BRITISH COUNCIL

ENGLISH LANGUAGE
ASSESSMENT RESEARCH GROUP

ASSESSING WRITING AT LOWER LEVELS: RESEARCH FINDINGS, TASK DEVELOPMENT LOCALLY AND INTERNATIONALLY, AND THE OPPORTUNITIES PRESENTED BY THE EXTENDED CEFR DESCRIPTORS

AR-G/2018/4

Eva Konrad, Franz Holzknecht, Veronika Schwarz and Carol Spöttl

ARAGs RESEARCH REPORTS ONLINE
SERIES EDITOR: VIVIEN BERRY
ISSN 2057-5203

ABSTRACT

This study investigated the current theory and practice of assessing writing at lower levels and the potential effects of the recently published extended CEFR descriptors. A mixed-methods design was applied that consisted of three main parts. First, a review of literature was conducted. Then, 45 sample tasks at A1, A2 or comparable low levels from 21 writing exams by 10 international test providers were collected and analysed. An online survey was conducted to collect information about the current practices of small-scale test development, practitioners' perceptions of and experiences with assessing low-level writing and the role of the CEFR.

The survey of the research literature conducted for this study, both seminal and recent, revealed that little guidance is currently offered for the development of low-level writing tasks, despite the practical need for assessing writing at this level. Results from the task analysis and the online survey showed that test developers face a number of challenges when assessing low-level writing. In particular, it was found that a clear differentiation between A1 and A2 tasks is often missing, that there are large differences in what is required of test-takers within the same level, and that existing task types and scales may be considered inadequate for assessing A-level writing.

It is suggested that the extended CEFR descriptors may provide more guidance for differentiating between the levels and for developing additional, authentic task types for writing at low levels.

Authors

Eva Konrad holds an MA in Language Testing from the University of Lancaster and an MA in English and American Studies from the University of Innsbruck. She works as a research assistant at the University of Innsbruck where she is a member of the Language Testing Research Group. She was previously involved in the development of language tests for the standardised Austrian school-leaving exam project, and has lectured at Austrian teacher training colleges. For her Master thesis, she conducted research into the potential effects of using a dictionary in a writing exam. She is currently enrolled in Lancaster's PhD program in Language Testing, where she plans to investigate diagnostic testing of L2 listening. Her research interests are the assessment of L2 writing and listening, diagnostic testing and reading and listening task development.

Franz Holzknecht holds an MA in Language Testing from Lancaster University and an MA in English Language Teaching and Sports Education from the University of Innsbruck. He works as a researcher for the Language Testing Research Group at the University of Innsbruck and is also pursuing his PhD on listening assessment at Lancaster University. His research interests are mainly in language assessment, particularly in listening assessment, writing assessment, and the use of eye-tracking. Franz was professionally involved in language test development for the standardised Austrian school-leaving examination. He has taught language testing courses at teacher training colleges in Austria and England and has been a consultant to a number of international language testing companies on test development issues. Franz has presented his research at international conferences and has published his work in *Assessing Writing*, *Papers in Language Testing and Assessment* and in several book chapters.

Veronika Schwarz holds a qualification from the University of Innsbruck as an academic expert of German as a foreign language, and is a qualified examiner for DTÖ at the A2/B1 level. She holds a MA in Romance Studies and a MA of Theatre, Film and Media Studies from the University of Vienna. She works as a research assistant for the Language Testing Research Group at the University of Innsbruck and as a dramatic adviser on theatre and performance projects. In her Master thesis, she wrote about the impact of performativity of speech acts in Spanish theatre on identity construction. Her research interests now are mainly in language assessment, particularly in the impact of task performance in educational and testing situations for German and Spanish as a foreign language.

Carol Spöttl is the co-ordinator of the Language Testing Research Group (LTRGI) at the University of Innsbruck's School of Education. She has degrees from the Universities of Edinburgh, East Anglia and Lancaster and has worked at the University of Innsbruck in the field of language teaching, testing and evaluation. From 2007–2015, she was the exam reform project leader in Austria. The government-funded project introduced a new school-leaving exam in the foreign languages. It saw the development of CEFR-linked tests at two CEFR levels, for the skills of reading, listening, writing and language in use and for four of the languages taught: English, French, Italian and Spanish. Current research projects in which the LTRGI is involved range from identifying anchor items for a large-scale language test, addressing assessment issues surrounding literature in ELT contexts, providing teacher trainers with benchmarked performances for speaking and writing, comparing writing assessment in two European countries to cross-disciplinary projects with the university's medical faculty.

CONTENTS

List of figures

List of tables

1. BACKGROUND

Most research into the assessment of English as a second language (ESL) writing so far has been conducted in the context of higher-level writing, such as academic writing, and has typically focused on multi-level tasks which are developed with the intention of eliciting a large variety of performances (Weigle, 2002). There is, however, very little literature on task development in the context of "a level-specific approach, where one task targets one specific proficiency level" (Harsch & Rupp, 2011, p. 1). Nevertheless, such an approach is necessary when trying to establish whether a test candidate has mastered a particular level of writing, as is the case in the Aptis Writing Test. In addition, as ESL learners are limited in what they can do with writing at the lower levels, test developers are naturally limited in what they can assess.

The Common European Framework (CEFR) has had a lasting impact on standardised language assessment, both in Europe and beyond (Council of Europe, 2006; Jones & Saville, 2009; Little, 2007; Martyniuk & Noijons, 2007). An increasing number of testing and educational organisations now claim linkage of their exams to the CEFR, including the lower levels A1 and A2. According to the 2001 version of the CEFR scales on overall written production, writers at A1 "can write simple isolated phrases and sentences", and writers at A2 "can write a series of simple phrases and sentences linked with simple connectors like 'and' and 'but'" (Council of Europe, 2001). A1 writers are further expected to be able to write short, simple postcards and fill in forms with basic personal information, and at A2, writers "can write short, simple notes and messages relating to matters in areas of immediate need" as well as "a very simple personal letter, for example thanking someone for something" (p. 26). Thus, although the CEFR was never intended to be a manual for test task design, the first version of the CEFR presents test developers with a very narrow frame for designing A-level writing tasks, particularly in terms of task type. Generally, test developers will aim to use a spread of different task types to balance out possible task type effects. Due to the restrictions of the CEFR's A-level descriptors, this does not seem possible. At the same time, Harsch (2007) states that the CEFR's can-do statements and scales do not provide a suitable basis for the development of writing tasks. This would suggest that the needs of test and task design might justify a more flexible interpretation of the CEFR scales.

An additional problem might arise in the context of A-level writing task design. In one of the few studies on level specific writing tasks, Harsch and Rupp (2011) investigated whether cut-scores could be empirically set to differentiate between the CEFR levels. Whilst they were able to establish cut-scores for higher levels, results for lower-level tasks were more ambiguous, suggesting that it might be difficult to draw a clear line between A1 and A2 tasks.

In practice, a number of tests assess (or claim to assess) A-level writing (e.g. Pearson PTE Elementary Level, Cambridge ESOL A1 and A2, EIKEN Grade 3 and Pre-2, to name but a few). The Aptis Writing Test also includes tasks at both A1 and A2 level. Both test and task design of the Aptis Test are linked to the CEFR at the development stage, and a standard-setting study by O'Sullivan (2015) has validated this link. The CEFR also forms the basis for the rating scales used in the writing test, although no rating scale is provided for A1 level in the Aptis General Technical Manual. The A1 level tasks target writing at word level (O'Sullivan & Dunlea, 2015, p. 16). The task types currently used to elicit the skills described in the test specifications include completing forms with basic personal information. At A2, tasks require test-takers to produce short written descriptions at sentence level, again by filling out a form.

However, the Council of Europe has been working on extended descriptors for the CEFR, which were published early in 2018, and include new descriptors especially in the area of A1 and pre-A1 writing, and particularly in the scale for Overall Written Interaction with the subscales for Written Interaction, Correspondence, and Notes, Messages and Forms. In addition, new scales have been developed for the area of online interaction, including scales for Online Conversation and Discussion and Goal-Oriented Online Transactions and Collaboration (Council of Europe, 2018). In light of this situation, it would be prudent to consider any potential impact from these extended descriptors on currently used A-level writing tasks for assessment purposes.

A lot remains unknown about the assessment of A-level writing. Recent developments regarding migration into Europe have given increased importance to low-level language testing in general (including the assessment of writing), and this, in combination with the Council of Europe's newly published extended CEFR descriptors, presents a window of opportunity for revising and updating existing practices of assessing low-level writing.

2. RESEARCH QUESTIONS

Based on the discussion above, the following research questions are addressed.

1. What is the current state of research regarding the development of task types assessing low-level writing?
2. What types of writing tasks are currently used in international, national and local contexts to assess low-level writing (CEFR A1 and/or A2 or comparable levels)?
3. What strategies do test developers employ to deal with the limitations of the CEFR in the context of low-level writing assessment (CEFR A1 and/or A2 or comparable levels)?
4. What potential do the expected changes to the CEFR offer to adapt assessment practices of low-level writing?

3. METHODOLOGY

To answer these research questions, a mixed-methods approach was employed. First, a thorough literature review was conducted in order to answer RQ1. Then, a number of publicly available sample tasks targeting low-level writing were collected from international test providers and subsequently analysed to answer RQ2. In addition, an online survey was developed to investigate how various institutions involved in language testing currently approach low-level writing assessment, and how the extended CEFR descriptors would affect their current practices. The outcomes of this survey informed RQs 2, 3, and 4. Finally, the additions to the CEFR writing scales were analysed in detail to answer RQ4. In the following, each of these steps will be outlined in more detail.

3.1 Literature review

For the literature review, an online literature search was conducted in two university libraries on varying combinations of key words such as: assessing low-level/A-level writing, development of writing skills, L2 writing at low levels, writing task development/design. In addition, we searched issues spanning the last 10 years of the following journals: *Assessing Writing*, *Journal of Writing Research*, *Journal of Second Language Writing*, as well as *Language Testing* and *Language Assessment Quarterly*. Course books and manuals that were collected for the task analysis were also checked for relevant information on task development. Finally, the resources section of the EALTA website was searched for relevant conference presentations (other testing organisations do not make these materials publicly available).

3.2 Task analysis

3.2.1 Tasks

The first step for this part of the study was to collect writing tasks targeting CEFR levels A1 and A2 or comparable levels. First, we collected 45 publicly available writing tasks from 21 writing exams by 10 international test providers. To analyse these tasks according to a grid described below, we also collected any additional material provided by the test developers, such as test specifications, rating scales, or CEFR-linkage and validation studies.

Second, we contacted a number of international educational institutions likely to work on lower levels writing assessment. These included the Swedish National Agency for Education, the Spanish Ministry of Education, the Finnish National Certificates of Language Proficiency, the Academie de Toulouse in France, the Dipartimento Istruzione e Formazione Italiana (The Italian Department for General and Vocational Education) in Bolzano, Italy, and the National Institute for Research, Innovation and Development in the Austrian School System (BIFIE). Although these institutions develop and administer writing tasks at the lower levels, the information provided with regards to different aspects related to task development or CEFR linkage was often incomplete or not made available in the public domain for understandable reasons. Therefore, the analysis in relation to these writing tasks was very restricted and the assessments could not be included in the study.

Third, classroom teachers in Austria, Italy, Spain, and Finland were contacted and invited to send us writing tasks used in this context for the assessment of A1 or A2 level writing. However, we decided not to include these in the analysis either due to the low number of responses or the lack of information regarding most aspects of task development in the responses.

3.2.2 Task analysis grid

In order to analyse the low-level writing tasks that were collected, the CEFR Grid for Writing Tasks v.3.1 (Council of Europe, 2009), developed by the Association of Language Testers in Europe, was used. The grid was developed in two formats: analysis and presentation. For the purpose of this study, the presentation grid was used, as the aim was "to provide a descriptive record of test tasks" (Council of Europe, 2008, p. 1). Only information relevant for this study was entered in the grid and this will be reported on in the Results section. This includes general information about the writing component, such as number of tasks, time to complete the tasks, integration of skills, channel (handwritten or word-processed) and targeted CEFR level, as well as more specific information such as types of tasks used, number of words targeted, and task authenticity. The complete grid is included as Appendix 1.

3.3 Survey

3.3.1 Survey design

An online survey was designed using the web-based tool eSurv (www.esurv.org). Expert evaluation was chosen as the most effective way of pre-testing the survey. This decision was based on two considerations. First, the number of possible respondents for the main survey was expected to be limited and a large-scale pilot might have further reduced our sample. Second, expert judgment was considered a practicable and fruitful way to reduce measurement error and identify in advance any items that may cause problems during the administration of the survey (Olson, 2010). All four experts who took part in the pretest have a background in language teaching; in addition, they have expertise in one or several of the following areas: statistics, language testing, and linguistics. For the pretest, they were given a set of questions to consider while taking a preliminary version of the online survey and gave their answers to the researcher directly after completion. The survey was then revised based on the experts' responses and published.

The final version contained questions on the following topics:

- assessment of low-level writing at participants' workplaces
- details about the tasks being used (i.e. task type, required length of output, time given for task completion, target level, design process)
- particular challenges of low-level writing that the participants had encountered
- rating of low-level writing
- the role of CEFR in participants' daily work, their perception of its usefulness and knowledge about the new descriptors
- participants' biodata.

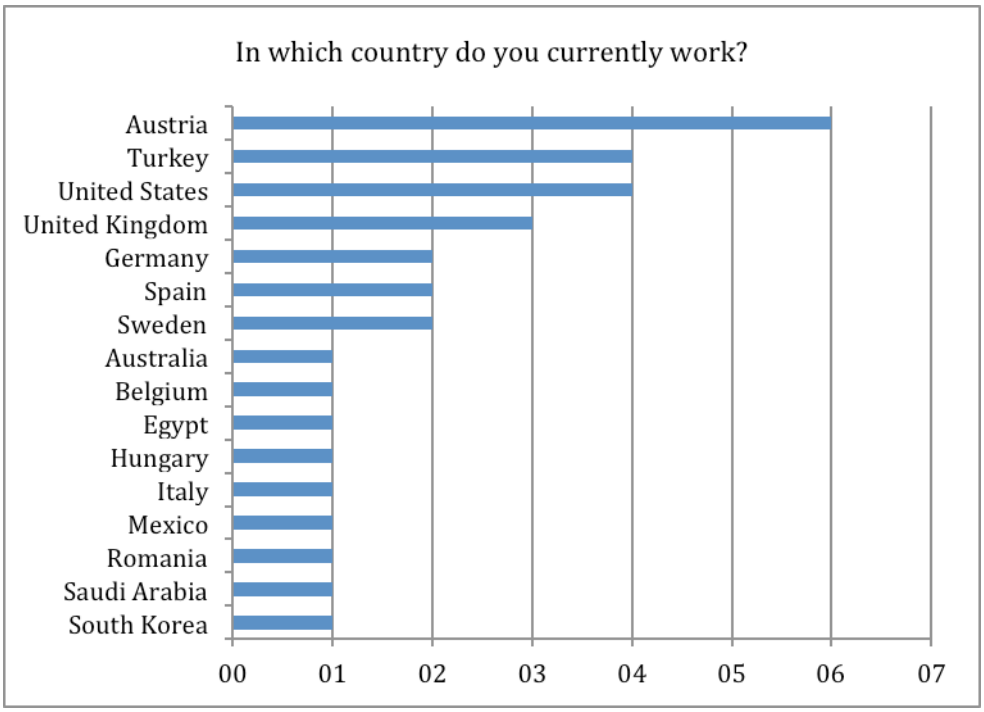
A full reproduction of the survey can be accessed at:
https://www.uibk.ac.at/fakultaeten/soe/ltrgi/repository/low-level-writing_survey.pdf.
The information sheet for the participants can be accessed at:
https://www.uibk.ac.at/fakultaeten/soe/ltrgi/repository/low-level-writing_survey-infosheet.pdf.

3.3.2 Subjects

The online survey was made available to members of the LTEST-L and EALTA mailing lists. In total, 87 people responded to the survey, however, more than half of them stopped after only the first question and could therefore not be included in the analysis. The final sample included in the analysis consisted of 42 responses, of which 32 completed the full survey. The remaining 10 dropped out at different stages during the survey.

The survey respondents were working in 16 different countries across the globe (Figure 1), with the majority being employed in Austria (six participants), Turkey (four participants), the US (four participants), and the UK (three participants).[1]

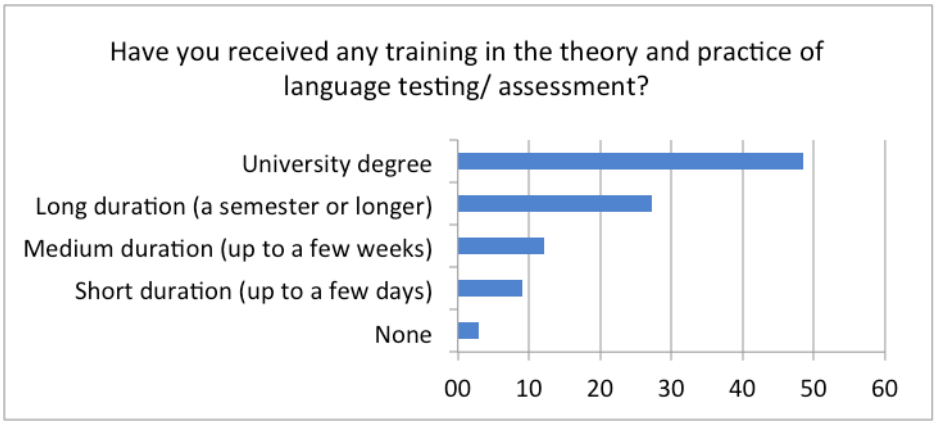
Figure 1: Participants' country of employment (N=32)



The majority of participants were language testing professionals who had obtained a university degree in that subject (N=16) or had been trained for an extended period of time (a semester or longer, N=9). Four participants had received training up to a few weeks, while three had been trained up to a few days. One participant had not received any training in language testing (see Figure 2).

