

Testing Writing

Specifications for the E8-Standards Writing Tests

LTC Technical Report 4

Otmar Gassner
Claudia Mewald
Guenther Sigott



bifie

bundesinstitut für bildungsforschung,
innovation und entwicklung des bildungswesens

bm:uk Bundesministerium für
Unterricht, Kunst und Kultur

CONTENTS

<u>EMBEDDING THE E8 WRITING TEST IN A NATIONAL AND INTERNATIONAL CONTEXT</u>	4
THE PLACE OF WRITING IN AUSTRIAN LOWER SECONDARY SCHOOL	4
<u>VALIDITY ASPECTS WITH REGARD TO THE E8 WRITING TEST CONSTRUCT</u>	4
TEST TAKER CHARACTERISTICS	4
COGNITIVE VALIDITY	7
WRITING THEORY IN BRIEF	8
COGNITIVE PROCESSING IN E8 STANDARDS	8
CONTEXT VALIDITY	10
SETTING: TASK	11
SETTING: ADMINISTRATION	12
LINGUISTIC DEMANDS: TASK INPUT AND OUTPUT	12
SCORING VALIDITY	13
CRITERIA AND RATING SCALE	14
RATER CHARACTERISTICS	15
RATING PROCESS	16
RATING CONDITIONS	16
RATER TRAINING	16
POST-EXAM ADJUSTMENTS	18
REPORTING RESULTS	19
CONSEQUENTIAL VALIDITY	19
<u>E8 WRITING TEST SPECIFICATIONS VERSION 02 (MAY 2008)</u>	20
1. PURPOSE OF THE TEST	20
2. DESCRIPTION OF TEST TAKERS	20
3. TEST LEVEL	20
4. TEST CONSTRUCT	20
5. STRUCTURE OF THE TEST	23
6. TIME ALLOCATION	23
7. ITEM FORMATS	23
8. RUBRICS	23
9. WRITING RATING SCALE	23
10. PROMPTS AND PERFORMANCE SAMPLES WITH JUSTIFICATIONS	29
<u>LITERATURE</u>	36
<u>APPENDIX</u>	38
PROMPT INTERPRETATION: LONG PROMPT	38
PROMPT INTERPRETATION: SHORT PROMPT	39

Embedding the E8 Writing Test in a national and international context

The place of writing in Austrian lower secondary school

There seems to be some agreement that speaking and listening are the skills most needed when trying to get around in a foreign language environment and that being able to read is next in priority. This leaves writing as the skill least necessary for survival. Nevertheless, writing is trained from year one on a regular basis. It starts off with model paragraphs that are personalized by the learners and leads on to open writing based on the content of the course book unit in progress. It goes without saying that lower ability learners are given more guidance with some of them hardly ever attempting an open-ended writing task.

In higher grades the importance attributed to writing increases. It seems to be a widespread belief among teachers of English that when writing skills are assessed, other dimensions of language competence like vocabulary and grammar knowledge are assessed automatically at the same time. Therefore, the writing grade goes a long way towards the overall English grade for that particular student (as speaking is hardly taken into account).

Whereas this belief might be responsible for the high regard teachers have for writing, the awareness of the complexity of assessment procedures for writing is still limited. There is no perceived need for shared standard levels, there is no agreement on how writing should be tested, marked and weighted in relation to the other skills (reading, listening, speaking), there are a great number of idiosyncratic marking schemes in place (even within one school), and there is no agreement on anything like pass marks or cut-off scores for grading.

In this situation there is room for constructive washback in the course of the introduction of E8 standards. It is hoped that the way the tests are constructed and assessed will impact on the way writing is taught and assessed in Austrian schools.

Validity aspects with regard to the E8 Writing Test construct

Shaw & Weir have designed a clear graphic to illustrate their “framework for conceptualising writing test performance” (see figure 1, p. 5). It takes all the relevant parameters into account and can serve as the blueprint for the description of the E8 Writing Tests. Within this framework the focus of the discussion will be on the following categories: test taker characteristics, cognitive validity, context validity, scoring validity, and consequential validity.

Test taker characteristics

It is obvious that test taker characteristics have an influence on the way a task is processed and a text is written. Three categories have been identified: physical/physiological, psychological and experiential characteristics (Shaw & Weir 2007, 5).

