
		1	2	3	4	5	6	7	8	9	10	11	12	13	
1	46	0	0	0	0	0	0	0	1	1	1	1	0	0	4
2	30.5	0	0	0	0	0	0	0	1	1	1	1	0	0	4
3	40	0	0	0	0	0	0	0	1	1	1	1	0	0	4
4	32.5	0	0	0	0	0	0	0	1	1	1	1	0	0	4
5	42.5	0	0	1	1	0	0	1	1	0	0	0	0	0	4
6	51.5	0	0	1	1	0	0	1	1	0	0	0	0	0	4
7	57	0	0	1	1	0	0	1	1	0	0	0	0	0	4
8	55	0	0	1	1	0	0	1	1	0	0	0	0	0	4
9	44.5	0	1	0	1	0	0	0	0	0	0	1	1	0	4
10	41	0	1	0	1	0	0	0	0	0	0	1	1	0	4
11	43	0	1	0	1	0	0	0	0	0	0	1	1	0	4
12	52	0	1	0	1	0	0	0	0	0	0	1	1	0	4
13	61.5	1	0	0	0	0	0	1	0	0	1	0	1	0	4
14	38	1	0	0	0	0	0	1	0	0	1	0	1	0	4
15	37	1	0	0	0	0	0	1	0	0	1	0	1	0	4
16	42	1	0	0	0	0	0	1	0	0	1	0	1	0	4
17	28.5	0	1	1	0	1	0	0	0	0	1	0	0	0	4
18	31	0	1	1	0	1	0	0	0	0	1	0	0	0	4
19	35	0	1	1	0	1	0	0	0	0	1	0	0	0	4
20	49	0	1	1	0	1	0	0	0	0	1	0	0	0	4
21	32.5	0	0	0	0	1	0	0	1	0	0	0	1	1	4
22	42.5	0	0	0	0	1	0	0	1	0	0	0	1	1	4
23	49	0	0	0	0	1	0	0	1	0	0	0	1	1	4
24	43	0	0	0	0	1	0	0	1	0	0	0	1	1	4
25	45	0	0	0	1	0	1	0	0	0	1	0	0	1	4
26	36	0	0	0	1	0	1	0	0	0	1	0	0	1	4
27	35	0	0	0	1	0	1	0	0	0	1	0	0	1	4
28	39.5	0	0	0	1	0	1	0	0	0	1	0	0	1	4
29	47	0	0	1	0	0	1	0	0	1	0	0	1	0	4
30	42	0	0	1	0	0	1	0	0	1	0	0	1	0	4
31	33	0	0	1	0	0	1	0	0	1	0	0	1	0	4
32	50	0	0	1	0	0	1	0	0	1	0	0	1	0	4
33	51	0	1	0	0	0	0	1	0	1	0	0	0	1	4
34	37	0	1	0	0	0	0	1	0	1	0	0	0	1	4
35	37.5	0	1	0	0	0	0	1	0	1	0	0	0	1	4
36	30.5	0	1	0	0	0	0	1	0	1	0	0	0	1	4
37	29.5	1	1	0	0	0	1	0	1	0	0	0	0	0	4

38	30	1	1	0	0	0	1	0	1	0	0	0	0	0	4
39	37	1	1	0	0	0	1	0	1	0	0	0	0	0	4
40	48	1	1	0	0	0	1	0	1	0	0	0	0	0	4
41	58	0	0	0	0	1	1	1	0	0	0	1	0	0	4
42	34	0	0	0	0	1	1	1	0	0	0	1	0	0	4
43	38	0	0	0	0	1	1	1	0	0	0	1	0	0	4
44	56	0	0	0	0	1	1	1	0	0	0	1	0	0	4
45	38	1	0	0	1	1	0	0	0	1	0	0	0	0	4
46	48	1	0	0	1	1	0	0	0	1	0	0	0	0	4
47	51	1	0	0	1	1	0	0	0	1	0	0	0	0	4
48	42	1	0	0	1	1	0	0	0	1	0	0	0	0	4
49	54	1	0	1	0	0	0	0	0	0	0	1	0	1	4
50	53	1	0	1	0	0	0	0	0	0	0	1	0	1	4
51	66	1	0	1	0	0	0	0	0	0	0	1	0	1	4
52	29.5	1	0	1	0	0	0	0	0	0	0	1	0	1	4
N		16	16	16	16	16	16	16	16	16	16	16	16	16	208

Appendix 4 Examples summaries discussion 2nd round standard-setting

Example English vwo (summary discussion + corresponding student work)