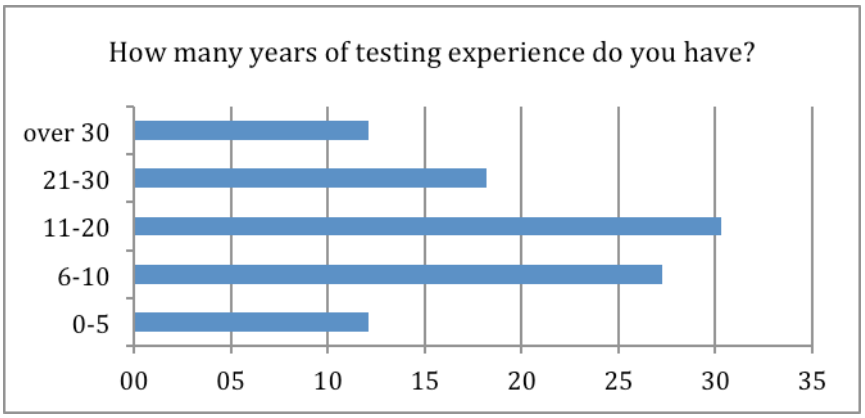
[1] The number of responses to questions about biodata was lowest as these questions were put at the end of the online questionnaire.

Figure 2: Participants' language testing training (N=33, in % of cases)



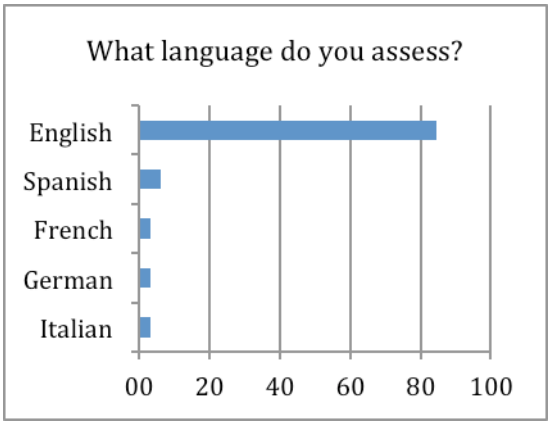
The respondents were mostly experienced language testers (Figure 3), with the majority (N=10) between 11 and 20 years of experience. Four participants had up to five years of testing experience, nine participants between six and nine years, and six participants between 21 and 30 years. The remaining four participants had more than 30 years of experience in language assessment. Nine participants did not respond to this question.

Figure 3: Participants' testing experience (N=33, in % of cases)

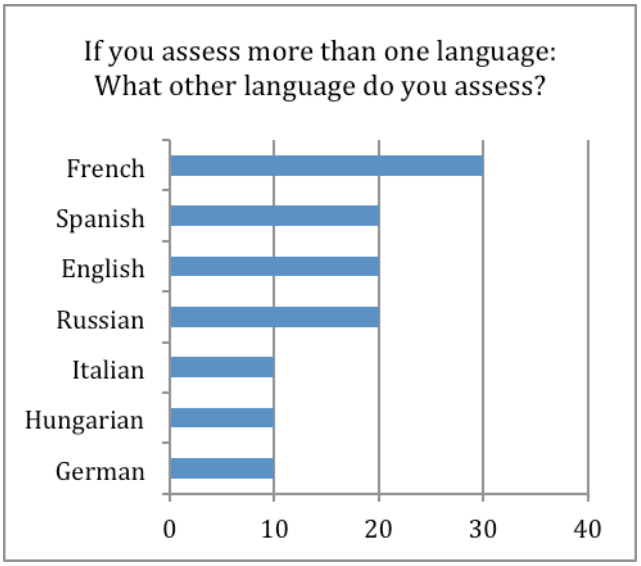


In terms of languages assessed (Figure 4), the vast majority of respondents (N=27) designed language tests for English, while the remaining five participants designed tests for German, Italian, French, or Spanish. Ten participants were assessing more than one language. In total, the sample designed language tests for seven different languages.

Figure 4: Languages assessed by participants (N=33, in % of cases)



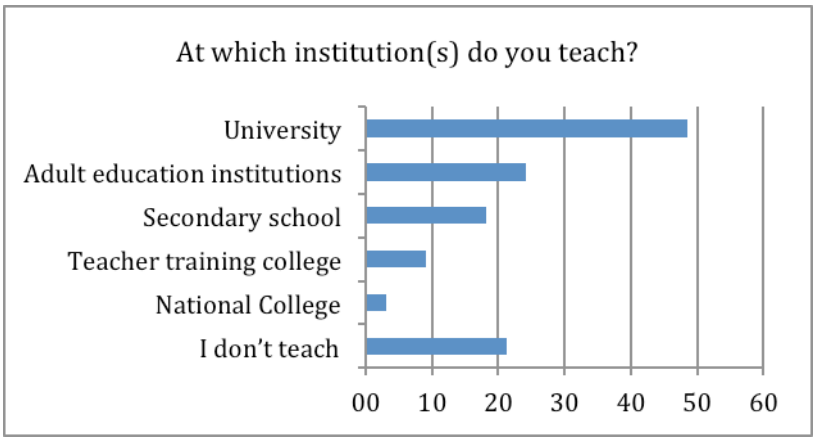
N=32



N=10

The majority of the sample was also involved in teaching (Figure 5), with more than half teaching at university (N=16). Participants were also teaching at adult education institutions (N=8), secondary schools (N=6), teacher training colleges (N=3), and at a national college (N=1). Seven participants did not teach, with nine answers missing.

Figure 5: Teaching by participants (N=33, in % of cases, multiple answers possible)



3.4 Comparison between original and extended writing related A1/A2 CEFR sub-scales

RQ4 asks what options are offered by the recently added CEFR descriptors for the development of A-level writing tasks. As a first step, the original and extended scales referring to writing were compared. Appendix 2 provides a full reproduction of the relevant extended scales, with the added descriptors highlighted in green for a better overview.

Drawing on further insights about current practices of assessing A-level writing gained through the survey and task analysis, we then discussed whether the new descriptors might be helpful in addressing potential difficulties that were identified.

3.5 Ethical consent

Before the beginning of the data collection, a Certificate of Good Standing was gained from the University of Innsbruck's Board for Ethical Issues. Participants of the survey were informed about the major aims of the research project via an online information sheet prior to answering any questions. Anonymity of results was guaranteed.

4. RESULTS

4.1 Literature review

4.1.1 Theoretical framework

Two models of writing have been very influential in writing research: Hayes and Flower (1980), which was later updated by Hayes (1996), and Bereiter and Scardamalia (1987). Hayes and Flower (1980) conceptualise writing as a recursive process that involves three major cognitive processes: planning, where ideas are generated drawing on the writing assignment and the writer's long term memory; translating, during which the planned text is put into words; and reviewing, which involves re-reading and editing of the written text.

The model by Bereiter and Scardamalia (Bereiter & Scardamalia, 1987) differentiates between unskilled and skilled writing. The first is described as knowledge telling, which is similar to unplanned speaking. L1 writing done by children or adolescents would mostly fall into this category (Weigle, 2002). Skilled writing, on the other hand, is also referred to as knowledge transforming. It is described as an interactive process in which the writer develops new knowledge and new insights through attempting to solve both a content problem and a rhetorical problem. This type of writing involves different strategies and different cognitive skills than unskilled writing. Further, it does not develop naturally but requires training and practice. This is the case both for native and non-native speakers/writers.

Although both models described above were originally developed in the context of L1 writing, they have also been widely used as a basis for research into L2 writing. Some modifications have been suggested to account for issues concerning low-level writing or L2 writing; these will be described in further sections of this literature review.

4.1.2 Developing writing skills

Historically, teaching writing has not been a main focus of L2 instruction (Valdés, Haro, & Arriarza, 2011). In this context, researchers have stated that there is no unified, coherent theory of writing in a second language (Kroll, 2003; Polio & Williams, 2009). Matsuda (2003) attributes this lack of theory to the historically determined focus of linguistics on spoken language. Illustrating this point in a description of the history of L2 writing instruction, he describes how the English Language Institute (ELI) as the first ESL program assumed that writing would naturally follow when students had mastered other aspects of language, and therefore did not include writing instruction in its goals (Matsuda, 2003). Harklau (2002), too, explains that the origins of (US) linguistics in anthropology, in combination with a strong influence of Saussure, has led to the notion that spoken language is the main object of linguistic inquiry. Similarly, Polio and Williams (2009) describe that in Europe, oral communication has generally been considered a more important goal of instruction. However, writing has since gained importance in L2 instruction, which has also led to a more prominent role of research into L2 writing. Silva (Silva, 2016) emphasises the growing number of publications relating to L2 writing since the field's origins in the 1950s. Polio (2003) defines four large areas of research interest: L2 writers' texts, the writing process, the participants, and the social context. It should be noted, though, that most research into L2 writing has focused on higher-level (e.g. academic) writing, while the development of L2 writing skills at low levels has received less attention. In an overview of research on EFL writing in schools, Lee (2016) also stresses the scarcity of literature in the field.

While spoken (L1) language is acquired naturally, writing always has to be taught and is generally linked with formal instruction (Weigle, 2002). Learning to write covers a very wide spectrum. At one end, there are very basic, mechanical aspects, like holding a pen and graphically producing letters. At the other end of the spectrum is the production of long, well-composed texts, aimed at a particular audience with a particular communicative intention. Most everyday writing, however, falls somewhere in the middle. Professional writers make up a minority of the population. Still, many professions require a functional level of writing, in particular because written communication has become more important with the increased use of electronic media. Even in private communication, writing has gained in importance.

The process of learning to write in an L1 usually starts when spoken language skills are already developed, at least to some degree. Berninger (1996) argues that young developing writers are not "merely scaled-down versions of skilled adult writers" (p. 196) and proposes a number of modifications to the Hayes and Flower (1980) model to explain beginning and developing writing/young L1 writers. She states that the cognitive processes described in the original Hayes and Flower (1980) model may develop at different rates in young writers, with translation being the first, planning next and revising last. This may explain why, for example, planning and translating are not clearly separated in young writers. These processes emerge systematically as the writers' cognitive skills and working memory mature.

4.1.3 Assessing L2 writing

L2 writing skills consist of two components: writing skills and L2 proficiency. Depending on the purpose of a test and the level being tested, both components will play a role in assessment to varying degrees. For example, Weigle (2002, p. 79) argues that "in a test of writing for low-proficiency foreign-language learners, we may decide that we are primarily interested in linguistic and textual knowledge, or knowledge about the grammar and vocabulary of the language and how sentences are organised into texts, rather than functional or socio-linguistic knowledge". This may determine one of the reasons behind the paucity of literature on low-level L2 writing. At the lowest levels, we may be looking at individual words or sentences, and it may be difficult to draw a clear line between grammatical knowledge and early writing skills. Researchers interested in writing skills will find more material to investigate at higher levels of writing.

Similarly, Alanen et al. argue that "the constructs and measures which seem particularly adapted for capturing the growth of proficiency across, for example, the beginning stages of L2 writing proficiency, may not be as suitable for the more advanced levels of writing" (Alanen et al., 2012, p. 27). This shows that the level of language proficiency is an important factor to consider in the development of instruments for assessing writing. Generally, Weigle (2002) cautions that most of the research on L1 and L2 writing assessment has been conducted with limited samples of writing by adults in academic contexts. She thus stresses that, when generalising from previous research results, individual differences between different groups of writers need to be considered. This argument has distinct pertinence to tasks designed to assess writing in an international test where target participants have much more diverse traits.

A special case of assessing L2 acquisition in general and writing, in particular, is the context of immigration. Many European countries find themselves faced with challenging assessment issues concerning the growing numbers of immigrants. Carlsen (2017) reports that immigrants in Norway are required to prove that they have relevant language skills as a prerequisite for residency. However, this group of test-takers faces the double challenge of low general literacy and low test literacy, placing added challenges on writing task developers and test providers.

4.1.4 Research on tasks used in writing assessment

Weigle (2002) states that although tasks used in writing assessment are meant to resemble real-world tasks, their main function is to elicit a performance that allows testers to draw conclusions about a candidate's writing ability as defined by the construct. They are the tool that operationalise the construct. As such, tasks are crucial instruments and the test-taker's interaction with the task may affect the test results in a number of ways. Although a wealth of literature is available providing guidelines on how to develop writing tasks, there is very little research on the role of tasks themselves. One of the few examples is Kuiken and Vedder (2008), who investigated effects of task complexity on written productions' syntactic complexity, lexical variation, and accuracy, but found no significant interactions. Previous research also suggests that the connection between task characteristics and test-taker performance is unclear (Alanen et al., 2012). For example, task types that expert judges considered easier elicited lower-rated test-taker performance, while other, seemingly more difficult tasks, resulted in higher scores (Hamp-Lyons & Mathias, 1994). Familiarity with the task type may also affect performance. Weigle (2002) states that if task types are familiar to the test-takers, they can be mastered even by non-expert writers, though they might not discriminate well. Similarly, if the task type is unfamiliar to the test-taker, even skilled writers may under-perform.

Harsch and Rupp (2011) investigated the alignment of level-specific writing tasks with the CEFR. They state that in writing assessment, it is common to employ a multi-level approach where tasks are expected to elicit a range of performances spanning several levels of proficiency. This may be problematic in two ways. First, it is sometimes necessary to target not many, but one specific level. Also, in a more homogenous test population, level-specific tasks may allow for more exact and, consequently, more reliable measurement and in addition reduce negative emotional reactions from test-takers who are confronted with tasks that are too difficult for them (Grotjahn, 2017). This is clearly an issue with A1 and A2 level assessment. Second, as discussed above, different levels of proficiency may call for different constructs and measures, and multi-level tasks spanning from A1 to the C levels may not do justice to the differences that exist between different groups of language learners.

This raises the question how to assess writers at low proficiency levels in the paradigm of communicative language testing. As writers at low levels cannot be expected to utilise higher-order writing skills in an L2, test developers are restricted to tasks that, on the one hand, might focus more on language ability than on communicative effectiveness, and on the other hand, are perhaps less authentic than is desirable. Bachman and Palmer (1996, p. 23) define authenticity, one of the qualities of test usefulness, as "the degree of correspondence between the characteristics of a given language test task to the features of a target language use (TLU) task". It is likely, however, that language learners at low proficiency levels would not encounter many situations outside the classroom (or the test) where they would need to communicate in writing. Thus, authenticity may not be a pressing concern in the assessment of low-level writing.

4.1.5 Further sources

The discussion of literature above is based on books, book chapters, journal articles and publicly available conference presentations that deal with relevant aspects of assessing writing. As stated in the project proposal, other sources were also consulted for the aims of this literature review.

Course books and test developers' manuals provided no relevant information on the development of writing tasks but were restricted to sample tasks. This will be included in the task analysis that follows later in this report, as will information on technical detail like number of tasks, task types, guidelines for rating, description of bands, and test administration attained from these sources.

4.2 Task analysis

As outlined above, for the task analysis, relevant sections of the CEFR grid for Writing Tasks (Council of Europe, 2009) were used to analyse tests targeting A1 and/or A2 writing (or comparable levels) by international test providers. The results of the analysis are presented in this section.

In total, the analysis included 42 publicly available writing tasks from 21 tests targeting lower levels (CEFR A1 and/or A2 or comparable levels) by 10 international test providers (Table 1). All of the tests were for English. Seven tests were designed specifically for A1; ten for A2 (the CAMLA MET test targets multiple levels from A2 to C, however one of the two tasks in the test is aimed at A2); one for A1 and A2; and two tests are not specifically linked to the CEFR, however they are comparable to A1 (Trinity QCF ESOL Entry 1) and A2 (Trinity QCF ESOL Entry 2).

4.2.1 Number of tasks, completion time, integration of skills, channel

Although targeting the same low proficiency level, the 42 tasks analysed revealed great diversity in the following features; the number of tasks set to measure the level, the time allowed to completed the various tasks, the approach to assessment (discrete skills or integrated approach or a combination of both) and the channel of response (pen and paper, computer-based or web-based). The tests vary in terms of number of tasks at the different levels, from one task to three tasks at each level. In terms of time allotted to complete the tasks at the lower levels, the information provided by the test providers differs widely. In many cases, only an overall completion time for the test including skills other than writing is given and the time to complete individual writing tasks is not specified. For those tests where there is a clear indication for the writing tasks alone, either in the test paper itself, in the test specifications, or in other supplementary documents for the test, at A1 the time given varies from 3 minutes for one task (Aptis) to up to 20 minutes (Cambridge ESOL A1). For A2, the recommended completion time spans between 7 minutes per task (Aptis) and 40 minutes (Trinity ISE Foundation[2]). The review also revealed divergences in the approaches adopted. A number of tasks at each level are integrated tasks where test-takers have to either read or listen (or both) to a passage first and then base their writing on this input. The Pearson tests, in particular, rely on integrated tasks at the lower levels, with all three exams including such task types. Of the 21 tests included in the analysis, 15 do not make use of integrated tasks but test writing only (see Table 1 for details).

Results, however, showed more homogeneity regarding the channel of response. The majority of the performances on the tasks included in the analysis are handwritten rather than word-processed by test-takers. Only the Aptis Test allows for both handwritten and word-processed responses. A summary of the information gathered on these issues is provided in Table 1.

[2] Only one of the tasks in Trinity ISE Foundation is a reading into writing task, the second task is a 'pure' writing task. Test-takers are given 40 minutes for each task.

Table 1: Task analysis: Targeted levels, number of tasks per test, completion time, integration, channel of response

Test	CEFR	Tasks	Time (minutes)				Integr.	Channel
			overall	task 1	task 2	task 3		
APTIS General Writing Test A1-C	A1	1	55[2]	3			-	both[6]
Cambridge ESOL A1[1]	A1	3	40	10	10	20	-	handw.
EIKEN Grade 3 (A1)	A1	1	50[3]	10			-	handw.
Euro Exam International (A1)	A1	2	20	5	15		-	handw.
Lang. Cert. Intern. Preliminary Level A1	A1	2	80[3]	n.s.*	n.s.		-	handw.
Pearson PTE, Foundation Level A1	A1	2	75[4]	n.s.	n.s.		list, read	handw.
TELC A1	A1	2	15	n.s.[5]	n.s.		read	handw.
TELC A1 junior	A1	3	20[3]	n.s.[5]	n.s.[5]	n.s.	visual	handw.
APTIS General Writing Test A1–C	A2	1	55[2]	7			-	both[6]
Cambridge ESOL A2[1]	A2	3	50	15	15	20	-	handw.
Cambridge KET, A2	A2	1	70[3]	n.s.			-	handw.
CAMLA (MET)	A2	1	45[2]	n.s.			-	handw.
ECL Hungary (A2)	A2	2	45	n.s.	n.s.		-	handw.
EIKEN Grade Pre-2 (A2)	A2	1	75[3]	10			-	handw.
Euro Exam International (A2)	A2	2	30	15	15		-	handw.
Lang. Cert. Intern. Access Level A2	A2	2	80[3]	n.s.			read	handw.
Pearson PTE, Elementary Level A2	A2	3	95[4]	n.s.			list, read	handw.
TELC A2 school	A2	1	15	15			-	handw.
Trinity ISE Foundation	A2	2	80	40	40		read	handw.
Pearson EFB English for Business	A1-A2	1	90[3]	n.s.			read	handw.
Trinity QCF ESOL Entry 1	no link	3	30	n.s.	n.s.	n.s.	-	handw.
Trinity QCF ESOL Entry 2	no link	3	50	n.s.	n.s.	n.s.	-	handw.