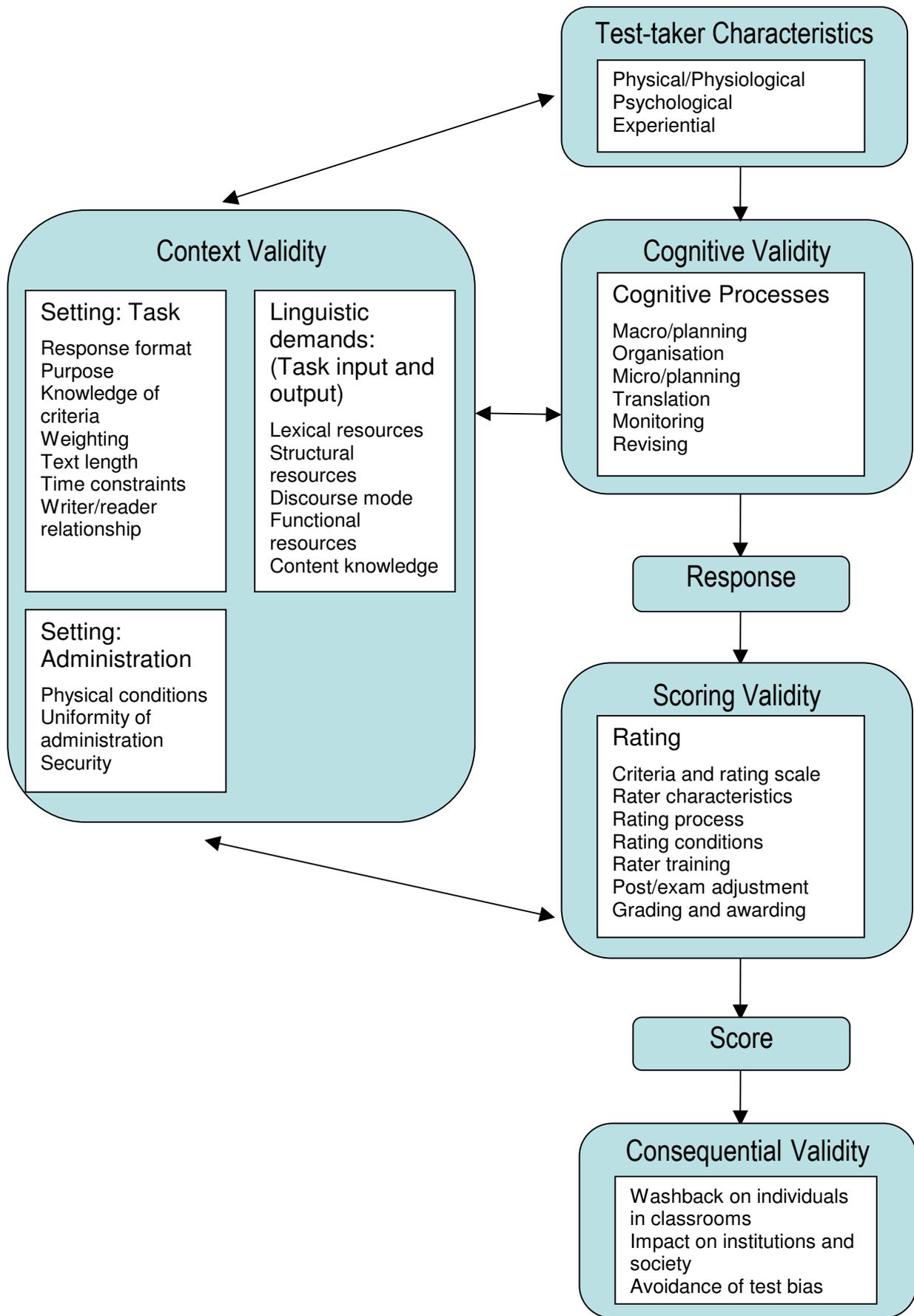


Figure 1: Adapted from Shaw & Weir 2007, 4

As regards the first category, any provisions made for schooling can be considered sufficient for the E8 test situation as all test takers are pupils in the last form of lower secondary schools in Austria. To put it simply, any pupil who is fit to attend English classes at an Austrian secondary school and to be assessed is fit to take the E8 test.

Psychological factors, however, are almost impossible to control. Most critical is motivation as E8 Standards is a low-stakes exam that has no influence whatsoever on the individual test takers' marks or on their school career. We can expect students from the lower segment of achievement to be more affected by lack of motivation. For this reason, test results might not fully represent the actual language competence of these students, but they might appear to be at a significantly lower level because a fair number of this group of test takers may choose not to show what they *can do* in English. As long as the test is only administered to a sample of the student population with no practical implications for the individual, it will be difficult to generate real interest and motivation in those that display a "could-not-care-less" attitude. Preferred learning styles and personality traits are other factors that are relevant, but cannot be catered for in the given test situation.

The third group of factors are experiential characteristics referring to familiarity with the test format. Whereas the test takers are all new to this particular type of testing, they should generally be familiar with the type of prompts used in the E8 Writing Test. As details from the test specifications below confirm (cf. pp. 21-22), all text types and domains used in prompt writing have been taken from the CEFR and are also based on the Austrian core curriculum (Lehrplan der Hauptschule 2007, 4-9) as described in *Bildungsstandards Englisch* (Gassner et al. 2005, 41). A brief survey of the course books with the largest circulation reveals that writing tasks are, as a rule, identical in topic and scope with the prompts in the E8 tests, but there is generally more scaffolding. In a great number of cases, writing is done on the basis of a given text which is slightly adapted by the learners so that it reflects their personal situation as in the example below.

STUDY and CHANGE

TEXT 1

Dear ...,

I've been in New York for a week now. I cannot tell you how great it is. We are staying in a hotel in Manhattan. There are hundreds of skyscrapers. When you look up at them, you feel like a mouse! Yesterday we went to see the Statue of Liberty. It is about a hundred metres tall and we went up to the top of it. From there we could see New York Harbour. Great! Today we are going to Central Park.

Love,

...

TEXT 2

I definitely would not like to live in New York. I think life in big cities is awful. There are lots of traffic jams and life is very dangerous. Some time ago, I saw a documentary about New York on TV. It showed that there are thousands of homeless people and street kids in New York. They have no place to live and they are hungry and depressed. There are also lots of burglaries and robberies in New York. No, New York would not be a place for me!

Figure 2: Adapted from Gerngroß et al. 1997, 78

Other writing tasks are more open and therefore often close to real E8 prompts, except for a number of pre-fabricated phrases and sentence starters that are also provided or example texts that fill in all necessary information on genre and a set of relevant language.

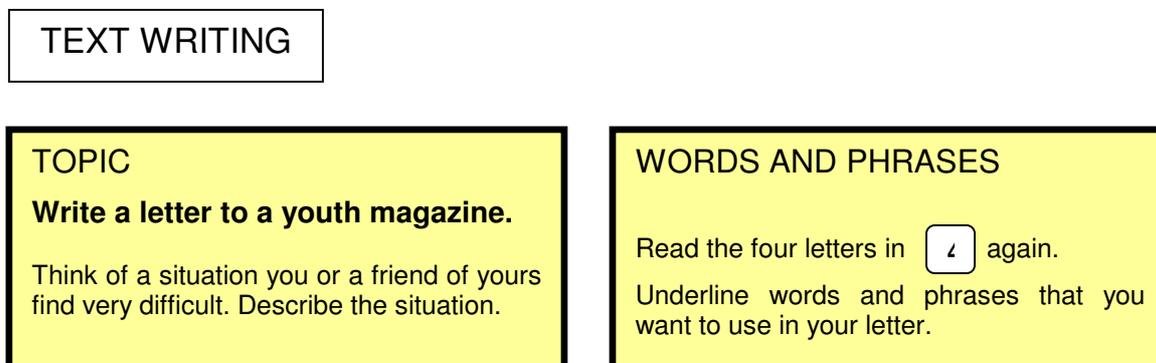


Figure 3: Adapted from Gerngroß et al. 1997, 87

Learners who have only done tasks of the type described above will find the E8 prompts challenging. Learners who have never faced open writing tasks in their learning history cannot be expected to perform well in the E8 Standards Tests or in international tests. We would consider it important washback if course book authors and, consequently, also teachers were to rethink the issues involved and also attempt unscaffolded writing tasks with ALL pupils. After four years of English at secondary school and some (very limited) writing at primary level amounting to more than 500 lessons, any student should be able to do a task like the one below successfully:

Your friend's birthday party was a few days ago. Write an **e-mail** to tell him/her that you liked the party.