Sufficient for B2?	
 <p>■ No ■ Yes</p>	
Written production/interaction general, range, and use of vocabulary	
YES	NO
<p><u>Interaction general:</u> Generally well-articulated, though not as convincing as no. 243. Clear and detailed. Expresses news and views effectively. Intelligible texts, though the message is not totally clear sometimes (1x).</p> <p><u>Vocabulary:</u> Wide range of vocabulary, albeit, not always accurate. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution. Sufficient range to give clear descriptions, express viewpoints, and develop arguments without too much conspicuous searching for words. Some complex forms are used.</p>	<p><u>Interaction general:</u> Doesn't always respond to task appropriately. Quite a few points where he/she fails to get message across due to incorrect use of syntax and grammar. Lack of accuracy in the language use makes the meaning unclear. Too many basic mistakes (and simple language), does not reach B2 level. Very productive B1: flows well and can make ideas clear, can write straightforward connected texts. Next to incomprehensible (1x).</p> <p><u>Vocabulary:</u> Quite basic with a high tendency of repeating him/herself. Inaccuracies, register problems. Not enough command of B2 vocabulary. Does try to use a range of vocabulary but bizarre spelling means the reader has to decode constantly or reread to substitute the correct collocations or more appropriate wording (<i>ratings/ concepts/ programmes</i>). Generally correct but with some comma mistakes (1x).</p>
Grammatical correctness, orthographic correctness (spelling), coherence and cohesion	
YES	NO
<p><u>Grammatical correctness:</u> Diverse grammatical weaknesses but still clear and very readable. Generally correct; some problems (<i>paids, those network, kill her=nature</i>), a couple of mistakes with tenses, but not systematic. Good grammatical control (2x).</p> <p><u>Orthographic correctness:</u></p>	<p><u>Grammatical correctness:</u> Relatively good grasp of basic grammatical structures, but with several basic mistakes. Some variation in sentence types, but also some incomplete. Severe grammatical errors which affect the reader's understanding of the text (e.g. <i>then/ than x2, less forest, there their</i>, tense sequences), which results to the texts failing in cohesion. Misuse of almost all structures (1x).</p>

<p>Quite poor spelling (<i>wich, there for, oxiden, dissasters, don not...</i>), but good enough at this level.</p> <p>Follows standard layout and paragraph conventions.</p> <p>Spelling and punctuation are reasonable accurate but with mother tongue influence.</p> <p><u>Coherence and cohesion:</u></p> <p>Good coherence, uses a variety of linking words efficiently. Can use a limited number of cohesive devices to link own utterances into clear, coherent discourse though at times with some 'jumpiness'.</p>	<p><u>Orthographic correctness:</u></p> <p>Spelling poor, too many errors. Errors impede understanding and put a strain on the reader.</p> <p>Spelling: generally correct (1x).</p> <p><u>Coherence and cohesion:</u></p> <p>Accuracy and coherence fall apart as it gets going (particularly T3). Dutch word order, poor coherence affects the meaning of the text.</p> <p>Not enough command of linking devices, little use of more advanced linking words. Relatively clear thread.</p>
---	--

Task 1

1. *I agree with your opinion about the programs Wipeout and America's got talent. Networks will get high rating for dumb programs and there for they are making more and more dumb programs. But those programs are also really enjoyable, so is it really bad that those networks make that kind of programs?*
2. *Your article about American TV in summer in the September issue of Rolling Stone caught my eye. It was like reading my own mind. I totally agree with you that networks are nowadays trying the most ridiculous concepts.*
3. *It is like every next show has to be even more ridiculous then the last one. As if America's got talent no longer is about talent but about wierd people doing wierd things. Worse auditions are far more populair then the good ones. Normal programs will all be changed to ridiculous programs.*
4. *Not only is this happening in the United States of America, here in the Netherlands is the same thing happening. The networks are making new concepts but to get high rating they have to make it ridiculous. The good programs are dissepairing and will be replaced by those new concepts.*
5. *Networks will no longer make concept for good programs. The only thing that the want to accomplish now is to get high ratings even if that means that they have to make a rubbish program.*
6. *The only networks that do not have that goal are the networks for the government. They do not want high ratings so bad that they are gonna make stupid programs. The government paids those network and there for they have to make programs that the government have approved.*
7. *In the summer I do not watch TV very often. The weather is to beautiful to stay inside and watch TV, but when I do watch TV I like to look at the more intelletent programs. The programs about science for example. It is fun to watch and they are good for your education.*