*not specified
[1] *this format was cancelled in July 2017 and will be continued in July 2018*
[2] *for the complete writing component*
[3] *for reading and writing*
[4] *for reading, listening, and writing*
[5] *Although these tasks are included in the writing part of the exam, one could argue that they are reading tasks rather than writing tasks, as test-takers only need to copy words from a reading text.*
[6] *handwritten and word-processed*

4.2.2 Task types and target response length

In the scales and sub-scales describing what writers can do at various levels, the 2001 version of the CEFR explicitly states short, simple post cards and form filling for A1 and short, simple formulaic notes, very simple personal letters, short, simple notes and messages and telling a story or describing something in a simple list of points for A2. As part of the task analysis, the second set of features were the various task types used to assess lower-level writing, as well as the target response length expected.

For A1 (or comparable levels) a number of different writing genres are included by the different test providers in their exams. Form filling was identified as the most frequent task type used to assess writing competence at A1 level exams, with four tests including this format. Short narratives in the form of a picture story or a brief email or letter were the next most common task type identified, with two exam boards employing these genres respectively. Other genres included a simple message, a note, a postcard with one provider including candidates writing single individual sentences.

For A2 (or comparable levels), the most prominent task type is an email or letter, with six exams including this genre in their assessment tasks, followed by form filling (four exams), a note, and essay (two exams respectively). Other task types include writing a forum post, a postcard, a diary entry, or a short essay based on pictures.

A study of the stipulated target response lengths revealed a general increase in the required word length between A1 and A2. At A1, the average maximum number of words to be produced by test-takers is 34 words per task, whereas at A2 the average is nearly twice as much with 65 words per task. However, there is also great variability within each CEFR level. For those exams at A1 where the number of words is specified, the maximum number of words to be produced by test-takers varies between 5 (Aptis) and 80 words (Pearson PTE A1) per task, with the majority of exams targeting between 20 and 50 words. For A2, the maximum number of words spans between 30 (Aptis) and 100 (Pearson PTE A2 and Trinity ISE Foundation). These numbers also correspond to the two Trinity exams which are not linked to the CEFR but are targeting comparable levels. The tasks' targeted genre and the number of words to be produced by test-takers are shown in Table 2.

Table 2: Task analysis: Task type and target response length

Test	CEFR	Genre			Words		
		task 1	task 2	task 3	task 1	task 2	task 3
Aptis General Writing Test A1–C	A1	message			1-5		
Cambridge ESOL A1[1]	A1	form	note	n.s.	n.s.	20	50
EIKEN Grade 3 (A1)	A1	n.s.			25-35		
Euro Exam International (A1)	A1	form	picture story		10-20	50	
Lang. Cert. Intern. Preliminary Level A1	A1	sentences	letter		30	30	
Pearson PTE, Foundation Level A1	A1	postcard	picture story		30-50	50-80	
TELC A1	A1	form[1]	email		n.s.[1]	30	
TELC A1 junior	A1	n.s.[1]	n.s.[1]	form	n.s.[1]	n.s.[1]	6
Aptis General Writing Test A1–C	A2	form			20-30		
Cambridge ESOL A2[1]	A2	form	letter	note	n.s.	50	80
Cambridge KET, A2	A2	note			25-35		
CAMLA (MET)	A2	n.s.	n.s.		n.s.	n.s.	
ECL Hungary (A2)	A2	forum post	n.s.		75	75	
EIKEN Grade Pre-2 (A2)	A2	essay			50-60		
Euro Exam International (A2)	A2	form	email/letter		33-48	50	
Lang. Cert. Intern. Access Level A2	A2	email	letter		30-50	30-50	
Pearson PTE, Elementary Level A2	A2	email	diary entry[2]		50-70	80-100	
TELC A2 school	A2	postcard			n.s.		
Trinity ISE Foundation	A2	article	email/essay		70-100	70-100	
Pearson EFB English for Business	A1-A2	n.s./message/form			n.s.		
Trinity QCF ESOL Entry 1	n.l.	form	postcard	n.s.	n.s.	30	30
Trinity QCF ESOL Entry 2	n.l.	form	article	email	n.s.	80	80

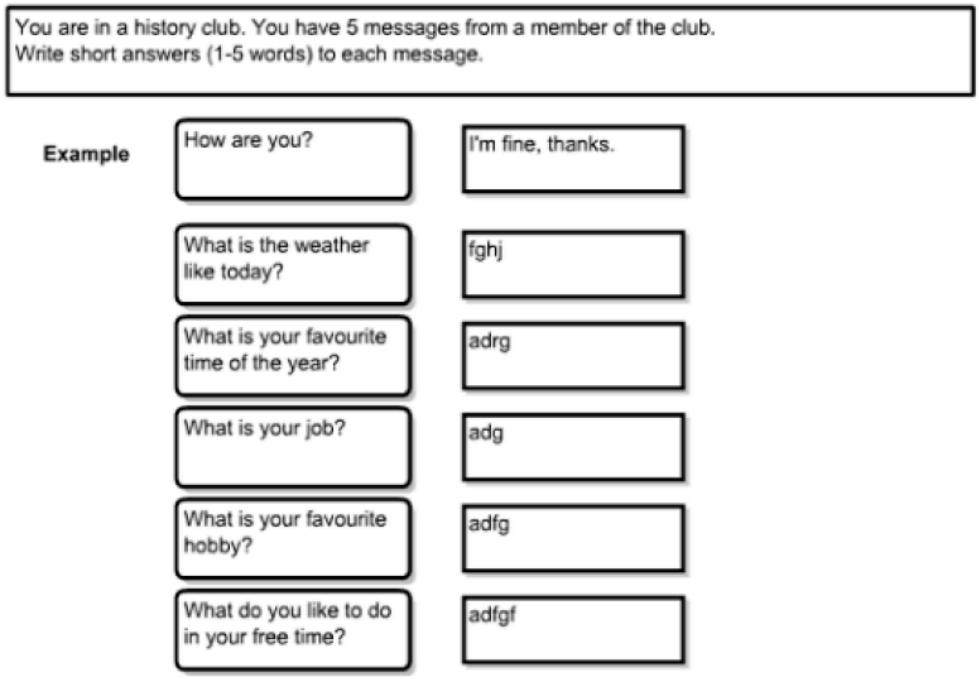
[1] *As mentioned in Table 1 above, these tasks are closer to being reading tasks rather than writing tasks.*
[2] *based on pictures*

4.2.3 Task authenticity

As was identified in the literature review, task authenticity presents special challenges in assessing writing at the lower levels. This study also analysed the tasks' situational and interactional authenticity. In the CEFR grid used for the analysis, situational authenticity is defined as "[t]he extent to which the task reflects a real-life activity a candidate could perform", while interactional authenticity is "[t]he extent to which interaction patterns are likely to mirror those in an equivalent, real-life task" (Council of Europe, 2009, p. 4; see also Appendix 1). For each of the two categories (situational authenticity and interactional authenticity), these definitions were adopted and applied as criteria on a three-point task authenticity scale of low, medium, or high. Initially, two researchers independently judged each of the sample task's authenticity. They agreed in 81% of all cases in their judgments. The researchers then discussed the small number of cases where there were discrepancies until agreement was achieved.

In order to illustrate how the tasks were judged, the judgment process will be outlined in detail for one Aptis A1 writing task (Figure 6), where both researchers agreed on the level of situational and interactional authenticity in their individual judgments. The basic situation of the task (i.e. replying to five messages on an online platform to introduce yourself as a new member of a history club) was considered authentic by both researchers. However, some of the specific questions asked were considered less authentic for this situation, for example "What is the weather like today?" or "What is your favourite time of year?" Thus, overall, situational authenticity was judged as "medium" by both researchers. For interactional authenticity, both researchers considered the interaction pattern of the task authentic (i.e. exchanging messages on an online platform with other members of the same club), so interactional authenticity was judged as "high".

Figure 6: Aptis A1 example task to illustrate authenticity judgments



As shown in Table 3, for situational authenticity the sample included tasks judged at the three levels, low, medium and high. In order to detect possible differences between A1 and A2, we assigned each label a number (1=low, 2=medium, 3=high) and calculated the mean level of authenticity for each level (A1 and A2, or comparable levels in the Trinity ESOL exams). For situational authenticity, the mean judgment for A1 was 2.05; for A2 it was 2.23. Thus, on average, the tasks of both levels were judged at medium situational authenticity, whereby A1 tasks were judged to be slightly less situationally authentic than A2 tasks.

For interactional authenticity, the judgments also ranged from low to high, however, 36% of the tasks (15 out of 42) were judged not to be interactive in nature, i.e. there is either no intended readership other than the writer him/herself for the text to be produced (e.g. diary entry), or the intended readership is not clear (e.g. "Write a story"), or the task is transactional in nature, meaning that the intended readership is not highly relevant for task completion (e.g. filling in personal details on a form). Again, the mean was calculated for each level (A1 and A2) by assigning each label a number, as outlined above. The tasks with no interaction were assigned 0. The mean judgment for A1 was 1.42; for A2 it was 1.77. Thus, on average, interactional authenticity was judged to be lower than situational authenticity, particularly for A1 tasks. For interactional authenticity, the increase between A1 and A2 was bigger than for situational authenticity (1.42 / 1.77 compared to 2.05 / 2.23).

Table 3: Task analysis: Task authenticity

Test	CEFR	Authenticity					
		Situational			Interactional		
		task 1	task 2	task 3	task 1	task 2	task 3
APTIS General Writing Test A1–C	A1	medium			high		
Cambridge ESOL A1	A1	high	high	high	n.a.	high	high
EIKEN Grade 3 (A1)	A1	low			low		
Euro Exam International (A1)	A1	medium	low		medium	n.a.	
Lang. Cert. Intern. Preliminary Level A1	A1	low	medium		n.a.	high	
Pearson PTE, Foundation Level A1	A1	medium	low		high	n.a.	
TELC A1	A1	medium[1]	medium		n.a.[1]	high	
TELC A1 junior	A1	low[1]	high[1]	high	n.a.[1]	n.a.[1]	n.a.
APTIS General Writing Test A1–C	A2	high			high		
Cambridge ESOL A2	A2	high	high	high	n.a.	high	high
Cambridge KET, A2	A2	medium			medium		
CAMLA (MET)	A2	low	low		n.a.	n.a.	
ECL Hungary (A2)	A2	high	low		high	n.a.	
EIKEN Grade Pre-2 (A2)	A2	low			low		
Euro Exam International (A2)	A2	high	medium		n.a.	high	
Lang. Cert. Intern. Access Level A2	A2	high	high		high	high	
Pearson PTE, Elementary Level A2	A2	medium	medium		high	n.a.	
TELC A2 school	A2	high			high		
Trinity ISE Foundation	A2	medium	medium		n.a.	high	
Pearson EFB English for Business	A1-A2	medium			high		
Trinity QCF ESOL Entry 1	n.l.	medium	high	medium	n.a.	high	high
Trinity QCF ESOL Entry 2	n.l.	medium	medium	medium	n.a.	high	high

[1] *Not applicable, i.e. the task is not interactive in nature (e.g. form filling)*
[2] *As mentioned above, these tasks are closer to being reading tasks rather than writing tasks*

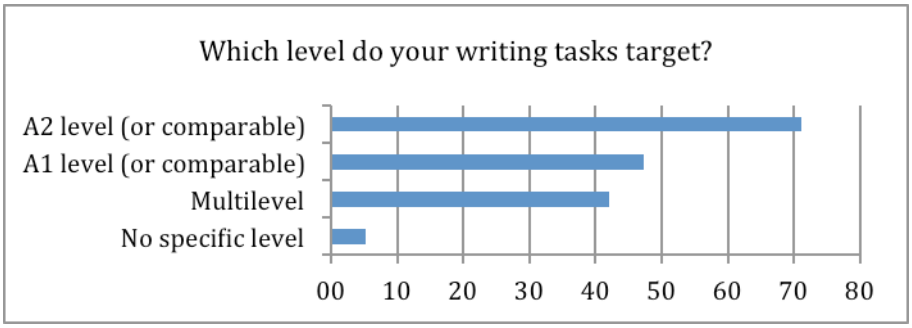
4.3 Survey

The findings are presented in four parts. The first part outlines results to questions about which lower levels the participants assess, how many tasks they include in their writing tests, and how the tests are designed, administered, and rated. The second part presents findings on questions about task types used in lower-level writing assessment and the target response length for these task types at the different levels. Part 3 outlines challenges and problems faced by the participants when assessing lower-level writing. Finally, in part 4 results to questions related to the CEFR are presented. Participants were asked whether they used the CEFR for assessment purposes, how useful they find it, whether their assessments are linked to the CEFR, and about the extended CEFR descriptors published in autumn 2017 and their potential for assessing lower level writing.

4.3.1 Targeted levels, number of tasks, task design, administration, and rating

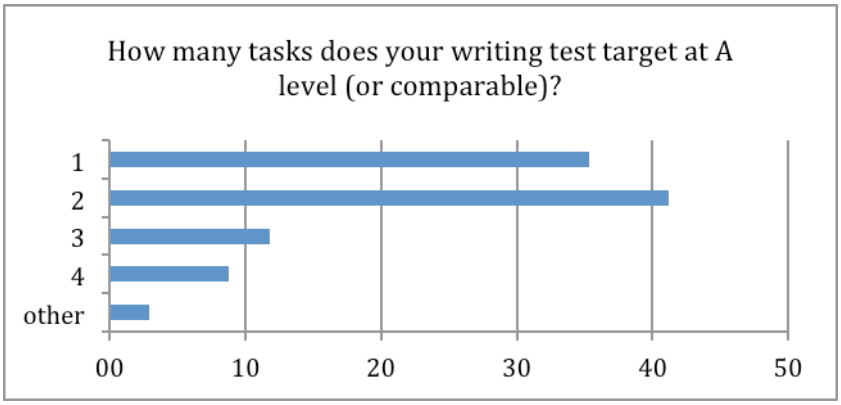
In terms of targeted levels (Figure 7), 71% of the participants designed tests for A2 (or a comparable level). About half of the sample (47%) designed writing tasks specifically targeting A1 (or a comparable level), while 42% developed tasks targeting more than one level. Only two participants stated that their tasks did not target any particular level.

Figure 7: Levels targeted by participants (N=38, in % of cases, multiple answers possible)



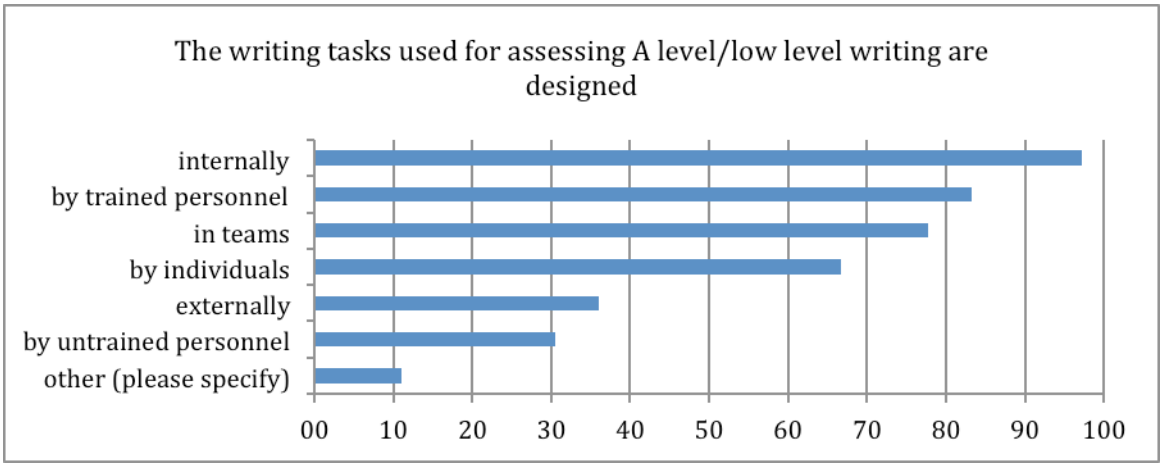
As shown in Figure 8, the majority of respondents include one or two tasks at the lower levels in their writing tests. Four of the participants stated that their writing tests include three tasks aimed at the lower levels, with another three participants designing writing tests with four tasks specifically aimed at lower-level writing. The remaining participant did not include tasks specifically targeting lower levels. From those who replied 'other', one participant specified that their test included two tasks at A1 and one task at A2. One participant replied that "all tasks are for all levels".

Figure 8: Number of tasks at lower levels (N=34, in % of cases)



When it comes to designing the tasks (Figure 9), the majority of participants stated that this is done internally at their institution (N=35), by trained personnel (N=30) and in teams (N=28). However, a large number of participants also stated that their lower-level writing tasks are designed by individuals (N=24). A number of participants also use tasks developed externally (N= 13) or tasks which are developed by untrained personnel (N=11). One participant specified that tasks were developed "by the department of quality analysis or, rarely, the inspectorate", and another stated that the tasks used for timed assessment were designed by the curriculum unit.