- Tell him/her why you liked the party.
- Tell him/her what you liked best.
- Ask your friend when you are going to meet again.
- Suggest something for the next weekend.

Figure 4: LTC Item archive

Cognitive validity

“The cognitive validity of a Writing task is a measure of how closely it represents the cognitive processing involved in writing contexts beyond the test itself, i.e. in performing the task in real life” (Shaw & Weir 2007, 34). Whereas it is notoriously difficult to describe the cognitive processes as they are not directly accessible, it seems important to describe a general writing model that accounts for writing in a real-life context as well as in an exam situation. However, one difference should be noted at the outset, namely that there is no time-constraint in most real-life situations whereas in the E8 testing situation time, topic, genre, and length of output are pre-

determined. This might impose limitations on the planning phase as well as on the writing and revision phases.

Writing theory in brief

In the given context, only sketchy references shall be made to various sources that present and discuss the writing process and models of L1 and L2 writing in detail. According to Grabe and Kaplan (1996, 230-232), the planning phase, which they call “goal setting”, involves these five factors:

- an assessment of the context
- a preliminary representation of the writing product
- an evaluation of possible problems in task execution
- an initial consideration of the genre required
- an organisational plan

Shaw and Weir (2007, 37) make a point of emphasizing the advantages of a more psycholinguistically oriented model of writing over the Grabe and Kaplan model and refer to Field (2004) and Kellogg (1994, 1996). Interested readers may wish to consult the detailed discussion there. The Field model (Field 2004, 329-331) involves

- macro-planning
- organisation
- micro-planning
- translation
- execution
- monitoring
- editing and revising

Scardamalia and Bereiter (1987) have described two different strategies used by skilled and less skilled writers in the planning phase: *knowledge telling* and *knowledge transformation*.

In knowledge telling, novice writers plan very little, and focus on generating content from within remembered linguistic resources in line with the task, topic, or genre. Knowledge transforming by the skilled writer entails a heightened awareness of problems as and when they arise – whether in the area of ideas, planning and organisation (content), or in those of goals and readership (rhetoric) [...] (Shaw & Weir 2007, 43).

Whereas this holds true for all writing, L2 writing poses additional cognitive demands on the writers as Field (2005) argues. Attention directed towards linguistic aspects like lexical retrieval, spelling, and sentence structures can impede the fluency of writing and the capacity to organise and structure the text as a whole. Some ideas might have to be abandoned in the execution phase on the grounds of language constraints and limitations.

Cognitive processing in E8 Standards

In the E8 context we suggest using a modified Grabe/Kaplan-Field model to illustrate the writing process, which will clearly be based on *knowledge telling* and thus has a very brief planning phase mainly consisting of selecting and considering relevant content points.

This model includes the following phases:

- assessment of the context (who writes about what to whom and why?)
- characteristic features of the genre required
- preliminary representation of the writing product
- selection of content points
- evaluation of possible problems in task execution
- micro-planning at paragraph and sentence level
- translation
- monitoring
- revising

In the E8 test situation, the planning phase is clearly not elaborate or extensive. After an assessment of the context, which includes identifying the topic, the situation of the writer, the text purpose and the addressee, most test takers will move straight to the consideration of the genre required and develop a “preliminary representation of the writing product”. Then the bullet points will pre-structure the content elements to be included and will initiate the process of selection if options are given. An organisational plan is not necessary as the tasks are fairly short and heavily scaffolded with little opportunity for deviation. Especially with the short task (40-60 words) planning efforts will be reduced to the bare minimum.

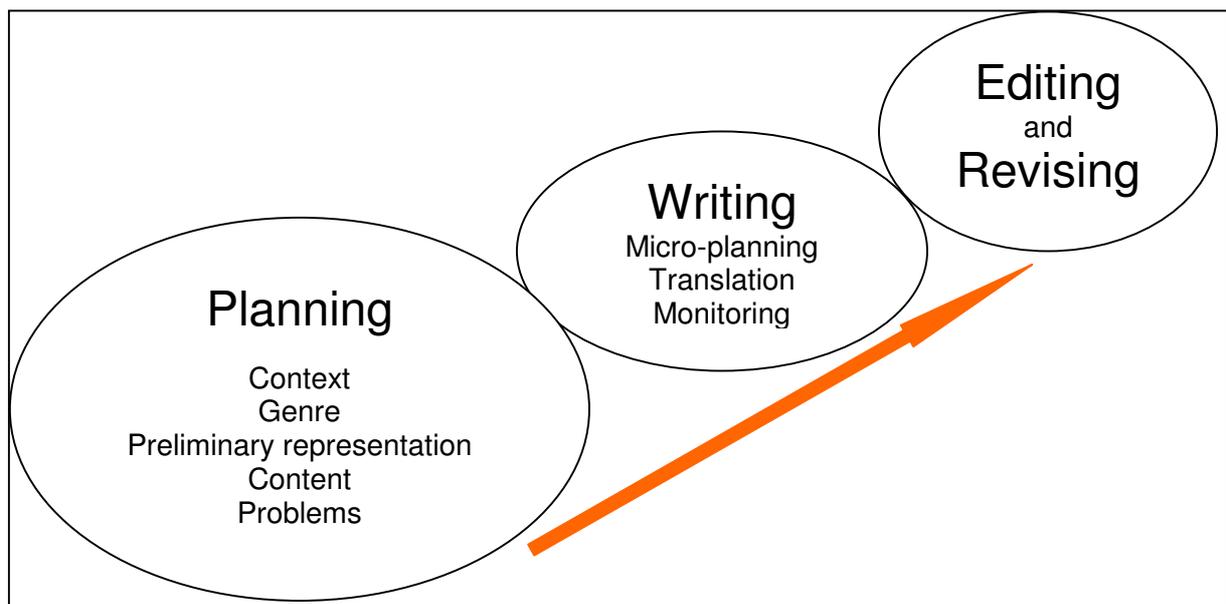


Figure 5

The writing is more closely scaffolded than in real life as a number of content points are given in the prompt, sometimes with the option to choose among them, and to include a clearly specified lower number of them in the text. On the one hand, this makes the writing process somewhat easier than in real life, on the other hand, it is a necessity if we want to ensure inter-rater reliability for the dimension of task achievement. In addition to this, providing a content schema for candidates at this level is necessary because the cognitive load for simultaneous activities on a macro and micro level would be too great and the task too demanding.