Task 2

<i>1. I do not understand the Sigmund that Ricky Toebe posted at 12:56 PM. I tried to understand it but I still do not get it.</i>
<i>2. I think the meaning of Sigmund is, that time is precious and you have to live while you can because time is running out.</i>
<i>3. Just like the Sigmund I attached to this comment. It is also about time and the importance of it.</i>
<i>4. Sigmund is a man of time. There is always an answer in time. So the solution to the woman her problem is also time.</i>
<i>5. Only the woman wants to get rid of her sadness immediately. So when Sigmund said it will take time to do that, she got even more sad, because time is running out and he has to spend some of that time to forget her old love.</i>
<i>6. When I see those strips side by side, it is my opinion that I like the Dutch one more. It is easier for people to recognize that situation. Everybody has been sad sometime in his life, so everybody understands what Sigmund says.</i>
<i>7. I would like some reaction of you readers. What do you guys think about the Dutch Sigmund? Do you share my opinion or do you have an other opinion? I hope I could read some of your comments soon.</i>

Task 3

Eight thousand years ago almost half of the Earth's land surface was covered with forest. Nowadays four fifths of that forest have already been irreplaceably degraded or destroyed. We are damaging mother nature and if we do not stop before it is to late we are going to kill her. But why are we damaging her, why do we destroy the forest. I am going to tell you why.

First of all, for ages we destroyed forest so we could use the ground. The ground of the forest is full of supplements wich are good for growing vegatables on them. The growing population demands us to cut even more forest down.

That leads to my next point, with a growing population comes a bigger question to infrastructure. To keep all those cities connected there need more infrastructure. The roads between cities go straight trough the woods.

With more infrastructure there are more people who are going to use it. Traveling gets easier and there for people are going to travel more. The industrie is growing fast and with that we need more fuel. The forests are cut and the trees are used as fuel. But we need more and more fuel every second, so more forest have to be cut down to foffill that demand.

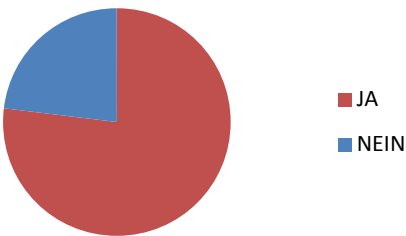
That we need to cut the forest down is a thing that is clear, but why is it then a problem? By damaging the forest we are taking the home of a lot of animals away. Some animals are nearly extingwished by our demand of damaging forest. What gives us the right to take there home

away and kill them just that we can get our work faster for example.

But we are not only damaging the forest or those animals. The whole earth is getting damaged by it. The climate is changing because there is far less forest. The forest uses carbondioxide and to create oxiden. That oxiden that we need. The carbondioxide is damaging the earth and is going to make a hole in the atmossphere. The earth will heat up and it will even be unlivable for humans. But the forest does not only do that. The forest also keeps a lot of water at his place. Without the forest we will get far more nature dissasteres. The water is going to flow and become a mudstream. When a mudstream is flowing nothing can stop it. It will damage everything on its way. That was just one nature dissaster.

I think we should do something about it. It is now time to do it before it is to late. Then we did not only kill mother nature, planet earth but also we would kill ourself. There are a lot of other solutions so that we don not have to damage the forest and I think the reasons I mentioned before are enough good reasons to do it, so just do it.

Example German *havo* (summary discussion + corresponding student work)

Ausreichend für B1?	
	