Figure 9: Task design parameters (N=36, in % of cases, multiple answers possible)

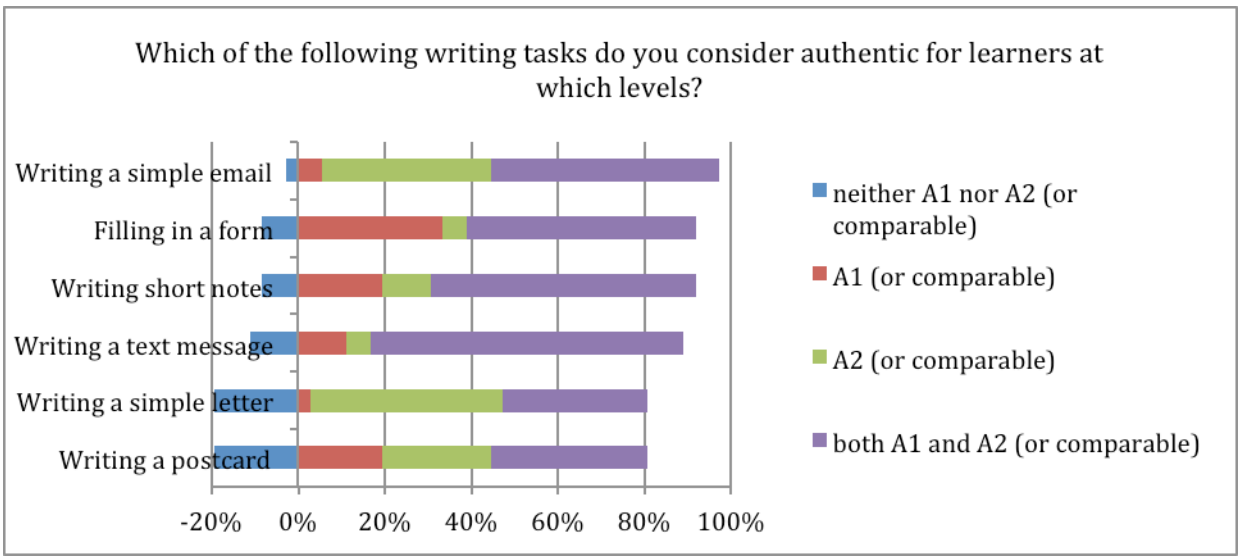


The majority of participants administer their writing tasks by pen and paper (N=24), while only a small group uses computer administration (N=3). Seven participants stated that they use both pen and paper and computer administration. When it comes to rating, the majority of respondents use a rating scale (88%, N=30), with 56% using an analytic scale and 32% using a holistic scale.

4.3.2 Task types and target response length

Respondents were asked to indicate which task types they consider authentic for lower-level writing assessment, from a given number of task types taken from the original A1 and A2 CEFR descriptors. The results are presented in Figure 10. As can be seen, all of the task types mentioned in the original CEFR descriptors are considered authentic by the majority of participants, for either A1 or A2 (or comparable levels), or both. The most authentic task type for A1 is considered to be "Filling in a form", with 33% agreement across respondents. For A2, "Writing a simple letter" (44% agreement) or "Writing a simple email" (39%) are regarded as the most authentic task types mentioned in the original CEFR. However, in general, the majority of respondents indicated that the task types are authentic for both A1 and A2 (or comparable levels), except for "Writing a simple letter", which was seen by most as an authentic task type for A2.

Figure 10: Task types participants consider authentic for lower-level writing assessment (N=36, in % of cases)



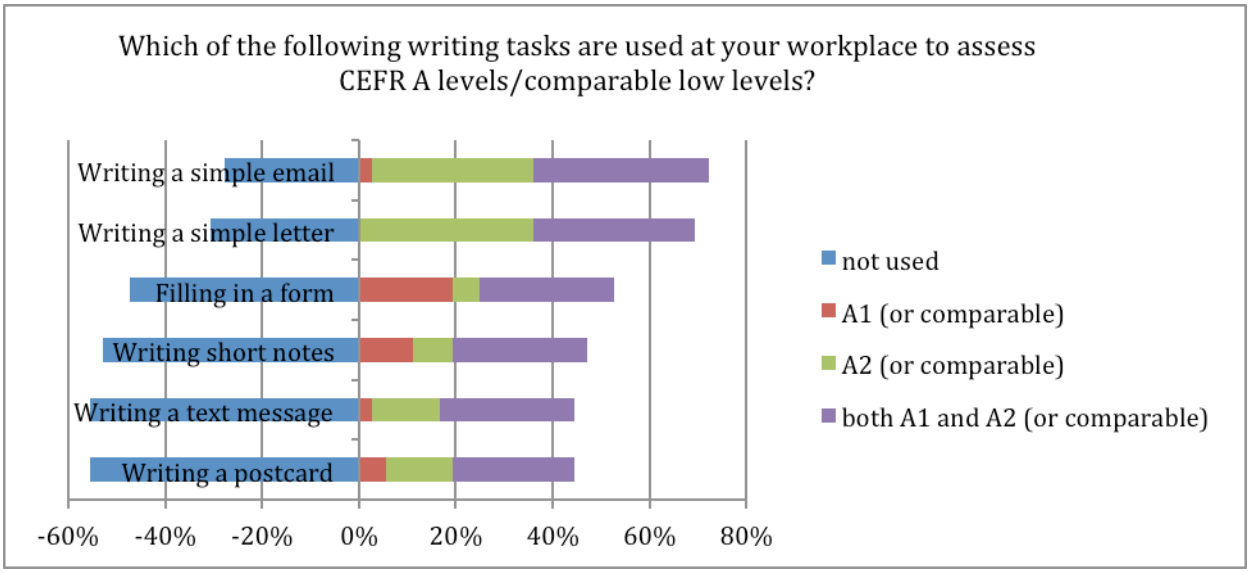
Following up this item, participants were further asked whether there were any other task types they considered authentic for use at the A-levels. Eighteen (18) task types were given (although the term "task type" may be loosely applied in this context), and for most of these it was not specified whether they were considered authentic for A1 or A2 level. Five responses referred to writing in combination with a visual stimulus: "writing a caption for photos", "labelling pictures into longer sequences", "writing a comment under a YouTube video", "adding labels, e.g. to a map to give directions", and "preparing a poster (A2)". The participants mentioned another 13 authentic task types:

- making a list
- writing an invitation (A2)
- recipes (A2)
- statements of personal views and opinions
- simple descriptions
- diary entries
- directions and instructions
- magazine articles
- simple blog entry
- creating mind maps to describe something
- writing short, structured narratives
- copying (with computer) complete sentences from a >300<600-word text to construct (usually by slightly reorganising) a kind of précis – level depends on text used – can be CEFR A
- writing something on the Internet.

Participants were also asked which task types they use at their workplace to assess lower-level writing. They could again choose from the task types mentioned in the original CEFR, or note down any additional task types they use. Interestingly, although the great majority of respondents agreed that the task types mentioned in the original CEFR are authentic for lower-level writing assessment (see Figure 10 above), three of these task types are used by less than half of the respondents (Figure 11). The task types used by most respondents are "Writing a simple email" and "Writing a simple letter". These are used to assess both A1 and A2 (or comparable levels) by about a third of the respondents, while another third uses them only for A2. One respondent stated that they use "Writing a simple email" for A1 (or a comparable level). "Filling in a form" is used by 19% (N=7) of respondents for A1 (or a comparable level), by 6% (N=2) for A2 (or a comparable level), and by 28% (N=10) for both A1 and A2 (or comparable levels).

The remaining three task types from the original CEFR are used by less than half of the respondents, and the majority of those who use them do so for both A1 and A2 (or comparable levels). Participants were further asked to specify which other task types were used at their workplace for the assessment of A-level writing. Results include descriptive or narrative texts like short narrations or picture/photo descriptions; concrete text types like recipes, CVs, lists; and task types focusing on the mechanics, rather than the content, of writing (dictation and "answering written questions").

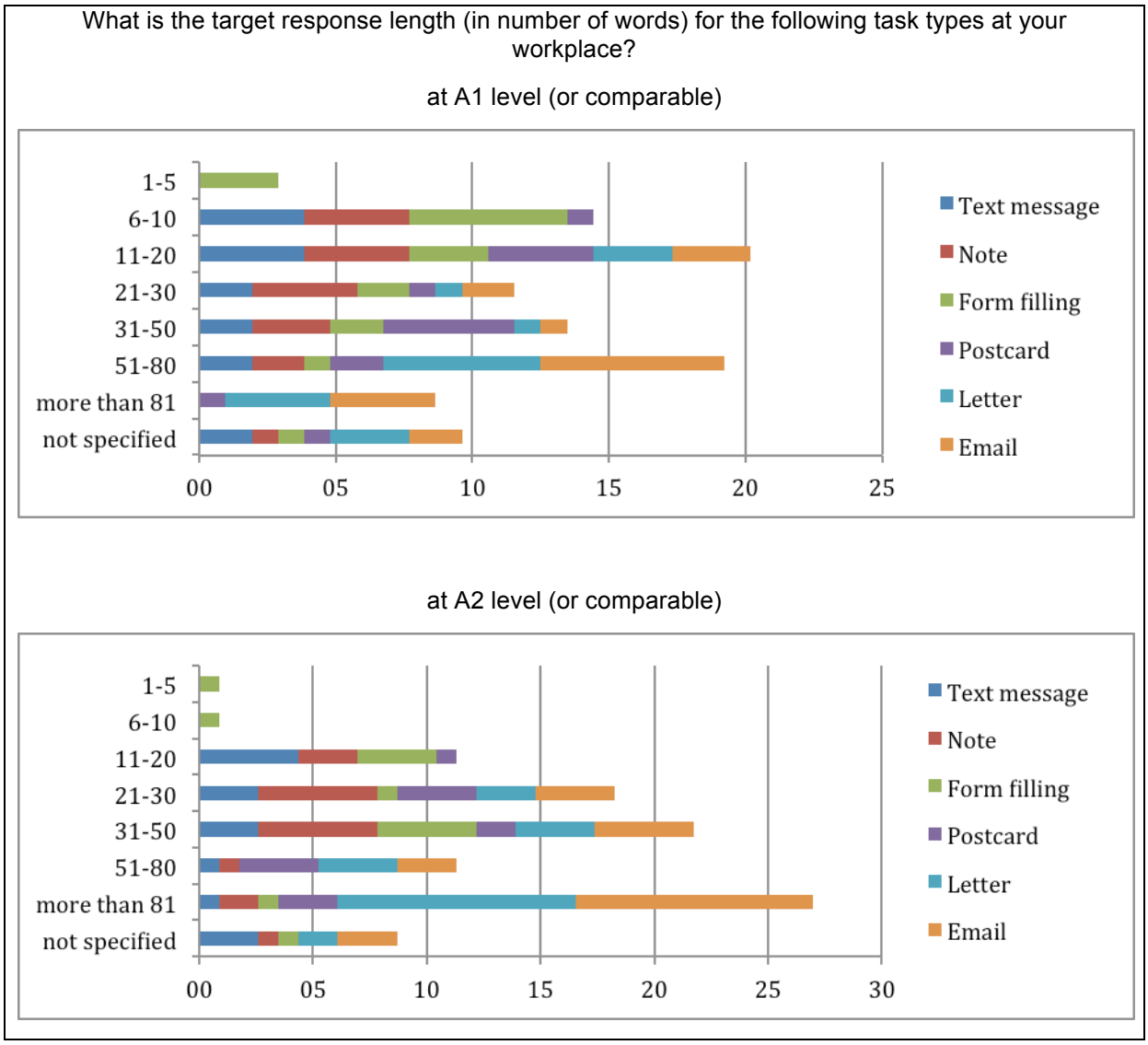
Figure 11: Task types in the original CEFR used by participants for lower-level writing assessment (N=36, in % of cases)



Respondents were not only asked which task types they use, but also about the tasks' target response length, both at A1 and A2, for the task types included in the original CEFR (Figure 12). As outlined above, about half of the participants did not use the task types mentioned in the original CEFR in their assessment, therefore the numbers in Figure 12 are based on the responses of 15 to 19 participants (depending on the task type). Percentages were calculated based on the total number of responses for each level. It can be seen that, as expected, the average target response length increases between A1 and A2. The majority of respondents target 11 to 20 words at A1 level (20%), but more than 81 words at A2 (27%). Still, a sizable number of participants stated that their writing tasks at A1 (or a comparable level) target 51 to 80 words (19%), while 22% indicated that they target 31 to 50 words at A2. Therefore, although the general trend seems to be to target a longer response length at A2 than A1, there seems to be some disagreement about how many words students should produce at the two levels. There was also considerable disagreement regarding the target response length of the different task types, as shown in Figure 12. For example, at A1, the target response length for "email" varies from 11 to 20 words up to more than 81 words, with a number of participants also choosing the intermediate steps. This trend is similar for the other task types, as well as the task types at A2.

For this set of questions, participants also had the option to give additional task types and text lengths. An analysis of these answers supports results from the closed questions insofar as the text lengths indicated for A1 (A1: 1–5 answering questions, 11–20 directions, story narration, 51–80 descriptions and short messages, more than 81 for stories) are shorter than those for A2 tasks (A2: 4–6 sentences narrative essay, 50–80 or 80–100 descriptive paragraph, 120 or more for story). One participant stated, "we don't specify how long they should write".

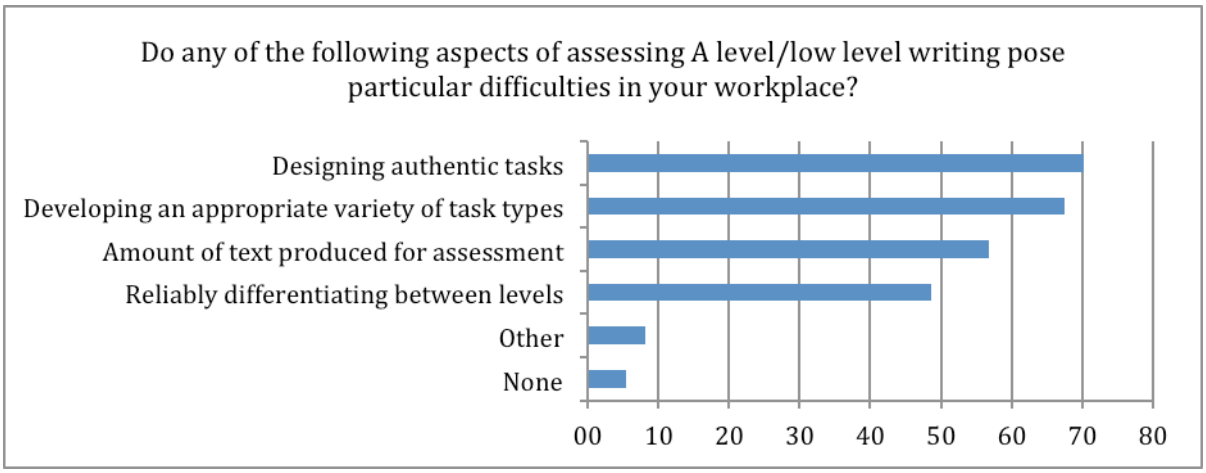
Figure 12: Target response length of task types in the original CEFR at A1 and A2 (or comparable levels) (N=15–19 (depending on task type), in % of cases of total number of responses at each level)



4.3.3 Challenges in assessing lower-level writing

In terms of challenges encountered when assessing lower-level writing (Figure 13), the areas participants struggle most are in finding authentic tasks (N=26) and developing an appropriate variety of task types (N=25). Another challenge indicated by more than half of the respondents is to specify an appropriate amount of text to be produced at the lower levels (N=21), which echoes the findings outlined above. Almost half of the participants also struggle with differentiating between the levels, while only two participants do not face any particular challenges when assessing lower-level writing.

Figure 13: Challenges of assessing lower-level writing (N=36, in % of cases, multiple answers possible)



The particular challenges of assessing A-level/low-level writing were also addressed in an open-ended question in the survey. The answers given by participants contained 57 individual responses that were coded into three broad categories: challenges relating to tasks, participants' interaction with the task, and challenges relating to the assessment process. The latter two categories are connected by the overarching problem of the limited proficiency of test-takers at this level.

Challenges relating to tasks were mentioned 11 times. The largest part of the comments referred to choosing appropriate tasks, as is illustrated by this example:

> *The first issue is how to construct a meaningful task for low levels, and how to provide a wide variety of topics.*

Although a high number of participants mentioned "finding authentic tasks" as a particular challenge in the corresponding closed item, only three participants commented on this issue in the open question. However, one of them suggested that authentic writing tasks were difficult to find because at this level:

> *people don't usually write in their daily lives.*

The second category of responses refers to test-takers' problems in responding to tasks, caused by their limited language proficiency. With 26 out of a total of 57 comments, this category has the highest number of responses. Four participants commented that test-takers might have problems understanding the task. Eight replies referred to the amount of text produced by test-takers as being problematic, as is illustrated by this example:

> *The test-takers often do not produce enough language to properly assess the different components of writing ability.*

Fourteen comments explicitly mentioned that the test-takers' limited proficiency affected task completion. As concrete examples, they mentioned: spelling, punctuation rules, L1 influence, and basic writing skills. One participant replied that:

> *Students have insufficient knowledge of vocabulary to write a text – in fact they write some basic sentences.*

Another participant stated that:

> *at low level, students mostly are not able to write coherent texts, or simply cannot express themselves in the foreign language (most assessment criteria simply don't take this into account).*

Finally, 18 responses concerned problems relating to the assessment process. Five participants mentioned that it was difficult to find or design adequate assessment scales for this level. Others referred to difficulties in rating performances of low-level writers:

> *Raters can place too much emphasis on errors.*

One participant also mentioned that:

> *fair assessment that summarises all necessary aspects of Writing at A1 without regarding other weaknesses is difficult.*

This was echoed by three other replies concerning difficulties in differentiating between levels. Two other participants, who were using the CEFR as a basis for their assessment, mentioned that they considered the CEFR inadequate for the assessment of children.

When asked whether they experienced any problems in rating lower level performances, 59% of participants said they did, while 41% did not. Asked to specify what problems they experienced, respondents referred in equal parts to inadequate rating scales and problems in reliably differentiating between levels. They also mentioned raters' expectations.

> *Teachers sometimes find it hard to estimate what a good performance is at A1/A2-level. They sometimes expect too much accuracy which is actually not expected yet at this level.*

Again, the length of output produced by test-takers at A-levels was mentioned as problematic.

> *A1 texts are often too short to be assessed with scale. Assessing coherence/cohesion especially difficult at low level.*

After giving their opinions on the particular challenges of assessing low-level/A-level writing, participants were further asked what they do to address these challenges. Replies again fall into three larger categories: (a) measures focusing on task or test design (16 comments); (b) measures referring to the assessment or rating process (18 comments); and (c) activities in teaching to combat the limitations of L2 proficiency and writing skills, as mentioned in the previous question (14 comments).