It has become clear from the present discussion that macro-planning and organization play no role in the given writing context and that the product delivered will be firmly set in the area of knowledge telling.

If we continue with the description of the various phases of the writing process, this might be the point where possible problems in task execution will be identified before the actual writing begins. The problems will be content-related and have to do with knowledge of the world and the decision which content points to select; they might also be connected to the attempt to recall the requirements of the genre in question and, furthermore, with the language necessary to express the ideas.

This stage of identifying language resources and their limitations is only a fraction away from actually putting pen to paper and certainly is a central aspect of micro-planning focusing

on the part of the text that is about to be produced. Here, the planning takes place on at least two levels: the goal of the paragraph, itself aligned with the overall goal of the writing activity; within the paragraph, the immediate need to structure an upcoming sentence in terms of information (Shaw & Weir 2007, 39).

Micro-planning merges with the translation phase where previously abstract ideas only accessible to the writer him/herself are translated into the public space defined by a shared language. In contrast to Shaw & Weir and Field, we see micro-planning and translation as two stages that are interlinked as the writer might oscillate between the one and the other at sentence level or at paragraph level (Shaw & Weir 2007, 39-40).

It is in the translation stage that language decisions have to be made and planning decisions have to be implemented. The actual production of text will be taking place under the constraints of content schemata, genre restrictions and the limitations of linguistic resources at hand in L2. What has been called “avoidance behaviour” (e.g. avoiding lexis or structures that seem unsafe) and “achievement behaviour” (e.g. using simpler structures, paraphrasing) by Field (2004, 66-67) needs to be taken care of in the marking scheme, as does the ability to produce coherent and cohesive texts.

The next step is monitoring although this is not necessarily sequential and might be oscillating with phases of translation. “At a basic level monitoring involves checking the mechanical accuracy of spelling, punctuation and syntax” (Shaw & Weir 2007, 41). At E8 level this is what can be expected, if not in the lowest segment of test takers. In addition, better writers will also check back on content and genre requirements. These monitoring activities will lead to editing and revising if some parts of the text have been found unsatisfactory. This might involve adding, deleting or modifying a content point, adding cohesive devices, replacing poor words and phrases with better ones, or simply correcting mistakes in spelling and structure.

In the E8 context, writing is certainly based on *the knowledge-telling model* (Scardamalia & Bereiter 1987); the Hyland summary of the model epitomizes E8 writing performances:

A knowledge-telling model addresses the fact that novice writers plan less often than experts, revise less often and less extensively, and are primarily concerned with generating content from their internal resources. Their main goal is simply to tell what they can remember based on the assignment, the topic, or the genre (Hyland 2002, 28).

Context validity

Tests should be as close as possible to authentic real-life situations. Writing is an activity that is normally performed by individuals at a time set aside for it. Writers

have a purpose and an audience; they have the freedom to interrupt the writing process and resume it at a time of their choice, especially for editing and revising; and they can normally use dictionaries and other resources. In the given test setting, some constraints will be operative.

Shaw & Weir 2007 (64-142) discuss a number of aspects of context validity related to three areas:

- Setting: task
- Setting: administration
- Linguistic demands: Task input and output

These points will structure the discussion of context validity of the E8 Writing Tests.

Setting: task

The aspects to be discussed here are response format, purpose, knowledge of criteria, weighting, text length, time constraints, and writer-reader-relationship. In the E8 Writing Tests authenticity is one of the most prominent aims of prompt construction. In contrast to real-life writing there is no provision for the use of any resource materials such as dictionaries.

As the response format may well play a significant role in test performance (Alderson et al. 1995), the decision has been taken to include two formats in the E8 Writing Test. There is a short task (40-60 words) and a long task (120-150 words), which are assessed separately. Both are open-ended writing tasks. Good writers have a better chance to show their best in the long task, whereas lower achievers are expected to do better in the short task, which is limited in scope and closely guided. As the E8 test does not provide an overall pass mark for E8 language competence in general, but reports back to the test takers a profiled feedback on each skill, the results achieved in the four skills are not weighted.

The prompts contextualise the task by defining the writer-reader-relationship, stating purpose and genre, and giving content points to be included in the text. The short task contains 3-5 content points, the long one 5-8. The rubric tells the test taker exactly how many of the content points listed must be dealt with in the text. For both tasks the test takers have 30 minutes of writing time plus 5 minutes for editing and revising in all.

The writing tasks are targeted at pupils of Austrian schools in year 8 and normally aged fourteen. The tasks are designed to appeal to this age group and to elicit scripts that show what test takers can do within the framework defined in *Bildungsstandards Englisch* (Gassner et al. 2005). The domains and genres have been carefully selected from the CEFR and filtered further on the basis of the Austrian curriculum.

So far the tests have only been administered in “pilot schools”, i.e. schools that volunteered to function as schools where the E8 Standards Tests could be piloted. These schools were provided with all the information and additional training materials on E8 Standards and have been involved in the process since 2005. Therefore, it can be safely assumed that the test takers have been familiarised with the test formats and also with the rating criteria, which have been publicly available on the LTC homepage since 2005. When the test goes national, which is expected to happen in 2009, special provisions will have to be made so that the schools involved in the testing are informed in good time and the pupils are given detailed information about the test formats and the assessment criteria.

Setting: administration

The E8 Standard Tests are set once a year, usually in May. The writing test was first piloted on a sample of ca. 700 test takers in 2006. The schools taking part in the tests were “pilot schools”. In 2009 a baseline study will be carried out and in 2012 a random sample of Austrian school children in grade 8 will be tested.