Schriftliche Produktion/Interaktion allgemein Wortschatzspektrum und -Beherrschung	
JA	NEIN
<p>ALLGEMEIN Texte entsprechen dem Schreibenanlass und sind umfangreich. Sehr flott, kreativ und interaktiv geschrieben. Aber von Kontext keine Ahnung (<i>Zwischen Alpental, Meer!</i>)! Kann Überlegungen ziemlich gut rüberbringen, v.a. in Aufg. 1 und 2. Im Chat wird der richtige Ton getroffen und gut auf den Gesprächspartner reagiert. Der Bewerbungsaufruf ist aktivierend. Der Weblog ist dagegen teilweise unverständlich wegen der Wortwahl. Textsortenspezifische Ausstieg nicht immer optimal (Aufg. 2: <i>Auf wieder sehen!</i>).</p> <p>WORTSCHATZ Ausreichendes Spektrum, es wird ausserdem gut argumentiert, vor allem in Aufg. 3; dabei jedoch einige Male unpassend bzw. missverständliche Unschärfen wegen Interferenzfehler aus der Muttersprache bzw. aus dem Englischen (<i>das Bildung ist hier sehr grosse, auf Mountainbikes gereden, allein Gemüse essen, das Wasser gemist</i>). Allerdings wird viel Kontext geboten, so daß doch klar wird, was gemeint ist. Lesefluss wird nicht besonders gestört oder behindert. Verwendet auch einige komplexere Ausdrücke (<i>Schwierigkeiten bereiten</i>).</p>	<p>ALLGEMEIN Kann auf Mitteilungen angemessen und verständlich reagieren (Aufg. 1). Die Verständlichkeit der Texte nimmt aber leider von Aufgabe zu Aufgabe ab.</p> <p>WORTSCHATZ Der Grundwortschatz ist weder groß noch wird er ausreichend gut beherrscht. Nicht ausreichend für B1 (<i>nach die Preise kückst, Uhren arbeiten, das Bildung, smäckt, stimmit da nicht mit, mogen essen</i>).</p>
Grammatische Korrektheit Orthographie, Kohärenz und Kohäsion	
JA	NEIN
<p>GRAMMATISCHE KORREKTHEIT Einfache, häufig verwendete Strukturen sind gut. In Aufg. 2 anspruchsvollerer Satzbau (größtenteils gelungen). Unsicherheiten im Präpositionalbereich. Verwechslungen wenn/als, Modalverb (mogen), Genus, Verbflexion (<i>ich weiße, geschwammen</i>), falsche Reflexivpronomen. Die Satzstellung ist aber meistens korrekt. Trotz erheblicher Patzer (<i>das Camping braucht du</i>) wird das Verständnis nicht wesentlich durch grammatische Schwächen gehindert.</p>	<p>GRAMMATISCHE KORREKTHEIT Nicht ausreichend für B1 (<i>ein andere Traum, mit das Geld, das,dass, ich lern, das Camping braucht du, denn brauchen wir dich, nach die Sport, ich weisse</i>). Texte durchgängig verständlich, die zahlreichen Fehler bei Aufg. 3 führen aber zu Missverständnissen.</p>

<p>ORTHOGRAPHIE Häufig Probleme bei Orthographie und Klein- und Großschreibung. Texte sind aber durchgehend verständlich.</p> <p>KOHÄRENZ UND KOHÄSION Niveauadäquat. Überwiegend gelungener Textaufbau mit zusammenhängenden Ausführungen. Kohärenz/Kohäsion jedoch schwankend. Wenig Verwendung von Konnektoren, eher Aneinanderreihung von einfachen Sätzen.</p>	<p>ORTHOGRAPHIE Noch ausreichend (<i>kueckst / ich schlaffe</i>).</p> <p>KOHÄRENZ UND KOHÄSION Kurze, einfache Elemente werden zwar zu zusammenhängenden Äußerungen verbunden, Konnektoren werden aber kaum benutzt. Kohärenz ist noch akzeptabel für B1.</p>
---	---

Auftrag 1

...

Wow, sehr toll! Ich freue mich darüber! Aber hattest du nicht ein andere Traum? Ich glaube etwas über ein neue Fahrrad ob neue Klamotten?

...

Er ist echt super, glaube ich auch. Leider ist mein Chef nicht so nett. Ich habe sehr nette Kollegen aber den Job selber bereit mir Schwierigkeiten, weil ich nur drei Euro pro Stunde verdiene!!!

...

Deine Mutter findet die schule sehr wichtig, dass stimmt auch, aber den Job selber ist auch sehr wichtig! Mann lernt was es mit das Geld kann tun! Und auch kennen sie andere Leute/Kollegen lernen.

...

ich bekomme Taschengeld, ungefähr €15 pro Woche! €15 ist nicht so viel, als du heute nach die Preise von Klamotten kückst. Aber der Vorteil ist das ich jede Woche Geld habe!

...

Ich lern zusammen fassungen und lern meine Hausaufgaben

Auftrag 2

Achtung! Das Camping braucht du. Bist du minimal 16 Jahre Alt und suchst du ein Job?

Denn brauche wir dich!!!

Du bekommst €8 pro Stunde. Und du musst 12 Stunde pro Woche arbeiten. Die Stunden sind verschieden jede Tag. Wir brauchen Hilfe in de Shop und natürlich ist es möglich mehr Uren zu arbeiten, wenn du das möchtest. Am liebsten süchen wir Jemand die Direct starten kann.

Für mehr Information und wenn du sich selbst in diese Job interessiert, könntest du Frey anrufen.