Most comments in category (a) refer to the development of writing tasks. Team work, piloting tasks, following guidelines, using CEFR descriptors as a basis for task design are mentioned as ways to improve task design. Other answers showed that participants tried to offer a variety of tasks and to align assessment tasks:

> *(…) to material used to teach at these levels in order to see what types of tasks students at these levels are exposed to.*

Five replies mentioned measures referring to test design, e.g. administering several tasks of varying difficulty, starting with the easiest, and using teamwork to design adequate tests.

Category (b), referring to the assessment or rating process, contains 18 comments. Ten of these refer to using rating scales or assessment criteria. Six participants state that they create their own assessment criteria/scale. Of the six participants who mentioned the rating process, half stressed the importance of rater standardisation and rater training. One participant states:

> *two raters would help but not available.*

Three respondents stress the importance of feedback to the test-takers in a school context.

Finally, category (c) subsumes activities to react to problems related to test-takers' language proficiency and writing skills. In equal parts, respondents stated that they ensured test-takers had understood what was required of them, "Using first language when necessary", or that they addressed particular weaknesses that affected the quality of writing:

> *I try to teach linking words, good transitions and vocabulary.*

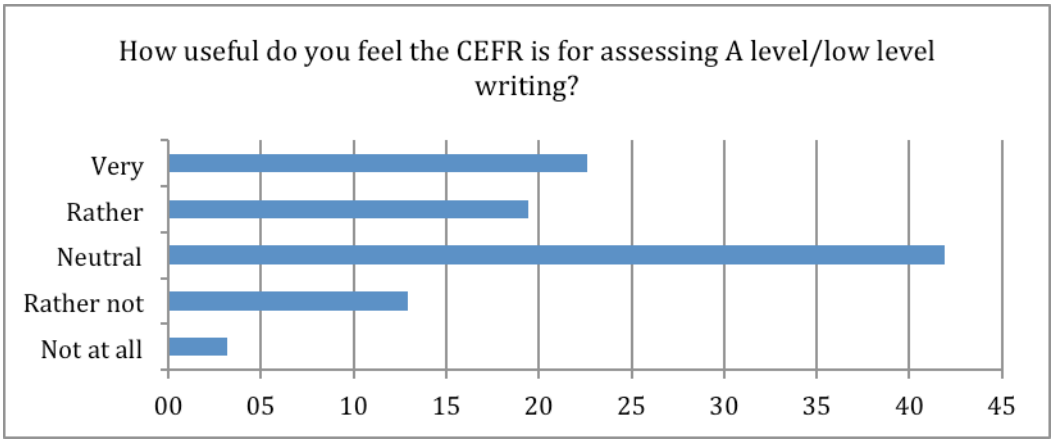
> *Punctuation/capitalisation exercises (…) and providing key sentence structures they might need or collocations.*

Nearly two thirds of all respondents in this category stated, sometimes in great detail, that they explicitly taught writing skills, which is surprising given that a lack of writing skills was not previously mentioned as a main challenge. For example, they mentioned providing their students with model sentences or paragraphs, training them in peer- and self-correction and writing draft versions of their texts, brainstorming for content before starting to write, and exercises in paragraph structuring.

4.3.4 CEFR linkage and extended CEFR descriptors

Participants were asked whether they use the CEFR as a basis for assessing writing. Of the participants, 76% (N=32) said they did, while 24% (N=10) did not. Of those who did, 42% (N=13) found the CEFR useful (Figure 14). The same number of participants felt neutral, while 16% (N= 5) indicated that the CEFR was not useful. One answer was missing.

Figure 14: Usefulness of CEFR for assessing lower-level writing (N=31, in % of cases)



When asked to specify how the CEFR was useful for assessing lower-level writing, participants mentioned that it gave them concrete ideas for the kind of content and activities that test-takers could be expected to cope with:

> *It helps set goals or learning outcomes. For example, at A1 students need to greet and meet, introduce oneself, use English for survival needs.*

Participants also commented that the CEFR is used as a basis for assessment scales and is helpful for distinguishing between levels. Those participants who had stated that they did not consider the CEFR useful for writing assessment at low levels gave as reasons that it was unspecific and "not designed with low-level writing in mind, much less kids". This echoes a sentiment that was mentioned by two participants in an earlier question on challenges of assessing low-level writing.

Those who use the framework were also asked whether they are required to link their assessment of writing to the CEFR (Figure 15) and which aspects they include in their linkage (Figure 16). The majority of respondents link their assessment to the CEFR, either by generally accepted practice (55%, N=16), by educational law (34%, N=10), or by internal policy documents (21%, N=6). Rating scales and text types are linked to the CEFR by 72% of respondents (N=21), test specifications by 69% (N=20), and choice of topic by 66% (N=19). Although 21% (N=6) stated that they are not required to link their assessment to the CEFR (Figure 15), only two participants do not link any specific aspects to the CEFR (Figure 16).

Figure 15: CEFR linkage – requirements (N=29, in % of cases, multiple answers possible)

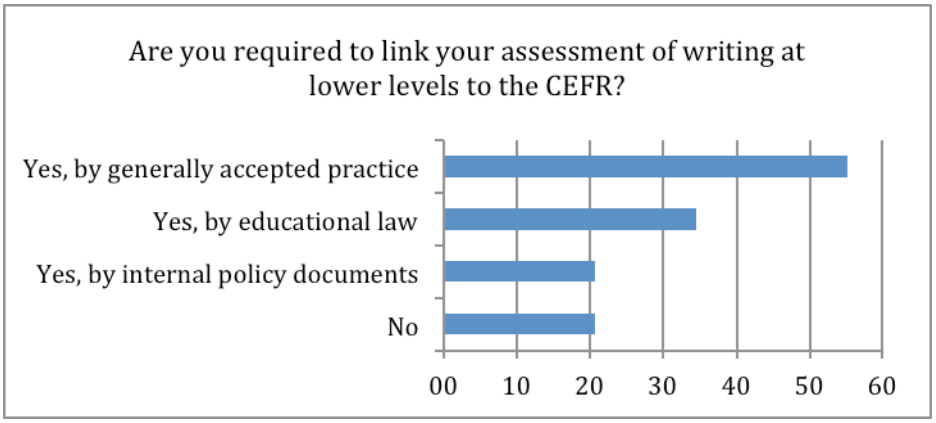
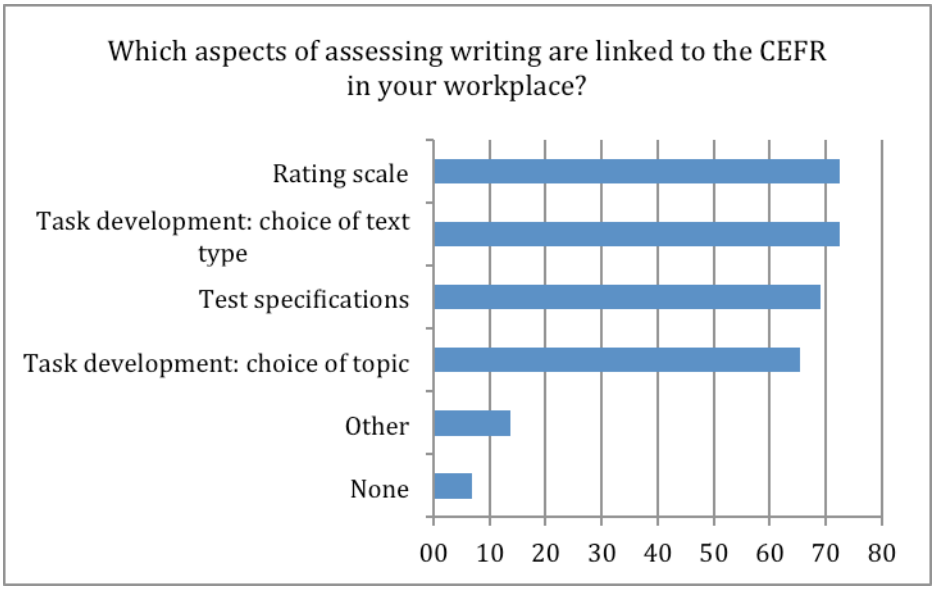


Figure 16: CEFR linkage – which aspects (N=29, in % of cases, multiple answers possible)



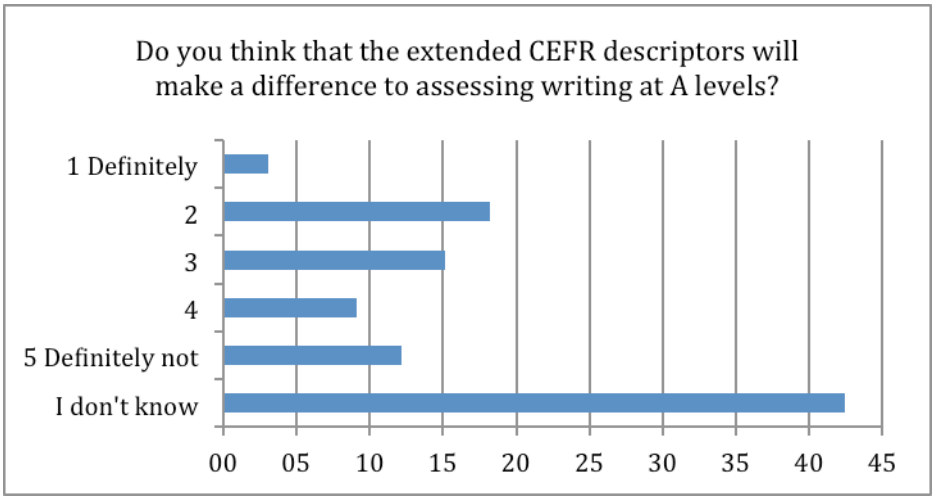
A number of questions related to the extended CEFR descriptors published earlier in 2018. Of the respondents, 74% (N=25) stated that they were familiar with the new descriptors, while 26 % said they were not (N=9), with eight answers missing. When asked whether they thought that the extended descriptors would make a difference to assessing writing at lower levels (Figure 17), most of the respondents did not know (42%, N=14). Twenty-one percent (21%, N=7) felt positive that the new descriptors would make a difference, while an equal number felt that they would not. Five participants (15%) felt neutral to the question. Those who were positive about an effect thought that the extended descriptors would be more specific and clarify the level descriptions. The reasons given by respondents who were more skeptical included criticism about the new descriptors, as is shown in these two examples.

> *They're too specific and not relevant to our teaching context.*
>
> *They seem to expect more from A-level than was previously the case – I think this may cause problems.*

Others expressed fundamental doubt about the usefulness of descriptors.

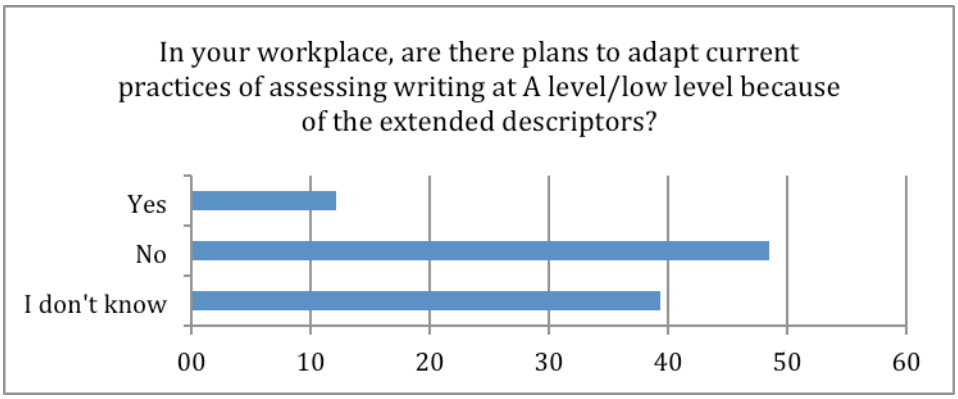
> *By their very nature descriptors aren't concrete; they're stereotypifications. In the end, they don't give useful information to either each individual learner or others. Concrete comments work.*

Figure 17: Impact of extended CEFR descriptors for assessing writing at lower levels (N=33, in % of cases)



Participants were also asked whether there are plans to adapt current practices of assessing lower-level writing at their workplace in light of the extended descriptors (Figure 18). The majority of respondents said that there were no plans to change their practices (49%, N=16). Twelve percent (12%, N=4) indicated that there are plans to change their practices, while 39% (N=13) did not know. Of those who answered that were plans to adapt practices, three stated that the concrete natures of changes had not yet been decided; while one participant indicated that new task types would be developed based on the new descriptors.

Figure 18: Changes in assessing writing at lower levels because of extended CEFR descriptors (N=33, in % of cases)



4.4 Comparison between original and extended writing related A1/A2 CEFR sub-scales

To get an overview of the new opportunities presented by the extended CEFR descriptors for assessing lower-level writing, we compared all writing related A1 and A2 CEFR scales of the original version of the CEFR with the extended version. Overall, the lower levels were extended by 43 new writing related descriptors, including the two new sub-scales on online interaction, as shown in Table 4 (for a detailed list of all new descriptors see Appendix 2). Most descriptors were added at A2 level, with 23 new descriptors across eight different sub-scales. At A1 level, a total of 11 new descriptors were added across eight different sub-scales. Finally, nine new descriptors across eight different sub-scales were formulated and validated for the new pre-A1 level.

Table 4: Number of new writing related descriptors for A1 and A2 in the extended CEFR

CEFR sub-scale	Number of new descriptors			
	pre-A1	A1	A2	
			A2.1	A2.2
global scale				
self-assessment grid				
overall written production	1	1		
creative writing		2	2	1
reports and essays			2	
overall written interaction	1			
correspondence	1	2	3	1
notes, messages and forms	1	1	2	
note taking			1	
processing text				
general linguistic range	1	1		
vocabulary range				
vocabulary control				
grammatical accuracy	1			
orthographic control		1		
sociolinguistic appropriateness				
flexibility				
thematic development			1	
coherence and cohesion				
online conversation and discussion	2	2	2	3
goal-oriented online transactions and collaboration	1	1	3	2
Subtotal	**9**	**11**	**23**	
Total	**43**			

Legend:
Lower levels (A1 and A2) remain the same in the extended CEFR (additions only in the higher levels)
No separation into lower and higher level in the original CEFR, but in the extended CEFR
No descriptors in the original CEFR, but in the extended CEFR

5. DISCUSSION AND RECOMMENDATIONS

Based on the findings presented above, each of the four research questions will be discussed in turn. Where appropriate, we add recommendations drawn from our findings.

5.1 Research question 1

What is the current state of research regarding the development of task types assessing low-level writing?

An extensive literature search identified two main research gaps. First, there is little published research on L2 writing at the lowest levels as equivalent to CEFR A1 and A2. One possible reason behind this phenomenon, as mentioned in the literature review, may be the historical development of linguistics as well as L2 instruction, which focused on spoken rather than written language. Also, research into L2 writing in general appears to focus on higher-level writing, particularly in the academic context. It could perhaps be postulated that in the absence of substantial linguistic production for analysis at lower levels in writing, the research community itself has neglected or overlooked the needs of practitioners and test providers involved in initial stages of writing development.

Second, this study found no published empirical research on the design of writing tasks and task types. Although writing tasks are being used to generate data for assessment as well as for research, neither their design nor their impact on the performance produced, have been adequately researched to date. This confirms the initial impression voiced at the outset of this report. Results from the survey and from the task analysis suggest, however, that more guidance would be useful. Despite clear descriptions of the two A-levels in the CEFR, operationalising the descriptors to design tasks appears to be challenging, as is discussed in more detail below.

The findings from this present research highlight the need for more research into both the link between the test construct and the task which is designed to elicit an appropriate performance, and also the relationship between the suitable task and task types at the lower levels and the linguistic and communicative features identifiable in the resultant performances, i.e. evidence of the construct.

5.2 Research question 2

What types of writing tasks are currently used in international, national, and local contexts to assess low-level writing (CEFR A1 and/or A2 or comparable levels)?

This research question was investigated through an analysis of 42 publicly available writing tasks from 21 writing exams by 10 international test providers, as well as a survey on lower-level writing assessment practices completed by 42 language test developer across the globe.

Before answering the research question, we will briefly explore how many writing tasks test developers deem appropriate to assess lower-level writing. The results of the task analysis mostly correspond to the information gathered in survey. The analysed tests vary from one task to three tasks at each level (A1 and A2 or comparable), while in the survey, participants indicated that their tests include one to four tasks at the lower levels. While multi-level tests such as Aptis only include one task per level, tests specifically targeting a lower level, for example, Pearson's PTE Elementary Level A2, include up to three tasks. However, there are also tests specifically targeting A1 or A2 which only include one writing task per level, such as EIKEN or TELC A2 School. For tests only including one task at each of the lowest levels, the number of task types as well as the amount of text produced, and thus ultimately the construct representation of the lower levels, are necessarily limited.

It might therefore be advisable to include at least two tasks at each of the lower levels, particularly for tests targeting a specific level. For multi-level tests such as Aptis, this might be difficult to achieve due to time constraints. In addition, the majority of Aptis test-takers may be more advanced and construct representation at the lower levels might therefore not have a high priority. Still, the test developers should be aware that the construct is only partially covered when using only one task.