The writing prompts have all been written by the prospective raters, moderated, edited, and screened by the testing group at the LTC Klagenfurt, pre-tested and stored in the item archive. The test booklets are designed at the LTC. The actual distribution of all test papers is handled centrally by the University of Education in Upper Austria.

The test papers are sent to the E8 Standards coordinator in each of the nine Austrian provinces, who is responsible for delivering the test papers to the schools selected for testing. Guidelines for school principals, test administrators, and teachers are in place (<http://www.uni-klu.ac.at/ltc/inhalt/266.htm>). On testing day an external test administrator, who has been trained for the job, comes to the selected school and administers the test in one of the fourth forms that sit the test. There is a test administrator for each class taking the test. The test administrator hands out the test booklets, reads out standardized instructions and clarifies time limits. When test takers have finished the writing test, they close their test booklets and leave the testing room. When time is up and all test takers have left the premises, the test booklets are collected and sent to the LTC at the University of Klagenfurt for the preparation of the rating plan and the rating session.

Linguistic demands: Task input and output

In the Austrian teacher community the communicative approach to language learning (Canale & Swain 1980 as an important precursor of the Bachman 1990 model of communicative language ability) is widely accepted, and it is also set down in writing in the national curriculum. As the learning tasks are modelled on real-life contexts, the learning environment aims to mirror real life as closely as possible. Exams set in the Austrian context need to share these premises and to reflect them in the tasks set.

Shaw & Weir (2007, 91), Alderson (2004, 13) and others complain that the CEFR remains vague and withholds any details when it comes to structures or vocabulary, using terms like “simple” in the descriptors. While this is true, reading the CEFR extensively rather than focusing only on the sections containing the scales proves helpful. In chapter 3, the development of the common reference levels is explained and it is made clear that they progress in a very coherent way from “the lowest level of generative language use” (CEFR 2001, 33) to social functions and “descriptors on getting out and about” (CEFR 2001, 34) based on *Waystage* (A2) and a simplified version of some transactional language from “‘The Threshold Level’ for adults living abroad” (CEFR 2001, 34). A2+ does not so much increase the range of topics, but focuses on “more active participation in conversation” and “significantly more ability to sustain monologues”. B1 reflects the Threshold level and involves “the ability to maintain interaction and get across what you want to, in a range of contexts” as well as “the ability to cope flexibly with problems in everyday life” (CEFR 2001, 34). B1+ makes increased demands on the quantities of information to be handled.

As this is the way the levels have been constructed (i.e. from *Waystage* to A2), it seems legitimate to move from A2 specifications back to *Waystage*. And there we have a vocabulary list and a list of structures considered characteristic of that level.

As UCLES have also oriented themselves on vocabulary lists from the Council of Europe Publications (Lexical Inventory in *Waystage*, 1980, 45-62; and in *Threshold*, 1979, 85-115), it can be considered a useful shortcut to pick up the vocabulary lists published on the web for KET (A2) and PET (B1), especially as these have been updated on the basis of frequency and usage data from language corpora. Generally, “[the] language specifications of KET are the same as those set out in *Waystage 1990*” (KET Handbook 2007, 1).

To resume the discussion of the vagueness of descriptors using words like “simple”, “basic” or “sufficient”, it may suffice to say that this vagueness needs to be contextualised. If the A2 descriptor on Grammatical Accuracy reads “Uses some simple structures correctly, but still systematically makes basic mistakes” (CEFR 2001, 112), we can expect learners to use the range of structures listed in the *Structural Inventory of Waystage* (63-83) or the *KET Handbook* (2-3) with severely restricted accuracy. In this sense, even vague terms like “simple” are reasonably well-defined so that raters know what to look for.

What has to be noted, however, is the basic orientation of the CEFR towards an adult learner and a dominance of tourist aspects of language learning. This is why the Austrian E8 Standards have also integrated the specifications set down in the Austrian curriculum and adapted the CEFR descriptors to the age group of the test population. This mainly reflects the selection of domains and transactional situations. It has no influence on the structures included, though it has some influence on the wordlist. Generally, the school books used in Austria take this into account. As the test is explicitly based on the Austrian curriculum, the linguistic demands of the test are fair for all test takers.

The writing prompts used always specify particular language functions to be performed, e.g. “invite..., apologize..., ask for..., give advice...”. A list of these functions has been made available to the teachers preparing the test takers so that these can be expected to be aware of them.

Several research papers have observed an interaction or even an interdependence of content knowledge on the one hand, and writing performance and test scores on the other (Read 1990, Papajohn 1999, Weir 2005). Provisions for this have been made by restricting topics to areas that can safely be assumed to be familiar to the test takers as they are set down in the Austrian core curriculum and must have been included in their English lessons. However, this still leaves the fact that some test takers might feel indisposed to deal with a particular topic for a number of reasons, the most common probably being lack of motivation and interest. The LTC team is considering offering a choice of topics in the future, which might help test takers to show their best although it gives rise to questions of comparability of performances.

Scoring validity

Scoring validity is concerned with all the aspects of the testing process that can impact on the reliability of test scores. [...] It is critical because if we cannot depend on the rating of exam scripts it matters little that the tasks we develop are potentially valid in terms of both cognitive and contextual parameters. Faulty criteria or scales, unsuitable raters or procedures, lack of training and standardisation, poor or variable conditions for rating, inadequate provision for post exam statistical adjustment, and unsystematic or ill-conceived procedures for grading and awarding can all lead to a reduction in scoring validity and to the risk of construct irrelevant variance (Shaw & Weir 2007, 144-145).

In this section we examine each of the relevant parameters in some detail: criteria and rating scale, rater characteristics, rating process, rating conditions, rater training, post-exam adjustments, and grading.

Criteria and rating scale

Before the actual construction of the rating scale information on existing scales was collected and the usefulness of the scales in the framework of E8 Testing was analysed: Jacobs et al. scoring profile 1981 (Weigle 2002, 116); TEEP attribute writing scales, Weir 1990 (Weigle 2002, 117); FCE Scoring rubric 1997 (Weigle 2002, 152); TOEFL writing scoring guide 2000 (Tankó 2005, 125); IELTS bands 2002 (Weigle 2002, 159); Analytic writing scale developed by the Hungarian School-Leaving English Examination Reform Project 2005 (Tankó 2005, 127).