Auf wieder sehen!

Auftrag 3

08.03.2013 Heute sind wir in die Zwischen Alpental. Das Bildung hier ist sehr große. Es liegt am See. Und ich sehe verschiedene Tieren.

09.03.2013

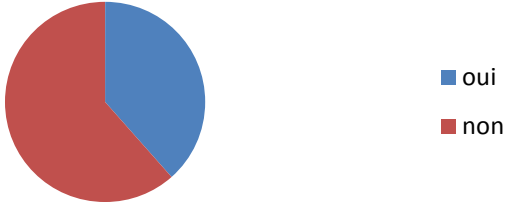
Das Programm war sehr interessant. Wir haben auf Mountainbikes gereden. Ich war sehr Müde. Nach die Sport haben wir gegessen. Das Essen smäckt sehr gut. Ich habe Pizza gegessen. Und Pizza mag ich sehr. Der Lehrer stimmt da

nicht mit. Er sagt das Pizza nicht Gesund ist. Das ist Normal, er ist ein Sportler! Er hat nur Gemüse gegessen. Aber ich mag das nicht. So, er kwam mit ein Plan. Wir mogen 2 mal pro Woche Fastfood essen und wir müssen die andere Tagen allein Gemüse essen. Das Vorteil ist das Gemüse sehr Gesund ist, aber ich mag das nicht so gern!!! Darum weiße ich noch nicht was ich machen soll.

11.03.2013

Morgen gehen wir nach das Meer. Ich freue mir darüber, weil ich Schwimmen wil. Ich habe for lange Zeit nicht geschwammen und ich habe das Wasser gemist! Jetzt gehe ich schlaffe und schreibe ich Morgens weiter!

Example French *havo* (summary discussion + corresponding student work)

Suffisant pour A2?	
 <p style="text-align: right;">■ oui ■ non</p>	
Production écrite/interaction générale, étendue et maîtrise du vocabulaire	
OUI	NON
<p>Interaction générale Il arrive à satisfaire ses besoins primordiaux, mais il a encore des problèmes pour exprimer des pensées plus complexes. Quelques problèmes dans exercice 4 (voc) mais sens global clair. Interaction générale fonctionne. Un petit A2, mais les éléments essentiels (sauf lors du dernier exercice) sont présents.</p> <p>Vocabulaire Bon vocabulaire, simple.</p>	<p>Interaction générale L'interaction est souvent correcte. Mais ne répond pas à certaines questions (Ex4 Q4). A du mal à donner des informations simples clairement malgré une relative aisance à poser des questions simples. Incohérences qui gênent le sens global du message.</p> <p>Vocabulaire Vocabulaire simple ; parfois incompréhension des mots. Ne possède pas un vocabulaire suffisant pour les besoins communicatifs élémentaires, certains mots ne semblent pas connus : « vacances », « tenniser », « la carte », « jouer a », « il est très vite ».</p>
Correction grammaticale, maîtrise de l'orthographe, cohérence et cohésion	
OUI	NON
<p>Correction grammaticale Bonne correction grammaticale. Oubli de l'accord, accepté à ce niveau.</p> <p>Orthographe Bonne correction orthographique. Mots de liaisons et orthographe phonétique permet l'accès au sens, phrases simples maîtrisées.</p> <p>Cohérence et cohésion Cohésion en concordance avec un niveau A2 Simple connecteurs (« et »). Le dernier exercice prouve qu'il s'agit d'un petit A2. Des problèmes de syntaxe et d'enchaînement.</p>	<p>Correction grammaticale Grammaire trop simple, ne maîtrise pas les structures de base. Aucun temps composé. Mauvaise maîtrise de « avoir » au Présent : « je n'a pas ». Verbes non conjugués : « je jouer ». Age exprimé avec « être ».</p> <p>La grammaire présente des lacunes importantes comme confusion de genre : « un question », « un grande maison », « quelle jours », « le culture », « mon langue français », et de nombre : « quelle jours », « les reaction ».</p> <p>Mauvais usage des prépositions de base: « en le Côte », « en mardi », « en midi », « pour 12 ans » (depuis). Ne sait pas poser les questions fondamentales élémentaires : « quel » / est-ce que, quand.</p> <p>Orthographe Orthographe dans l'ensemble correcte excepté « grandmère/grandpère ». Ponctuation correcte.</p>

Exercice 1:

*Je m'appelle Je téléphone pour ma grandmère.
J'ai un question...
Ma grandmère est seule. Elle est 85 ans, mon grandpère est mort. Elle a
un grande maison.
Nous pouvons un rendez-vous en la maison ma grandmère. C'est en
Saint Germain.*

Exercice 2:

Je fais du tennis pour douze ans. J'ai trois cours chaque semaine.