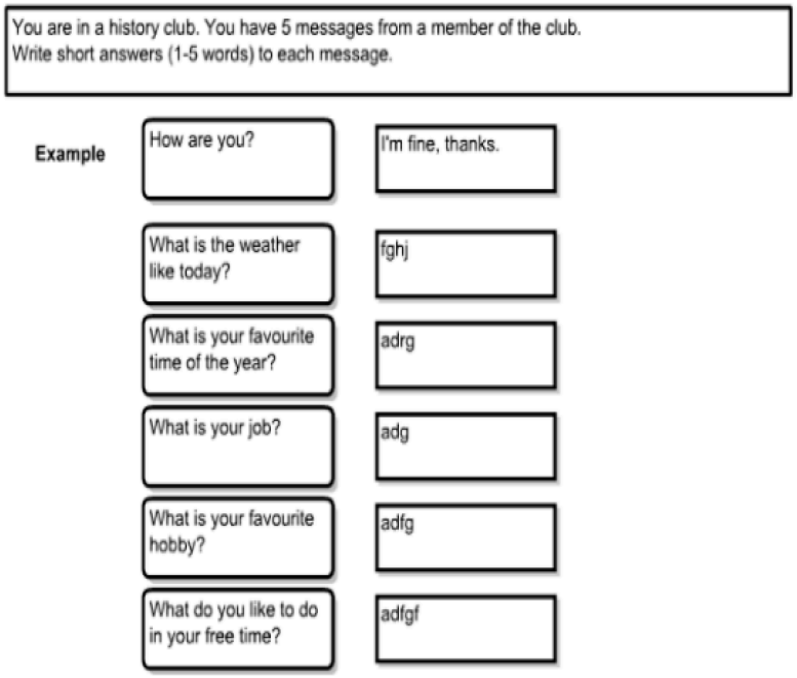
When it comes to the demands of the tasks on the test-taker in terms of task completion time and the number of words to be produced, there is great variation among the analysed tests as well as the survey responses. Although the task completion time generally increases between A1 and A2, at A1 it varies between 3 minutes (Aptis) and 20 minutes (Cambridge ESOL A1) per task and at A2 between 7 minutes (Aptis) and 40 minutes (Trinity ISE Foundation). A general increase was also observed in the target number of words between A1 and A2 both in the task analysis as well as the survey results; however, variation was again substantial. At A1, the maximum target response length spans between 5 words (Aptis) and 80 words (Pearson PTE A1) per task, with the majority of exams targeting between 20 and 50 words. At A2, the maximum number of words to be produced spans between 30 words (Aptis) and 100 words (Pearson PTE A2 and Trinity ISE Foundation). These numbers are very similar to the information provided by the survey respondents and were also confirmed by their responses on the question about which aspects pose particular difficulties, where 57% indicated that they found it challenging to specify the amount of text to be written. Thus, overall, our findings indicate that test providers are uncertain about how much they can expect from test-takers for lower-level writing assessment in terms of the target response length. This has implications for fairness towards test-takers. Depending on which test they take, they are faced with very different demands within the same levels, even if these tests claim CEFR linkage. This also raises doubts about how comparable these results are, seeing that A1 in one test is clearly interpreted very differently from A1 in another test.

It should also be mentioned that the Aptis writing tests at A1 and A2 have the lowest demands on test-takers in relation to the time test-takers should spend on the tasks (3 minutes on the A1 task and 7 on the A2 task) as well as the number of words they need to produce (5 words on the A1 task and 20 in the A2 task), compared to all other tests included in the analysis. This might again have to do with the fact that the Aptis writing test is the only exam included in the analysis that targets multiple levels and includes tasks up to the C levels, with the higher levels necessarily taking up a bigger proportion of the overall exam time. Still, the Aptis test developers may want to consider including tasks with higher demands in terms of number of words to be produced at the lower levels. Although the CEFR does not include specific guidance in terms of number of words, only targeting five words in a form filling task might not meet the demands of some of the extended A1 descriptors, as described below.

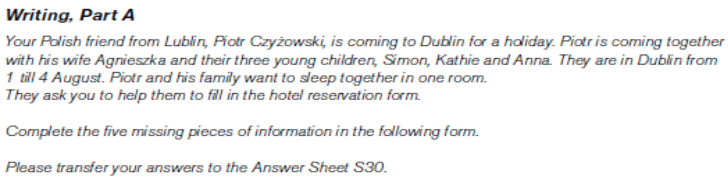
The task analysis has also shown that, with an increase in the target number of words, the tasks generally become more complex and the demands on the test-taker increase, particularly at A1 level. For example, in the exemplary Aptis and TELC A1 tasks in Figure 19, test-takers only need to answer a number of short questions in a few isolated words. While in the Aptis task, test-takers at least need to come up with the answers themselves, in the TELC task they only need to copy information from the input and change “Polish” to “Poland”, which makes it more of a reading task than a writing task. In contrast, the demands on the test-taker are much higher in the Euro Exams International and Pearson PTE Foundation A1 tasks, which ask test-takers to write a story based on a picture in 50 to 80 words (Figure 20). Not only do candidates need to produce more text, but they also need to demonstrate their skills in all processes involved in writing (planning, translation and reviewing, as suggested by Hayes & Flower, 1980). This shows that test developers seem to be uncertain about how much they can expect from test-takers, as they interpret the A1 CEFR descriptors very differently.

Figure 19: 'Simple' A1 tasks: Aptis A1 and TELC A1

Aptis A1



TELC A1



Writing, Part A

Your Polish friend from Lublin, Piotr Czyżowski, is coming to Dublin for a holiday. Piotr is coming together with his wife Agnieszka and their three young children, Simon, Kathie and Anna. They are in Dublin from 1 till 4 August. Piotr and his family want to sleep together in one room.
They ask you to help them to fill in the hotel reservation form.

Complete the five missing pieces of information in the following form.

Please transfer your answers to the Answer Sheet S30.

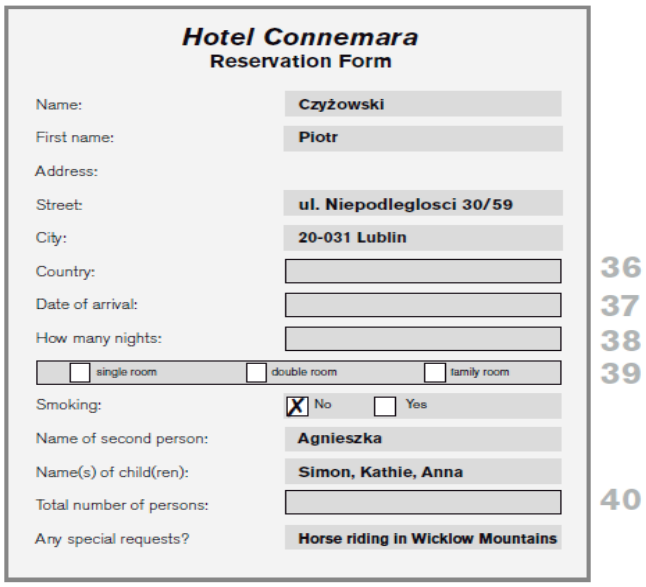
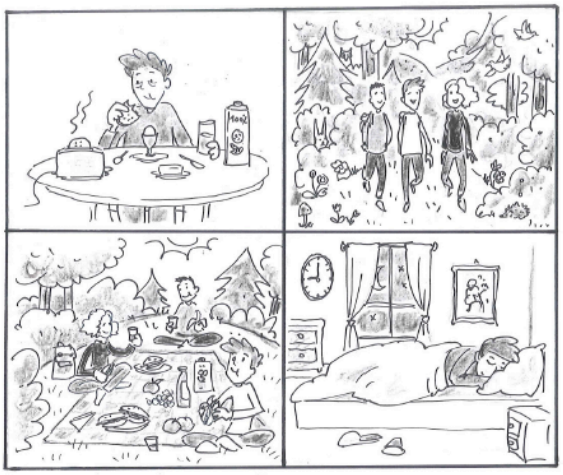


Figure 20: 'Complex' A1 tasks: Euro Exam International A1 and Pearson PTE Foundation A1

Euro Exams International A1

Task Two: Picture Story *(15 minutes)*

Write a story based on the pictures in at least 4 sentences in about 50 words.



The story starts like this:

One Sunday morning Ken got up early.

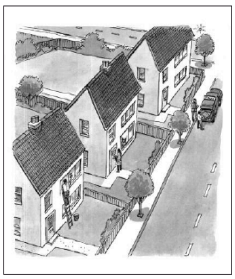
Pearson PTE Foundation A1

Choose one of the pictures below and write your answer in 50 – 80 words.

48. A) Write a short story about the picture.

Or

48. B) Write a short story about the picture.



In terms of types of tasks used, the results of the task analysis again echoed the findings of the survey. The original CEFR suggests short, simple postcards and filling in of basic forms as suitable task types for A1, whilst at A2 language learners can additionally be expected to write "short, simple notes and messages relating to matters in areas of immediate need" (Council of Europe, 2001, p. 26), as well as very short personal letters. The task analysis showed that large test providers generally appear to stay close to these recommendations; however, they also use additional task types not mentioned in the CEFR. For A1, form filling, postcard, picture story, notes, and email/letter are used. At A2, these task types continue to be used, but additional task types including article, diary entry, forum post, and essay are now also introduced. Responses from the survey were similar. Participants indicated that they found form filling, postcard, notes, messages, and email/letter to be authentic task types for both A1 and A2. However, interestingly, three of these task types (postcard, notes, and message) are used by less than half of the respondents in their assessments. There are also a number of additional task types participants found authentic for assessing lower-level writing, such as article, diary entry, forum post ("writing something on the Internet") or blog post, although participants did not indicate for which level these would be more suitable and whether or not they use them in their assessments. Overall, the findings suggest that the task types specifically mentioned in the original CEFR descriptors, at either A1 or A2, are considered suitable for either level, but that there are also a number of additional task types not mentioned in the CEFR which can be used.

For A2, test developers seem to use a greater variety of task types, many of which are not mentioned in the original CEFR descriptors.

5.3 Research question 3

What strategies do test developers employ to deal with the limitations of the CEFR in the context of low-level writing assessment (CEFR A1 and/or A2 or comparable levels)?

Two aspects need to be considered before discussing RQ3. First, results from the survey show that 42% of participants considered the CEFR useful for assessing low-level writing, and the same percentage felt neutral. Only 16% responded negatively to this item. This suggests that respondents did not feel limited by the CEFR. At the same time, the survey clearly showed that the assessment of low-level writing poses particular challenges, as is illustrated by the large number of answers given to the open-ended questions addressing this issue. From this, we may conclude that the limitations and challenges that are undeniably present do not come from the CEFR, but are inherent in the level that is being assessed. Therefore, it may be necessary to rephrase to original research question and ask: **What strategies do test developers employ to deal with the challenges of assessing low-level writing?** The difficulties mentioned by the survey's participants can be grouped into four categories: authenticity, task types, test-taker performance, and rating. Each of these issues will now be addressed in more depth.

Authenticity

The biggest challenge, according to the survey results, is the notion of authenticity. Situational authenticity, as defined by the CEFR, compares the given writing task to a real-life situation that a test-taker could have to perform in a non-test situation. Particularly for writers at A1 there are only very few situations that would authentically require written communication. The task analysis also showed that many of the tasks used displayed little to no interactional authenticity, in particular those task types that have been added to those suggested by the CEFR. Weigle (2002) argues that in some circumstances a trade-off between authenticity and other qualities of test usefulness, like reliability, may be necessary. From our research, it appears that although test developers involved in classroom testing claimed to look for authentic tasks, they tended to value variety of task types higher. Large-scale test developers worked with a smaller number of task types, but it was shown in the analysis that these were not necessarily more authentic, particularly in terms of interactional authenticity. One conclusion to draw from these results is that authenticity at low levels may not be of central concern. It may, however, be advisable for test developers to decide on minimal requirements for authenticity of their tasks, and to observe and collect publicly available examples of low-level writing, or, to take a step back, investigate whether and to what amount A-level writing can be observed, e.g. in online forums or similar platforms.

Task type

Related to authenticity is the issue of task types. Nearly 70% of respondents perceived it as difficult to design an appropriate variety of task types. This may be explained by the fact that most survey participants were also teaching, hence their assessment would most likely be conducted in a classroom context with a higher need for variety than in large-scale assessment, which is an additional challenge faced in this particular assessment context. Interestingly, also 70% of survey respondents claimed to refer to the CEFR for choosing appropriate task types, but the results outlined above show little evidence of this.

In addition, nearly all of the task types used at the respondents' workplaces are used for both A1 and A2, and survey participants indicated that differentiating between levels was a problem for them, which also confirms findings by Harsch and Rupp (2011). Using the same task types for both levels may contribute to the difficulty of differentiation. This result may again indicate that test developers either do not have a clear idea of what can be expected at each level or, as was suggested by one participant in personal communication, knowingly exceed the requirements of A1 writing because

they feel that at this level, there is not enough L2 writing competency to be assessed. To some extent, this was also confirmed by the task analysis, where a number of test providers seem to expect too much from test-takers at A1 level (see also Figure 20 above). This seems a rather problematic strategy of dealing with the limitations of assessing low-level writing.

Despite these findings, a number of limitations with regards to authenticity and task types need to be acknowledged. First, survey respondents were not asked whether they were able to or saw a need to differentiate between situational and interactional authenticity, so authenticity may have been variously interpreted by the respondents. Second, although most large-scale test providers included in the task analysis seem to have a very broad definition of their target test-takers, it was not investigated in detail for which specific target group of learners some of the assessments are designed (whether, for example, immigrants to a country or young learners), partly because this information was often not available. This could also impact on task design and hence on the perceived authenticity and appropriateness of task. Third, a more detailed look at the writing prompt may have shed additional light on the issue of authenticity, for example whether the prompt limits the type of tasks that are used at lower levels of writing. Future research on authenticity could also focus more on the test-takers and their perceptions of the tasks, as suggested by Lewkowicz (2000). This, however, was beyond the scope of the paper.

Test-taker performance

The third category of challenges refers to test-taker performance. A high number of responses from survey participants referring to this issue suggests that they perceive it as a pressing problem to fairly rate A-level writers, either because the texts they produce are very short, or because obvious linguistic limitations make it difficult to give due credit to each performance. This, again, is an inherent problem of A-level writing, which is characterised by very short and very simple texts, very basic expressions, some basic structures, and “only limited control of a few simple grammatical structures” (Council of Europe, 2001, p. 114). From the survey responses, it appears that participants are aware of this; nevertheless, this puts them in a difficult situation when assessment is expected to focus on communicative effectiveness and on what the test-taker can (rather than cannot yet) do. Rather than solving this dilemma, it appears that the chosen strategy is to work on increasing test-takers’ performance; for teachers, this is an obvious choice. Unfortunately, the task analysis does not give us any data about the strategies employed by large-scale test developers, or whether this issue is perceived as problematic in large-scale testing at all.

Rating

Finally, rating emerges as a challenging aspect of low-level writing from the survey. This includes difficulty in finding or designing adequate assessment scales and reliably differentiating between levels. Of the 88% of survey participants who use rating scales, 56% use an analytic and 32% a holistic one. The strategies mentioned in the survey include creating one’s own scales/rating criteria, rater standardisation, and rater training. However, it appears that the scales available are not perceived as adequate. Large-scale test providers also tend to use analytic scales. Pearson, for example, uses an analytic scale containing the criteria range, accuracy, coherence and cohesion, and orthographic control, with the addition of a global section score for written interaction or production. The descriptors are taken from the A1 CEFR descriptors; however, it seems that the task types used by Pearson at A1 would fit better with the CEFR’s recommendations for A2 (see also discussion below). The MET writing scale is also analytic, applying the following categories: grammatical accuracy, vocabulary, mechanics, cohesion and organisation, and task completion. Yet, survey participants frequently mentioned that the performances they received were too short to be assessed with a scale. It may be of interest to follow up whether holistic scales at low levels, particularly at A1, are more user friendly, and possibly more adapted to the purpose, than analytic ones. Although analytic scales have advantages, particularly regarding the reliability of ratings, feedback from the survey suggests that writing at A1 particularly may be too short and too restricted for the application of several criteria. A holistic scale that considers the performance’s general communicative success may present a more authentic approach (Weigle, 2002). The Aptis task specifications, for example, also stress that for A1 tasks “the focus in on communicative competence”.

5.4 Research question 4

What potential do the extended CEFR descriptors offer to adapt assessment practices of low-level writing?

In order to answer this research question, the extended CEFR descriptors were revisited in light of the task analysis and survey results. A number of potential improvements to assessment practices could be identified.

First, the differentiation between A1 and A2, which was listed as a challenge by test developers and was also found to be problematic for large-scale test providers as identified in the task analysis, may become clearer with the additions to the original CEFR descriptors. With 43 supplementary descriptors, the writing related A1 and A2 scales now offer more guidance for task design. The exemplary tasks in Figure 19 and Figure 20 above will be used to illustrate the potential for clearer differentiation between levels. The 'simple' A1 tasks in Figure 20, when judged against selected extended CEFR descriptors, would perhaps be closer to targeting pre-A1 than A1 writing (pre-A1: "Can use isolated words and basic expressions in order to give simple information about him/herself"; A1: "Can use some basic structures in one-clause sentences with some omission or reduction of elements") (Council of Europe, 2018, p. 131). The tasks in Figure 20, on the other hand, may now be closer to targeting A2 rather than A1 writing, according to selected extended CEFR descriptors (A2: "Can tell a simple story (e.g. about events on a holiday or about life in the distant future"; A1: "Can use simple words and phrases to describe certain everyday objects (for example, the colour of a car, whether it is big or small") (Council of Europe, 2018, p. 76). It is also noteworthy that the two new sub-scales on online interaction not only include a total of 16 new descriptors for the lower levels, but that the descriptors for A2 are also split into "lower A2" and "higher A2" descriptors, which might again be useful for test providers to differentiate more clearly between the lower levels.

Another potential improvement relates to the number of different task types at the two lower CEFR levels, particularly at A2. As identified in the survey responses, many test developers struggle to find an appropriate variety of task types for assessing writing at the lower levels. The extended descriptors now include two additional task types at A1, (text) messages and online postings, and eight additional task types or varieties of task types at A2: story, diary entry, text message, email, greetings card, notes on a presentation, notes on events, and personal and other details on a form. In addition, the new scale on "reports and essays" now also includes two descriptors for A2, and the two new scales on online interaction also include descriptors which can be included in revised test specifications and potentially be turned into new task types. While a number of these new task types have already been used by exam boards as identified in the task analysis (diary entry, essay), the specific mention of them in CEFR descriptors might help test developers in refining these task types and also in better differentiating between A1 and A2.

Finally, the two new sub-scales on online interaction might also help to mitigate some of the problems associated with authenticity. Particularly interactional authenticity, which was identified as problematic in the task analysis, could benefit from the additional descriptors provided in the extended CEFR. As described above, some of the new descriptors could be developed into new interactive task types. However, achieving full authenticity might still be difficult, particularly at pre-A1 and A1 level, as test-takers at this proficiency level may not often communicate in writing in real life.

6. CONCLUSION

This study looked at low-level writing assessment from a number of different perspectives. A literature review revealed that research into assessing writing at the lower levels has been sparse, both into task design and into the language produced, despite a clear need by practitioners. To address this need, an analysis of 42 writing tasks targeting CEFR A1 and A2 (or comparable levels) from 10 international test providers was conducted and a survey on test development practices on lower-level writing assessment was distributed online and filled in by 42 test developers across the globe. In addition, the extended CEFR descriptors (Council of Europe, 2018) were examined in light of the task analysis and survey findings to identify potential improvements to assessment practices.