Lumley reports findings from Weigle 1994, who used an analytic scale to have 30 compositions assessed by novice and expert raters. Weigle focused on novice raters, which is relevant to the E8 situation in Austria where a rating culture is only just evolving.

She found that rater reliability increased as a result of training, and that the improved agreement was the result of raters gaining better consensual understanding of the terms and levels represented in the scale. She found evidence that training helped clarification of the rating criteria (Lumley 2005, 44).

This supports the view of the testing team that in the given context an analytic scale would be preferable over a holistic scale. This view is also supported by Weigle 2002, who mentions several advantages of analytic over holistic scoring:

- It is more useful in rater training as inexperienced raters can more easily understand and apply the criteria in separate scales.
- It is particularly useful for second-language learners, who are more likely to show a marked or uneven profile.
- A scoring scheme in which multiple scores are given to each script tends to improve reliability (Weigle 2002,120).

Another reason for ruling out a holistic approach was the fact that rating procedures for scripts within the Austrian school system are not regulated, show great variety and are to a large extent holistic, even impressionistic. As assessment procedures for writing in place in Austrian schools cannot be taken as a basis for a disciplined approach towards rating scripts, breaking with this tradition seemed to best guarantee a fresh approach to assessment.

Taking the general background of Austrian traditions in assessing writing into account and inspired by the Hungarian scale (Tankó 2005, 127), the decision was taken to design an analytic scale measuring four dimensions: Task Achievement, Coherence and Cohesion, Grammar, and Vocabulary. Whereas three of these four dimensions have a strong recognition value for Austrian teachers, Coherence and Cohesion might appear unusual and reflects the high importance given to this dimension by the CEFR. These four dimensions promised to yield enough detail for a constructive feedback profile on individual test takers' performance as well as informative data for system monitoring. The construction of the scale was informed by the fact that the overall majority of performances could be expected to be around A2/B1. This meant that A1 and B1 descriptors needed to be included while anything above B2 could be neglected. We are aware of the fact that this kind of scale cannot measure B2 or C1 performances and we have settled for stating that performances above the upper end

of the descriptors in the E8 scale are called “above B1”. But, generally, the applicability of a particular descriptor does not automatically signal that a script is at that CEFR level. Firstly, bands consist of more than one descriptor, and secondly, linking written performances to the CEFR is a complex procedure that is beyond the scope of this paper and will be discussed in the LTC Technical Report 6.

The second consideration in scale construction was the cognitive load that raters can manage in the rating process. The decision to use four dimensions is also in agreement with the CEFR recommendation to reduce the number of possible categories to “a feasible number” as “more than 4 or 5 categories starts to cause cognitive overload”. (CEFR 2001, 193) We take it that this warning also applies to the number of bands and descriptors that raters can handle, so we have opted for four bands supplied with descriptors and three empty bands in between, making it a seven-band scale plus a zero band. The scale contains three columns with the left one giving the descriptors, the middle columns showing an extended version of the descriptors and the right columns giving the exact wording of the relevant CEFR descriptor. This is especially useful for training purposes when the deflated scale might not carry enough information for the rater. The scales themselves have been fine-tuned in the training process in an ongoing dialogue with the raters. It follows from this that the scales are what has been called assessor-oriented (Weigle 2002, 122; CEFR 2001, 38).

An important decision in the process of scale construction was the removal of the CEFR levels at the end of the CEFR descriptors. This was the logical step to take when some raters awarded band 7 to a script and argued that the script was a B2 performance. However, such an argument is inadmissible as the prompts used in the test are written on the basis of A2 or B1 descriptors and responses to these prompts simply cannot measure performances above B1 as one basic factor is the scope of a performance together with the given limitations of domains and genres. So when the B2 descriptor for grammatical accuracy “Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.” (CEFR 2001, 114) describes the performance well, it does not mean that it is B2, but that the A2/B1 task has been solved very well and that the (grammar) performance is a very good A2 or B1 performance respectively.

The writing scripts are assessed on four dimensions: Task Achievement, Coherence and Cohesion, Grammar, and Vocabulary. Whereas the last three are informed by the CEFR and the Austrian *Bildungsstandards Englisch* (Gassner et al. 2005), the CEFR does not contain anything on task achievement. In our view, however, the content aspect of writing is central and largely responsible for the overall quality of a script. Nevertheless, all four dimensions are rated separately and are reported as a profile, which more often than not is uneven or marked.

Rater characteristics

It has been reported that “Subject specialists and language trained EFL teachers demonstrate a tendency to employ rating instruments differently” (Elder 1992, in Shaw & Weir 2007, 169). In this respect the present situation in Austria is uncomplicated as all raters are teachers of English who teach in lower secondary schools. Some of these are native speakers now living and working in Austria, some have a university background, others were educated at Colleges of Teacher Education.

Although the raters go through a specific training that familiarizes them with the rating scales and the rating procedures, differences in their experiential background and in their professional training and development may lead to differing assessments of scripts. Harshness and leniency of raters is taken care of through Rasch modelling.

Rating process

Milanovic et al. (1996) identified a number of approaches raters take in the process of rating a script. In our training sessions we generally advise against the 'read through' and the 'provisional mark approach', both of which are based on one reading of the script. Raters are encouraged to adopt a 'principled two scan/read approach' to the process with a focus on task achievement and coherence and cohesion in the first reading and on grammar and vocabulary in the second. The length of the scripts seems to support this approach.

We are aware of group effects on rater reliability as described by Shaw & Weir (2007, 174-175) and have made an effort to use them to advantage in the standardization meetings at the beginning of the training sessions and the rating session. In addition to the procedures recommended for standardization meetings (Alderson, Clapham & Wall 1995, 112-113) a considerable amount of time is spent on the detailed interpretation of the prompts (see appendix, pp. 38-40) and an open discussion of any questions that might be raised by the raters. This takes into consideration that all raters have also been involved in the writing of prompts and their piloting.

Rating conditions

Knowing that rating is a strenuous and demanding job, we have attempted to create a supportive environment and optimal external conditions. The rating sessions take place in a conference hotel with excellent seminar facilities and a number of options for recreation. All scripts are rated on-site. Raters are given batches of scripts for rating within a reasonably tight time frame. Experience has shown that the vast majority of raters work faster, which lengthens their phases for recharging their batteries and restoring good energy levels.