*Je veux jouer a Djokovic. Il est beau et il est un bon tenniser.
il est très vite !*

Je veux la carte pourquoi, je jouer longtemps tennis, et c'est ma droom.

Exercice 3:

*C'est une challenge pour travailler chez vous ! Pour la langue et
le culture.*

*Je voyage avec le train en mardi en midi et j'arrive en
16.00 heure.*

*Je travaille quelle joures ? et mon emploi du temps est ? et je dois
travailler en le nuit ?*

*C'est un idee super ! Mon langue français est bon. Est c'est une bon
chance pour étudier le langue.*

C'est possible mes parents rester en l'hotel ?

Exercice 4:

*Je veux voyager en le Côte d'Azur pour ma vacation.
Je na pas understand les reaction. C'est trop erg.
C'est damage pour le stakingen.
C'est une idee pour regler un train pour le voyageurs ?
Merci,
...*

Appendix 5 Interview guidelines

- The interview guidelines are semi-structured. This means that the guidelines are flexible, so that new questions may be asked based on the answers given by the teachers.
- It is not necessary to follow the order given here. However, make sure to touch on every theme.
- In the first part, the curricular spider web is used as a reference framework.
- In the second part, validity, usefulness and practicability of the test and the assessment method are entered into.

a) Factors affecting the results.

1. Please indicate three factors, which, in your school, affect the results for your subject most.

In what category would you range these?

1. Rationale
2. Attainment targets
3. Learning content
4. Learning activities, work forms
5. Teacher role
6. Materials and resources
7. Learning environment
8. Time
9. Assessment
10. Grouping arrangements

2. What do you feel are the bottlenecks regarding the teaching of your subject?

In what category would you range these?

- | | |
|-----------------------------------|--|
| 1 Rationale | how does the department / school view the learning of a language? |
| 2 Attainment targets | targets for writing skills in a foreign language? |
| 3 Learning content | skills vs. knowledge, productive/receptive, literature... |
| 4 Learning activities, work forms | which are efficient? which are practicable? |
| 5 Teacher role | what role suits you best? and where the attaining of the learning target is concerned? |
| 6 Materials and resources | course, training material, extra materials (including digital ones) |
| 7 Learning environment | classrooms, ICT facilities, school environment |

8	Time	lesson duration, preparation time, the lesson's time of day, time for doing homework...
9	Assessment	what do you test? how do you test?
10	Grouping arrangements	do students work individually, in couples, or in small groups?

3. A young teacher English/German/French is hired in your school. What would be your most important tip for him/her?

b) Testing.

1. What should be tested regarding writing skills?

- what should a student be able to do?
- role of grammar / vocabulary?

2. How do you test writing skills?

- content
- form
- frequency
- assessment method

3. Why do you choose for this method?

- Are you happy with it?
- Are there any bottlenecks or points of improvement? If so, what are they?

4. This year, will you be using a similar writing test to the one developed by Cito for this study project?

- why?

5. In this project, two assessment methods have been discussed: the Cito assessment model has been made available to all schools. Two assessment sessions have been organised using the so-called 'anchor method'. What is your opinion of these two methods?

- do they give a good idea of the performance level of the students? (validity)
- do they lend themselves for objective use? Or, rather, do you believe that two different teachers may arrive at the same scores? (reliability)
- are they practicable to be used in school? (practicability)

6. What is your golden tip for a good writing-skills test?

SLO is the Netherlands institute for curriculum development. We are bridging the contexts of policy, research, and practice. Our expertise focuses on the development of curricular goals and content for various educational levels, from national policy to classroom practices. We closely collaborate with many different stakeholders from policy circles, schools (boards, principals, teachers), research, civic organizations, and the society at large. This allows us to design and validate relevant curriculum frameworks, to elaborate exemplary materials and to evaluate these in school practices. Our products and services support both policy makers and schools and teachers in making substantive curricular decisions and in elaborating these into relevant, inspiring and effective education.

PO Box 2041
7500 CA Enschede
The Netherlands

T + 31 (53) 484 08 40
E info@slo.nl
www.slo.nl

 [company/slo](https://www.linkedin.com/company/slo)

 [@slocommunicatie](https://twitter.com/slocommunicatie)

slo