The results revealed a number of challenges associated with lower-level writing assessment. It was shown that test developers often struggle to differentiate between A1 and A2 in writing task design, as they seem to interpret the original CEFR descriptors differently. The problem of differentiation also resonated in our finding that the same task types were used to assess both A1 and A2 writing. There was, however, some indication that test developers may react to the limitations of A1 writing by consciously exceeding requirements in their assessment. This, in particular, would warrant further research in this area. Responses from the survey also indicated that small-scale test developers found it challenging to apply existing scales to the performances of L2 writers at low levels. Particularly challenging was the principle of authenticity, especially interactional authenticity, as indicated by both the task analysis and the survey responses.

The analysis of the extended CEFR descriptors revealed a number of potential improvements in assessment practices. It was argued that the new descriptors could help to better differentiate between the lower levels, to find additional task types, and to mitigate some of the shortcomings in relation to interactional authenticity. The clearer differentiation may increase test fairness, if test providers come to a more uniform understanding about the demands they can put on test-takers at the A-levels.

Overall, our investigation into current practices of assessing A-level writing has exposed a number of challenges. It may be possible that some of these stem from treating low-level writing as a scaled-down version of the kind of writing that is done at higher levels, which, despite the CEFR's focus on can-do statements, is defined more by its deficiencies. Coming back to Berninger's (1996) appeal to acknowledge young developing (L1) writers in their own right, it may be time to take a fresh look at L2 low-level writing as having its own, distinctive set of characteristics.

REFERENCES

Alanen, R., Huhta, A., Martin, M., Tarnanen, M., Mäntylä, K., Kalaja, P., & Palviainen, Å. (2012). Designing and assessing L2 writing tasks across CEFR proficiency levels. In *Learning to Write Effectively: Current Trends in European Research* (pp. 221–223). http://doi.org/10.1163/9781780529295_050

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A Process Model of Writing Development across the Life Span. *Educational Psychology Review*, *8*(3), 193–218.

Carlsen, C. H. (2017). Language testing with a bitter aftertaste. Paper presented at the *EALTA Conference*, Sèvres, France.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe. (2006). *Survey on the use of the Common European Framework of Reference for Languages (CEFR). Synthesis of results.* Strasbourg, France: Council of Europe. Retrieved from: http://www.coe.int/T/DG4/Linguistic/Source/Surveyresults.pdf

Council of Europe. (2008). *The CEFR Grid for Writing Tasks v.3.1 (presentation)*. Strasbourg, France: Council of Europe.

Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. Strasbourg, France: Council of Europe.

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment companion volume with new descriptors*. Strasbourg, France: Council of Europe.

Grotjahn, R. (2017). Testkonstrukt und Testspezifikationen [Test construct and test specifications]. In B. Akukwe, R. Grotjahn, & S. Schipolowski (Eds.), *Schreibkompetenzen in der Fremdsprache. Aufgabengestaltung, kriterienorientierte Bewertung und Feedback* (pp. 71–116). Tübingen: Narr.

Hamp-Lyons, L., & Mathias, S. P. (1994). Examining expert judgments of task difficulty on essay tests. *Journal of Second Language Writing*, *3*(1), 49–68. http://doi.org/10.1016/1060-3743(94)90005-1

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, *11*(4), 329–350. http://doi.org/10.1016/S1060-3743(02)00091-7

Harsch, C. (2007). *Der Gemeinsame Europäische Referenzrahmen Für Sprachen : Leistung Und Grenzen. [The Common European Framework of Reference for Languages: Strengths and Limitations].* Saarbrücken, Germany: VDM.

Harsch, C., & Rupp, A. A. (2011). Designing and scaling level-specific writing tasks in alignment with the CEFR: A test-centered approach. *Language Assessment Quarterly*, *8*(1), 1–33. http://doi.org/10.1080/15434303.2010.535575

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & E. R. Steinberg (Eds.), *The Science of Writing*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 31–50). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Jones, N., & Saville, N. (2009). European language policy: Assessment, learning, and the CEFR. *Annual Review of Applied Linguistics*, *29*(2009), 51–63. http://doi.org/10.1017/S0267190509090059

Kroll, B. (2003). Introduction. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing*. Cambridge: Cambridge University Press.

Lee, I. (2016). EFL writing in schools. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of Second and Foreign Language*. Berlin.

Lewkowicz, J. A. (2000). Authenticity in language testing: Some outstanding questions. *Language Testing*, *17*(1), 43–64. http://doi.org/10.1177/026553220001700102

Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, *91*(4), 645–653.

Martyniuk, W., & Noijons, J. (2007). *Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe Member States*. Strasbourg, France: Council of Europe. Retrieved from http://www.coe.int/T/DG4/Linguistic/Source/Survey_CEFR_2007_EN.doc

Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 15–34). Cambridge: Cambridge University Press.

O'Sullivan, B. (2015). *Linking the Aptis Reporting Scales to the CEFR*. Technical report. English Language Assessment Research Group. London: British Council. https://www.britishcouncil.org/exam/aptis/research/publications/reporting-scales

O'Sullivan, B., & Dunlea, J. (2015). *Aptis General Technical Manual, Version 1.0*. Technical report. English Language Assessment Research Group. London: British Council. https://www.britishcouncil.org/aptis-general-technical-manual-version-10

Olson, K. (2010). An Examination of Questionnaire Evaluation by Expert Reviewers. *Field Methods*, *22*(4), 295–318. http://doi.org/10.1177/1525822X10379795

Polio, C. (2003). Research on second language writing: An overview of what we investigate and how. In B. Kroll (Ed.), *Exploring the Dynamics of Second language Writing* (pp. 35–69). Cambridge: Cambridge University Press.

Polio, C., & Williams, J. (2009). Teaching and Testing Writing. In C. J. Long, Michael H.; Doughty (Eds.), *The Handbook of Language Teaching* (pp. 486–517). Blackwell Publishing Ltd.

Silva, T. (2016). An overview of the development of the infrastructure of second language writing. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of Second and Foreign Language Writing* (pp. 19–43). Berlin: De Gruyter.

Valdés, G., Haro, P., & Arriarza, M. P. E. (2011). The Development of Writing Abilities in a Foreign Language: Contributions toward a General Theory of L2 Writing. *The Modern Language Journal*, *76*(3), 333–352.

Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

APPENDIX 1: The CEFR Grid for Writing Tasks v. 3.1 (presentation)

*The original template for this grid was developed by ALTE members. www.alte.org

This grid has been developed in order to assist test providers in their work with the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)* and the *Manual for Relating Language Examinations to the CEFR*, both available from the Language Policy Division of the Council of Europe.

There are two varieties of this grid: the **analysis** grid and the **presentation** grid (this one). The **analysis** grid is intended to be used in workshops, benchmarking events and for other activities where participants at the event are asked to complete the grid. In these cases, the activities are likely to be part of the *standardisation of judgments* stage described in the *Manual*. Where the grid is to provide a descriptive record of test tasks, the **presentation** grid may be more suitable. If completed grids are intended to be offered as illustrative samples, their final use will relate to *standardisation of judgments* stage of the *Manual*. On the other hand, if test providers wish to analyse test content and specifications for their own purposes, the relevant stage is *specification*.

Sample Test Tasks

Report on analysis of	
Target language of this test	
Target level (CEFR) of this test	
Task number/name	

General Information – the whole test

1	Total test time	minutes
2	Purpose	

3 Background to the examination
4 Candidature
5 Structure of the test

General Information – the writing component

6	Number of tasks in the writing paper	
7	Total component time	minutes
8	Integration of skills	
9	Channel	
10	CEFR level of this component	

11 The writing component format
12 Specific Information – example task
13 Mark distribution
14 Task rating
15 Effective level

16 Sample task:

– sample task here --

i) Task input/prompt		
17	Language of input/prompt	
18	CEFR level of input/prompt	
19	Time permitted or suggested for this task	minutes
20	Control/guidance	
21	Content	
22	Genre	
23	Rhetorical function(s) of input	
24	Imagined audience	
25	Mode of input/prompt	
26	Topic or theme of input	
27	Integration of skills for input	

ii) Response (description of written response elicited by the prompt(s)/input)		
28	Number of words expected	
29	Rhetorical function(s) expected	
30	Text purpose	
31	Register	
32	Domain	
33	Grammatical competence expected	
34	Lexical competence expected	
35	Discoursal competence expected	
36	Authenticity: situational	
37	Authenticity: interactional	
38	Cognitive processing	
39	Content knowledge required	

iii) Rating of task		
40	Known criteria	
41	Task rating method	
42	Assessment criteria	
43	Number and combination of raters	

iv) Feedback to candidates		
44	Quantitative feedback	
45	Qualitative feedback	

46 Example answer
47 Commentary
48 Score allocated

All references to the *CEFR* are to the document on the Council of Europe's Language Policy Division's website.

2 The purpose of the test may be **general proficiency**, for a specific purpose. State the purpose if specific (**English for Legal Purposes**, **German for Academic Purposes**, etc.).

3 The description of test background may contain the reasons for developing the test, a description of the suite of which this test is a part, or other such details.

4 Describe the size and demographic profile of the candidature.

5 Describe the other components of the test (e.g. the speaking component, the reading component).

6 In the case that the number of tasks depends on which options are chosen, specify in the introductory text (point 5).

8 Skills, in addition to writing, which are involved in the completion of this task (regardless of whether they are explicitly recognised at the rating stage). Choose from: **none**, **reading**, **speaking**, **listening, a combination**.

9 The method by which the candidate's response is recorded. Choose from **handwritten**, **word processed**, **either**.

10 *CEFR*, Ch. 3.

11 The description may include information such as the number of subsections, task types in each subsection, time allowed for each subsection.

12 You may wish to include a short description of the task here. The description could include the aims of the task, what candidates have been asked to do and would constitute a full completion of the task.

13 Describe how marks are distributed in this section of the task and what candidates would need to include to achieve full marks on this task.

14 Explain how the task is rated (e.g. **clerically**, **machine marked**), what instruments are used and what aspects are considered when deciding the grade.

15 Describe the measures taken to ensure Writing tasks are set at the appropriate level. This description may include the process of question paper production and trialling.

16 Insert the sample task, including rubric and prompt/input.

18 Choose *CEFR* level: **A1**, **A2**, **B1**, **B2**, **C1**, **C2**.

19 If not specified, expected time.

20 The extent to which the rubric, prompt or input determines the nature and content of the response. Choose from: **controlled**, **semi-controlled** or **open-ended**.

21 Whether the content of the response is specified in the rubric. Choose from: **specified** or **not specified**.

22 Choose from: **letter (business)**, **letter (personal)**, **review**, **academic essay**, **composition**, **report**, **story**, **proposal**, **article**, **form**, other (specify).

23 The functions which might be expected in the response. Choose from: **describing (events)**, **describing (processes)**, **narrating**, **commentating**, **expositing**, **explaining**, **demonstrating**, **instructing**, **arguing**, **persuading**, **reporting events**, **giving opinions**, **making complaints**, **suggesting**, **comparing and contrasting**, **exemplifying**, **evaluating**, **expressing possibility/probability**, **summarising**, other (specify). *CEFR*, p125 – 130.

24 The imagined audience for the input. Choose from: **friend/acquaintance**, **teacher**, **employer**, **employee**, **committee**, **board**, **business**, **students**, **general public** (e.g. with a newspaper article), other (specify).

25 Choose from: **oral**, **written** or **visual**, or **a combination**.

26 The topic or theme. Choose from: **personal identification**, **house and home/environment**, **daily life**, **free time/entertainment**, **travel**, **relations with other people**, **health and body care**, **education**, **shopping**, **food and drink**, **services**, **places**, **language**, **weather**, other (specify). *CEFR*, p 51 – 53.

27 The language skills the candidate needs to understand the rubric and prompt/input. Choose from: **reading**, **listening**, or **a combination**.

29 The functions which might be expected in the response. Choose from: **describing (events)**, **describing (processes)**, **narrating**, **commentating**, **expositing**, **explaining**, **demonstrating**, **instructing**, **arguing**, **persuading**, **reporting events**, **giving opinions**, **making complaints**, **suggesting**, **comparing and contrasting**, **exemplifying**, **evaluating**, **expressing possibility/probability**, **summarising**, other (specify). *CEFR*, p125 – 130.

30 The expected purpose(s) of the response. Choose from: **referential** (to give 'objective' facts about the world), **emotive** (to describe the emotional state of the writer), **conative** (to persuade the reader(s)), **phatic** (to establish or maintain social contact with the reader(s)), **metalingual** (to clarify or verify understanding), **poetic** (writing for aesthetic purposes).

31 The register the candidate is expected to adopt in their response. Choose from: **informal**, **unmarked to informal**, **unmarked**, **unmarked to formal**, **formal**. *CEFR*, p 118 – 122.

32 The domain to which the expected response is imagined to belong. Choose from: **personal**, **public**, **occupational**, **educational/academic**. CEFR, p 45 – 46.

33 Choose *CEFR* level: **A1**, **A2**, **B1**, **B2**, **C1**, **C2**. *CEFR*, p 112 – 116.

34 Choose *CEFR* level: **A1**, **A2**, **B1**, **B2**, **C1**, **C2**. *CEFR*, p 110 – 112.

35 Choose *CEFR* level: **A1**, **A2**, **B1**, **B2**, **C1**, **C2**. *CEFR*, p 123 – 125.

36 The extent to which the task reflects a real-life activity a candidate could perform.
Choose from **low**, **medium**, or **high**.

37 The extent to which interaction patterns are likely to mirror those in an equivalent, real-life task.
Choose from **low**, **medium**, or **high**.

38 The difficulty in performing the task from a non-linguistic point-of-view.
Choose from: **reproduction of known ideas**, **knowledge transformation**.

39 The kind of extra-linguistic knowledge required to complete the task. Choose from: **personal/everyday life knowledge areas**, **general/non-specialised knowledge areas**, **specialised knowledge areas** (scientific, study-related, etc.), **a wide range of knowledge areas**.

40 Describe the rating criteria made available to the candidate, either before or during the test. If the criteria are not available together with the paper, state where they can be viewed.

41 Choose from: **impressionistic/holistic**, **descriptive scale**, **analytical scale.**

42 State the criteria used in marking. Choose from: **grammatical range, grammatical accuracy**, **lexical range lexical accuracy**, **cohesion and coherence**, **content/task fulfilment**, **development of ideas**, **orthography**, other (specify).

43 If clerically marked, the number or raters will be **1** or more. However, responses may only be second- or third-marked in some cases and by fellow raters, or by more senior raters. If so, insert '+ more in selected cases' after the base number of raters.

44 Quantitative feedback routinely given (for the writing component). Choose from: **raw score**, **percentage score**, **ranking in candidature**, **CEFR level**, **exam-specific grade**, **pass/fail status**, other (specify).

45 Qualitative feedback routinely given (for the writing component). Choose from: **comments for each of the rating criteria**, **holistic comments**, other (specify).

46 Insert a sample response to the task.

47 An explanation or justification of the grade awarded to the sample response.

48 The grade (or score) awarded to this sample response.

APPENDIX 2: The extended writing related CEFR descriptors at the lower levels

Note: The new descriptors for the lower levels (A1 and A2) are highlighted in green.

	OVERALL WRITTEN PRODUCTION
C2	Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.
C1	Can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion. Can employ the structure and conventions of a variety of written genres, varying the tone, style and register according to addressee, text type and theme.
B2	Can write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources.
B1	Can write straightforward connected texts on a range of familiar subjects within his/her field of interest, by linking a series of shorter discrete elements into a linear sequence.
A2	Can write a series of simple phrases and sentences linked with simple connectors like 'and,' 'but' and 'because'.
A1	Can give information in writing about matters of personal relevance (e.g. likes and dislikes, family, pets) using simple words and basic expressions. Can write simple isolated phrases and sentences.
Pre-A1	Can give basic personal information in writing (e.g. name, address, nationality), perhaps with the use of a dictionary.

	CREATIVE WRITING
C2	Can write clear, smoothly flowing and engaging stories and descriptions of experience in a style appropriate to the genre adopted. Can exploit idiom and humour appropriately to enhance the impact of the text.
C1	Can write clear, detailed, well-structured and developed descriptions and imaginative texts in an assured, personal, natural style appropriate to the reader in mind. Can incorporate idiom and humour, though use of the latter is not always appropriate. Can write a detailed critical review of cultural events (e.g. plays, films, concerts) or literary works.
B2	Can write clear, detailed descriptions of real or imaginary events and experiences marking the relationship between ideas in clear connected text, and following established conventions of the genre concerned.
	Can write clear, detailed descriptions on a variety of subjects related to his/her field of interest. Can write a review of a film, book or play.
B1	Can clearly signal chronological sequence in narrative text. Can write a simple review of a film, book or TV programme using a limited range of language.
	Can write straightforward, detailed descriptions on a range of familiar subjects within his/her field of interest. Can write accounts of experiences, describing feelings and reactions in simple connected text. Can write a description of an event, a recent trip – real or imagined. Can narrate a story.
A2	Can write about everyday aspects of his/her environment e.g. people, places, a job or study experience in linked sentences. Can write very short, basic descriptions of events, past activities and personal experiences. Can tell a simple story (e.g. about events on a holiday or about life in the distant future).
	Can write a series of simple phrases and sentences about their family, living conditions, educational background, present or most recent job. Can write short, simple imaginary biographies and simple poems about people. Can write diary entries that describe activities (e.g. daily routine, outings, sports, hobbies), people and places, using basic, concrete vocabulary and simple phrases and sentences with simple connectives like 'and,' 'but' and 'because'. Can write an introduction to a story or continue a story, provided he/she can consult a dictionary and references (e.g. tables of verb tenses in a course book).
A1	Can write simple phrases and sentences about themselves and imaginary people, where they live and what they do. Can describe in very simple language what a room looks like. Can use simple words and phrases to describe certain everyday objects (for example the colour of a car, whether it is big or small).
Pre-A1	*No descriptors available*

	REPORTS AND ESSAYS
C2	Can produce clear, smoothly flowing, complex reports, articles or essays which present a case, or give critical appreciation of proposals or literary works. Can provide an appropriate and effective logical structure which helps the reader to find significant points. Can set out multiple perspectives on complex academic or professional topics, clearly distinguishing his/her own ideas and opinions from those in the sources.
C1	Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues. Can expand and support points of view at some length with subsidiary points, reasons and relevant examples. Can write a suitable introduction and conclusion to a longer report, article or dissertation on a complex academic or professional topic provided that the topic is within his/her field of interest and there are opportunities for redrafting and revision.
B2	Can write an essay or report that develops an argument systematically with appropriate highlighting of significant points and relevant supporting detail. Can write a detailed description of a complex process. Can evaluate different ideas or solutions to a problem.
	Can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. Can synthesise information and arguments from a number of sources.
B1	Can write short, simple essays on topics of interest. Can write a text on a topical subject of personal interest, using simple language to list advantages and disadvantages, give and justify his/her opinion. Can summarise, report and give his/her opinion about accumulated factual information on familiar routine and non-routine matters within his/her field with some confidence.
	Can write very brief reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions. Can present a topic in a short report or poster, using photographs and short blocks of text. Can write a short, simple description of a familiar device or product.
A2	Can write simple texts on familiar subjects of interest, linking sentences with connectors like 'and,' 'because,' or 'then'. Can give his/her impressions and opinions in writing about topics of personal interest (e.g. lifestyles and culture, stories), using basic everyday vocabulary and expressions.
A1	*No descriptors available*
Pre-A1	*No descriptors available*

	OVERALL WRITTEN INTERACTION
C2	Can express him/herself in an appropriate tone and style in virtually any type of formal and informal written interaction.
C1	Can express him/herself with clarity and precision, relating to the addressee flexibly and effectively.
B2	Can express news and views effectively in writing, and relate to those of others.
B1	Can convey information and ideas on abstract as well as concrete topics, check information and ask about or explain problems with reasonable precision.
	Can write personal letters and notes asking for or conveying simple information of immediate relevance, getting across the point he/she feels to be important.
A2	Can write short, simple formulaic notes relating to matters in areas of immediate need.
A1	Can ask for or pass on personal details in written form.
Pre-A1	Can write short phrases to give basic information (e.g. name, address, family) on a form or in a note, with the use of a dictionary.