The rating session takes place in May or June and lasts three to four days. It starts with a standardization meeting and then continues with rating sessions based on a detailed rating plan prepared by the LTC Klagenfurt.

Rater training

According to Alderson, Clapham & Wall, rater training is one of the most essential aspects in an effort to obtain reliable assessments (1995, 105). Lumley refers to

a growing body of work that shows the effectiveness of the training process in allowing raters to develop adequate reliability or agreement in the use of individual scales in relation to specified test tasks (2005, 62).

The E8 Testing Team at the LTC Klagenfurt has taken this very seriously and developed a one-year training programme for raters starting in October and preparing the raters for the rating session in May or June. This programme is described in some detail below.

RECRUITMENT

In the recruitment phase teachers in Austrian lower secondary schools are approached. As the test takers come from two different types of lower secondary schools, the General Secondary School (Hauptschule) and the Academic Secondary

School (Gymnasium), care has been taken to balance the intake of raters accordingly.

TRAINING SESSION 1: OCTOBER

As the CEFR is the most relevant background document for the E8 Standards, the starting point of the first training session is *The Common European Framework* in general and the *Overall Writing Scales for Production and Interaction* in particular. The familiarization with the CEFR is implemented on the basis of the recommendations made in the *Manual on Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF) (2003, 70-77)*, including sorting tasks. It is made clear at this point that one aspect of writing is related to communicative activities and strategies, another one to linguistic competences.

The Test Specifications are presented and discussed in detail: domains, genres, model of writing, prompt format. In this phase there is a focus on prompt production. There are *Guidelines for prompt writers* that provide assistance in the process of prompt writing.

Each prospective rater takes on the task of writing one long and one short prompt in the weeks following the first training session. All prompts are sent in to the testing team for moderation. Once the prompts have been screened, they are piloted by the prompt writers. Behind this procedure there is the belief that raters need to know about the qualities of prompts and what elements they must contain. This prepares them for better interaction with the test prompts in the actual rating phase.

With regard to differential validity it is important to state that all precautions have been taken at the early stage of prompt writing to avoid test bias. Varying cultural backgrounds and knowledge of the world have been taken into account as well as the given variety of cognitive characteristics, mother tongue, ethnicity and gender.

After an introduction to the Austrian E8 Standards the rating scale is presented and explained. Unfolding the seven bands with four of them defined and working through the four dimensions takes time. The procedure adopted is to look at the seven bands of one dimension, which leads to some theoretical understanding of the scales, but connections to actual scripts are still tenuous. So the raters get two scripts each for individual marking on the first dimension (task achievement). They discuss their assessment in groups of three and the trainer discloses the benchmarked assessments and argues the case. This procedure is repeated for the other three dimensions.

In the second phase of the training workshop the participants get sets of benchmarked scripts which they rate on one dimension before they get a final set of scripts which is rated on all four dimensions. After discussions and argumentation of the judgements for the benchmarked scripts the participants have a rough idea of the complexity of the rating process and the effort it takes to standardize judgements.

The rating sheets filled in by the participants provide a first set of data for the analysis of intra-rater reliability.

TRAINING SESSION 2: NOVEMBER – FEBRUARY

The second training phase is an open one with a fairly loose structure. All participants first write their prompts and get them back from the testing team as screened prompts (in some cases the prompts are returned to the writers for repair). The prompts are piloted in one of their classes so that every participant has around 20 scripts based on the long prompt and another 20 based on the short prompt.

The first online rating phase starts in November with the testing team sending out scripts to the raters for individual rating on each of the four dimensions. The raters have about four weeks to do this and send in their ratings. When all rating sheets have been transmitted to the testing team, the benchmarks are sent out to the raters. This procedure is repeated in December/January.

In February the raters practise their rating skills on their own pilot scripts. They select three long and two short scripts (or vice versa) to be fed into the training process. They rate these scripts and write justifications for their ratings. The digitalised scripts are sent to the LTC team together with the ratings and the justifications. The trainers go through these scripts and select interesting samples for the upcoming training workshop in March.

TRAINING SESSION 3: MARCH

After the standardization meeting and a phase when open questions from Training Session 2 are discussed, the time is spent rating scripts and discussing particular problems arising in the process. As the raters have to handle a number of different prompts in this phase, it normally emerges that raters need more guidance in the analysis of the prompts.

Depending on the number of participants between 50 and 100 scripts are available for training purposes. The LTC team selects scripts with strong agreement between participant ratings and team benchmarks for the first phase, whereas in the second phase there is also work on scripts for which the original rating departs from the benchmarks. This procedure helps to deepen the participants' understanding of the scales.

Inter-rater reliability and intra-rater reliability are monitored and pertinent data on each rater is collected systematically. After this training session, the testing team finalizes the prompts to be used in the May testing.

RATING SESSION: MAY/JUNE

The rating session begins with a standardization meeting, which involves rating a number of scripts based on familiar prompts. The second important part of this meeting is the analysis and interpretation of the prompts used in the actual test. Raters are given detailed information and are invited to discuss any issues that are still unclear.

Then the actual rating begins. There is a rating plan with overlap for double rating and inclusion of already rated scripts for monitoring intra-rater reliability. The rating is done in batches of 20 scripts per half day with free time management for the raters within that half day.

Post-exam adjustments

Although considerable efforts are taken in the training programme to minimize discrepancies in rater behaviour, the ratings are adjusted for any remaining differences in rater severity by means of multifaceted Rasch analysis after the scripts have been marked. This becomes possible by having a certain proportion of scripts (ca. 10 per cent) marked by more than one rater, or, preferably, by all raters, so that rater behaviour can be assessed in terms of model fit as well as severity. Extreme rater behaviour, rare at this stage, is analysed and communicated to the raters as well as discussed in rater training sessions for subsequent test administrations.

Reporting results

The purpose of the E8 Standards consists in diagnosing strengths and weaknesses in the writing competence of Austrian pupils in grade 8. The aim, therefore, is system monitoring rather than certification or selection at the level of individual test takers. Consequently, while the test results are linked to the CEFR, critical cut grades on which to base selection decisions need not be established by the test constructors. Instead, great attention is paid to the way the results are fed back to test takers and other stakeholders via an internet platform, which can be accessed by means of an identification code.