	CORRESPONDENCE
C2	Can write virtually any type of correspondence necessary in the course of his/her professional life in an appropriate tone and style.
C1	Can express him/herself with clarity and precision in personal correspondence, using language flexibly and effectively, including emotional, allusive and joking usage.
	Can, with good expression and accuracy, write formal correspondence such as letters of clarification, application, recommendation, reference, complaint, sympathy and condolence.
B2	Can maintain a relationship through personal correspondence using the language fluently and effectively to give detailed descriptions of experiences, pose sympathetic questions and follow up issues of mutual interest. Can in most cases understand idiomatic expressions and colloquialisms in correspondence and other written communications and use the most common ones him/herself as appropriate to the situation. Can write formal correspondence such as letters of enquiry, request, application and complaint with appropriate register, structure and conventions. Can write a forceful but polite letter of complaint, including supporting details and a statement of the desired outcome.
	Can write letters conveying degrees of emotion and highlighting the personal significance of events and experiences and commenting on the correspondent's news and views. Can use formality and conventions appropriate to the context when writing personal and professional letters and emails. Can write formal emails/letters of invitation, thanks or apology with appropriate register and conventions. Can write non-routine professional letters, using appropriate structure and conventions, provided these are restricted to matters of fact. Can obtain, by letter or email, information required for a particular purpose, collate it and forward it by mail to other people.
B1	Can write personal letters giving news and expressing thoughts about abstract or cultural topics such as music, films. Can write letters expressing different opinions and giving detailed accounts of personal feelings and experiences. Can reply to an advertisement in writing and ask for further information on items which interest him/her. Can write basic formal emails/letters, for example to make a complaint and request action.
	Can write personal letters describing experiences, feelings and events in some detail. Can write basic emails/letters of a factual nature, for example to request information or to ask for and give confirmation. Can write a basic letter of application with limited supporting details.
A2	Can exchange information by text message, email or in short letters, responding to questions the other person had (e.g. about a new product or activity).
	Can convey personal information of a routine nature, for example in a short email or letter introducing him/herself. Can write very simple personal letters expressing thanks and apology. Can write short, simple notes, emails and text messages (e.g. to send or reply to an invitation, to confirm or change an arrangement). Can write a short text in a greetings card (e.g. for someone's birthday or to wish them a Happy New Year).
A1	Can write messages and online postings as a series of very short sentences about hobbies and likes/dislikes, using simple words and formulaic expressions, with reference to a dictionary. Can write a short, simple postcard. Can write a short, very simple message (e.g. a text message) to friends to give them a piece of information or to ask them a question.
Pre-A1	Can write short phrases and sentences giving basic personal information with reference to a dictionary.

	NOTES, MESSAGES AND FORMS
C2	*No descriptors available; see B2*
C1	*No descriptors available; see B2*
B2	Can take or leave complex personal or professional messages, provided he/she can ask clarification or elaboration if necessary.
B1	Can take routine messages that are likely to occur in a personal, professional or academic context. Can take messages communicating enquiries, explaining problems.
	Can write notes conveying simple information of immediate relevance to friends, service people, teachers and others who feature in his/her everyday life, getting across comprehensibly the points he/she feels are important. Can take messages over the phone containing several points, provided that the caller dictates these clearly and sympathetically.
A2	Can take a short, simple message provided he/she can ask for repetition and reformulation.
	Can write short, simple notes and messages relating to matters in areas of immediate need. Can prepare simple notes on events and related dates and times that he/she needs to remember (e.g. arrangements made, a work schedule, a study timetable). Can fill in personal and other details on most everyday forms, e.g. to request a visa or visa waiver, to open a bank account, to send a letter recorded delivery, etc.
A1	Can write numbers and dates, own name, nationality, address, age, date of birth or arrival in the country etc. such as on a hotel registration form. Can leave a simple message giving information on e.g. where he/she has gone, what time he/she will be back (e.g. 'Shopping: back at 5 p.m.').
Pre-A1	Can fill in very simple registration forms with basic personal details: name, address, nationality, marital status.

	NOTE-TAKING (LECTURES, SEMINARS, MEETINGS ETC.)
C2	Can, whilst continuing to participate in a meeting or seminar, create reliable notes (or minutes) for people who are not present, even when the subject matter is complex and/or unfamiliar. Is aware of the implications and allusions of what is said and can make notes on them as well as on the actual words used by the speaker. Can make notes selectively, paraphrasing and abbreviating successfully to capture abstract concepts and relationships between ideas.
C1	Can take detailed notes during a lecture on topics in his/her field of interest, recording the information so accurately and so close to the original that the notes could also be used by other people. Can make decisions about what to note down and what to omit as the lecture or seminar proceeds, even on unfamiliar matters. Can select relevant, detailed information and arguments on complex, abstract topics from multiple spoken sources (e.g. lectures, podcasts, formal discussions and debates, interviews etc.), provided that standard language is delivered at normal speed in one of the range of accents familiar to the listener.
B2	Can understand a clearly structured lecture on a familiar subject, and can take notes on points which strike him/her as important, even though he/she tends to concentrate on the words themselves and therefore to miss some information Can make accurate notes in meetings and seminars on most matters likely to arise within his/her field of interest.
B1	Can take notes during a lecture, which are precise enough for his/her own use at a later date, provided the topic is within his/her field of interest and the talk is clear and well structured.
	Can take notes as a list of key points during a straightforward lecture, provided the topic is familiar, and the talk is both formulated in simple language and delivered in clearly articulated standard speech. Can note down routine instructions in a meeting on a familiar subject, provided they are formulated in simple language and he/she is given sufficient time to do so.
A2	Can make simple notes at a presentation/demonstration where the subject matter is familiar and predictable and the presenter allows for clarification and note-taking.
A1	*No descriptors available*
Pre-A1	*No descriptors available*

	GENERAL LINGUISTIC RANGE
C2	Can exploit a comprehensive and reliable mastery of a very wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. No signs of having to restrict what he/she wants to say.
C1	Can use a broad range of complex grammatical structures appropriately and with considerable flexibility. Can select an appropriate formulation from a broad range of language to express him/herself clearly, without having to restrict what he/she wants to say.
B2	Can express him/herself clearly and without much sign of having to restrict what he/she wants to say.
	Has a sufficient range of language to be able to give clear descriptions, express viewpoints and develop arguments without much conspicuous searching for words, using some complex sentence forms to do so.
B1	Has a sufficient range of language to describe unpredictable situations, explain the main points in an idea or problem with reasonable precision and express thoughts on abstract or cultural topics, such as music and films.
	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events, but lexical limitations cause repetition and even difficulty with formulation at times.
A2	Has a repertoire of basic language, which enables him/her to deal with everyday situations with predictable content, though he/she will generally have to compromise the message and search for words.
	Can produce brief everyday expressions in order to satisfy simple needs of a concrete type: personal details, daily routines, wants and needs, requests for information. Can use basic sentence patterns and communicate with memorised phrases, groups of a few words and formulae about themselves and other people, what they do, places, possessions etc. Has a limited repertoire of short memorised phrases covering predictable survival situations; frequent breakdowns and misunderstandings occur in non-routine situations.
A1	Has a very basic range of simple expressions about personal details and needs of a concrete type. Can use some basic structures in one-clause sentences with some omission or reduction of elements.
Pre-A1	Can use isolated words and basic expressions in order to give simple information about him/herself.

	GRAMMATICAL ACCURACY
C2	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).
C1	Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.
B2	Good grammatical control. Occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.
	Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding. Has a good command of simple language structures and some complex grammatical forms, although he/she tends to use complex structures rigidly with some inaccuracy.
B1	Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.
	Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.
A2	Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.
A1	Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.
Pre-A1	Can employ very simple principles of word order in short statements.

	ORTHOGRAPHIC CONTROL
C2	Writing is orthographically free of error.
C1	Layout, paragraphing and punctuation are consistent and helpful. Spelling is accurate, apart from occasional slips of the pen.
B2	Can produce clearly intelligible continuous writing, which follows standard layout and paragraphing conventions. Spelling and punctuation are reasonably accurate but may show signs of mother tongue influence.
B1	Can produce continuous writing which is generally intelligible throughout. Spelling, punctuation and layout are accurate enough to be followed most of the time.
A2	Can copy short sentences on everyday subjects – e.g. directions how to get somewhere. Can write with reasonable phonetic accuracy (but not necessarily fully standard spelling) short words that are in his/her oral vocabulary.
A1	Can copy familiar words and short phrases, e.g. simple signs or instructions, names of everyday objects, names of shops and set phrases used regularly. Can spell his/her address, nationality and other personal details. Can use basic punctuation (e.g. full stops, question marks).
Pre-A1	*No descriptors available*

	THEMATIC DEVELOPMENT
C2	Can use the conventions of the type of text concerned with sufficient flexibility to communicate complex ideas in an effective way, holding the target reader's attention with ease and fulfilling all communicative purposes.
C1	Can use the conventions of the type of text concerned to hold the target reader's attention and communicate complex ideas. Can give elaborate descriptions and narratives, integrating sub themes, developing particular points and rounding off with an appropriate conclusion. Can write a suitable introduction and conclusion to a long, complex text. Can expand and support main points at some length with subsidiary points, reasons and relevant examples.
B2	Can develop an argument systematically with appropriate highlighting of significant points, and relevant supporting detail. Can present and respond to complex lines of argument convincingly.
	Can follow the conventional structure of the communicative task concerned, when communicating his/her ideas. Can develop a clear description or narrative, expanding and supporting his/her main points with relevant supporting detail and examples. Can develop a clear argument, expanding and supporting his/her points of view at some length with subsidiary points and relevant examples. Can evaluate the advantages and disadvantages of various options. Can clearly signal the difference between fact and opinion.
B1	Can clearly signal chronological sequence in narrative text. Can develop an argument well enough to be followed without difficulty most of the time. Can compare and contrast alternatives.
	Shows awareness of the conventional structure of the text type concerned, when communicating his/her ideas. Can reasonably fluently relate a straightforward narrative or description as a linear sequence of points. Can briefly give reasons and explanations for opinions.
A2	Can tell a story or describe something in a simple list of points. Can give an example of something in a very simple text using 'like' or 'for example.'
A1	*No descriptors available*
Pre-A1	*No descriptors available*

	ONLINE CONVERSATION AND DISCUSSION
C2	Can express him/herself with clarity and precision in real-time online discussion, adjusting language flexibly and sensitively to context, including emotional, allusive and joking usage. Can anticipate and deal effectively with possible misunderstandings (including cultural ones), communication issues and emotional reactions occurring in an online discussion. Can easily and quickly adapt his/her register and style to suit different online environments, communication purposes and speech acts.
C1	Can engage in real-time online exchanges with several participants, understanding the communicative intentions and cultural implications of the various contributions. Can participate effectively in live, online professional or academic discussion, asking for and giving further clarification of complex, abstract issues as necessary. Can adapt his/her register according to the context of online interaction, moving from one register to the other within the same exchange if necessary. Can evaluate, re-state and challenge arguments in professional or academic live online chat and discussion.
B2	Can engage in online exchanges, linking his/her contributions to previous ones in the thread, understanding cultural implications and reacting appropriately.
	Can participate actively in an online discussion, stating and responding to opinions on topics of interest at some length, provided contributors avoid unusual or complex language and allow time for responses. Can engage in online exchanges between several participants, effectively linking his/her contributions to previous ones in the thread, provided a moderator helps manage the discussion. Can recognise misunderstandings and disagreements that arise in an online interaction and can deal with them, provided that the interlocutor(s) are willing to cooperate.
B1	Can engage in real-time online exchanges with more than one participant, recognising the communicative intentions of each contributor, but may not understand details or implications without further explanation. Can post online accounts of social events, experiences and activities referring to embedded links and media and sharing personal feelings.
	Can post a comprehensible contribution in an online discussion on a familiar topic of interest, provided that he/she can prepare the text beforehand and use online tools to fill gaps in language and check accuracy. Can make personal online postings about experiences, feelings and events and respond individually to the comments of others in some detail, though lexical limitations sometimes cause repetition and inappropriate formulation.
A2	Can introduce him/herself and manage simple exchanges online, asking and answering questions and exchanging ideas on predictable everyday topics, provided enough time is allowed to formulate responses, and that he/she interacts with one interlocutor at a time. Can make short descriptive online postings about everyday matters, social activities and feelings, with simple key details. Can comment on other people's online postings, provided that they are written in simple language, reacting to embedded media by expressing feelings of surprise, interest and indifference in a simple way.
	Can engage in basic social communication online (e.g. writing a simple message on a virtual card for a special occasion, sharing news and making/confirming arrangements to meet). Can make brief positive or negative comments online about embedded links and media using a repertoire of basic language, though he/she will generally have to refer to an online translation tool and other resources.
A1	Can write very simple messages and personal online postings as a series of very short sentences about hobbies, likes/dislikes, etc., relying on the aid of a translation tool. Can use formulaic expressions and combinations of simple words to post short positive and negative reactions to simple online postings and their embedded links and media, and can respond to further comments with standard expressions of thanks and apology.
Pre-A1	Can post simple online greetings, using basic formulaic expressions and emoticons. Can post online short simple statements about him/herself (e.g. relationship status, nationality, occupation), provided he/she can select them from a menu and/or refer to an online translation tool.

	GOAL-ORIENTED ONLINE TRANSACTIONS AND COLLABORATION
C2	Can resolve misunderstandings and deal effectively with frictions that arise during the collaborative process. Can provide guidance and add precision to the work of a group at the redrafting and editing stages of collaborative work.
C1	Can coordinate a group who are working on a project online, formulating and revising detailed instructions, evaluating proposals from team members and providing clarifications in order to accomplish the shared tasks. Can deal with complex online transactions in a service role (e.g. applications with complicated requirements), adjusting language flexibly to manage the discussion and negotiation. Can participate in complex projects requiring collaborative writing and redrafting as well as other forms of online collaboration, following and relaying instructions with precision in order to reach the goal. Can deal effectively with communication problems and cultural issues that arise in an online collaborative or transactional exchange by reformulating, clarifying and exemplifying through media (visual, audio, graphic).
B2	Can take a lead role in online collaborative work within his/her area(s) of expertise, keeping the group on task by reminding them of roles, responsibilities and deadlines in order to achieve established goals. Can engage in online collaborative or transactional exchanges within his/her area(s) of expertise that require negotiation of conditions and explanation of complicated details and special requirements. Can deal with misunderstandings and unexpected problems that arise in online collaborative or transactional exchanges by responding politely and appropriately in order to help resolve the issue.
	Can collaborate online with a group that is working on a project, justifying proposals, seeking clarification and playing a supportive role in order to accomplish shared tasks.
B1	Can engage in online transactions that require an extended exchange of information, provided the interlocutor(s) avoid complex language and are willing to repeat and reformulate when necessary. Can interact online with a group that is working on a project, following straightforward instructions, seeking clarification and helping to accomplish the shared tasks.
	Can engage in online collaborative or transactional exchanges that require simple clarification or explanation of relevant details, such as registering for a course, tour, event or applying for membership. Can interact online with a partner or small group working on a project, provided there are visual aids such as images, statistics and graphs to clarify more complex concepts. Can respond to instructions and ask questions or request clarifications in order to accomplish a shared task online.
A2	Can use formulaic language to respond to routine problems arising in online transactions (e.g. concerning availability of models and special offers, delivery dates, addresses, etc.). Can interact online with a supportive partner in a simple collaborative task, responding to basic instructions and seeking clarification, provided there are some visual aids such as images, statistics, or graphs to clarify the concepts involved.
	Can make simple online transactions (such as ordering goods or enrolling on a course) by filling in an online form or questionnaire, providing personal details and confirming acceptance of terms and conditions, declining extra services, etc. Can ask basic questions about the availability of a product or feature. Can respond to simple instructions and ask simple questions in order to accomplish a shared task online with the help of a supportive interlocutor.
A1	Can complete a very simple online purchase or application, providing basic personal information (such as name, email address or telephone number).
Pre-A1	Can make selections (e.g. choosing a product, size, colour) in a simple online purchase or application form, provided there is visual support.

British Council Assessment Research Awards and Grants

If you're involved or work in research into assessment, then the British Council Assessment Research Awards and Grants might interest you.

These awards recognise achievement and innovation within the field of language assessment and form part of the British Council's extensive support of research activities across the world.

ASSESSING WRITING AT LOWER LEVELS: RESEARCH FINDINGS, TASK DEVELOPMENT LOCALLY AND INTERNATIONALLY, AND THE OPPORTUNITIES PRESENTED BY THE EXTENDED CEFR DESCRIPTORS

AR-G/2018/4

Eva Konrad, Franz Holzknecht, Veronika Schwarz and Carol Spöttl

ARAGs RESEARCH REPORTS ONLINE

ISSN 2057-5203