The diagnostic information that results from the writing test is reported on the four dimensions of the Writing Scale (Task Achievement, Coherence and Cohesion, Grammar, Vocabulary). The results for each dimension are reported on a scale from 0 to 7, which enables reference to the CEFR up to B1. Ratings are adjusted for differences in rater severity and task difficulty by means of multifaceted Rasch analysis. The results are therefore comparable across all test takers regardless of which rater rated the performance and what particular prompt the performance is based on. In addition, percent ranks are reported within each of the four ability groups (3 ability groups in General Secondary School plus one ability group in Academic Secondary School). This enables norm-referenced comparisons within each ability group. Thus, for instance, teachers may compare the performance of their class to the performance of all classes of that ability group in the entire sample from all over Austria.

In compliance with political requirements, only the test takers themselves have access to their individual results through a code they are given when sitting the exam. Teachers, school principals and educational authorities only receive aggregated data, which enables them to compare the group relevant to them (class, school, federal state) with other groups at the same level.

Consequential validity

Shaw & Weir (2007, 218) take the term 'consequential validity' from Messik 1989 and interpret it in the light of recent literature to include washback (influences on teaching, teachers, learning, curriculum and materials) and impact (influences on the community at large). The E8 Standards can be envisaged as an instrument to initiate changes in the direction of positive or beneficial washback.

In 2008 two new course books for teaching English to the target group will be launched and both of them claim to be informed by the CEFR and the E8 Standards. This means that text book writers are well aware of the E8 Standards Tests and are adapting their materials towards them. To take an example from *Testing Reading* (Gassner et al. 2007), the emphasis in the E8 Testing on reading strategies is expected to lead to a re-orientation in the way reading lessons are planned and executed. Awareness-raising even at early stages of learning will eventually lead to a different approach to reading and to improved reading competence.

The requirements for the writing test are clearly laid down in this document and will demonstrate what kinds of writing our learners are expected to deliver. It is hoped that this will lead to less scaffolded writing, thus enhancing learner empowerment. The emphasis given to coherence and cohesion in the CEFR and the E8 Standards might also focus teacher attention on this area and entail improvements.

E8 Writing Test Specifications Version 02 (May 2008)

1. Purpose of the test

The purpose of the writing test is to diagnose strengths and weaknesses in test takers' writing competence. This feedback is of interest to the test takers themselves, their parents, teachers, school principals, regional and federal school boards, and the Ministry of Education and Culture.

2. Description of test takers

The test takers are Austrian pupils in General Secondary School (*Allgemeinbildende Pflichtschule (APS)*) and Academic Secondary School (*Allgemeinbildende Höhere Schule (AHS)*) towards the end of grade 8 (8. Schulstufe). Pupils from all three ability groups in APS will be tested. The majority of test takers will be aged 14.

3. Test level

The difficulty level of the test is supposed to encompass levels A2 to B1 in the *Common European Framework of Reference*.

4. Test Construct

Since the purpose of the test is diagnosis, the most significant competences needed for writing have to be identified for assessment purposes. This is first and foremost the communicative competence demonstrated in an appropriate response to the task. Secondly, it is the ability to produce fluent text by using adequate devices to create coherence and cohesion on paragraph and text level. Thirdly, a good knowledge of a range of grammatical structures and the ability to use them accurately, and fourthly, the choice of vocabulary that is relevant to the content, accurate and has a certain range.

The model of writing discourse which is to be used to construct items is specified in Figure 6. It lists a number of writing purposes and links these with text types and descriptors from the *Bildungsstandards Englisch* (Gassner et al. 2005)

The test is designed to elicit language samples that allow the candidates to be assessed in four areas: Task Achievement, Coherence and Cohesion, Grammar, and Vocabulary. An analytic rating scale is used by trained assessors. A sufficiently large sample of scripts is double-rated to ensure reliability. Differences in rater severity will be adjusted for in the process of multi-faceted Rasch analysis.

Model of the writing discourse

Dominant Intention/ Purpose	Primary Audience	Writing Activity / Text Types	Deskriptoren aus <i>Bildungsstandards FS Englisch, Version Sept. 05*</i>	Descriptors from the CEFR
To convey emotions, feelings	Self Others	<ul style="list-style-type: none"> • Personal story • Personal diary • Personal letter • Postcard • Notes 	<ul style="list-style-type: none"> • Kann Erfahrungsberichte schreiben, in denen Gefühle und Reaktionen in einem einfachen, zusammenhängenden Text wiedergegeben werden. • Kann eine Geschichte erzählen. • Kann Karten, persönliche Briefe und E-Mails schreiben und darin auch über Erfahrungen, Gefühle, Ereignisse berichten. 	<ul style="list-style-type: none"> • <i>Can write accounts of experiences, describing feelings and reactions in simple connected text. – B1</i> • <i>Can narrate a story. – B1</i> • <i>Can write very short, basic descriptions of events, past activities and personal experiences. – A2</i> • <i>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. – B1</i> • <i>Can write personal letters describing experiences, feelings and events in some detail. – B1</i>
To inform	Others	<ul style="list-style-type: none"> • Narrative report • Fill in a form • Give directions • Simple technical description • Notes 	<ul style="list-style-type: none"> • Kann Karten, persönliche Briefe und E-Mails schreiben und darin auch über Erfahrungen, Gefühle, Ereignisse berichten. • Kann kurze, einfache Notizen und Mitteilungen schreiben, die sich auf unmittelbare Bedürfnisse beziehen. • Kann kurze, einfache fiktive Biographien und einfache Gedichte über Menschen schreiben. 	<ul style="list-style-type: none"> • <i>Can write very short, basic descriptions of events, past activities and personal experiences. – A2</i> • <i>Can write short, simple formulaic notes relating to matters in areas of immediate need. – A2</i> • <i>Can write short, simple imaginary biographies and simple poems about people. A2</i>

* In *ÖSZ 2007 (Praxishandbuch)* only minor changes were made in the formulation of two descriptors. In order to maintain consistency in our work, the original descriptors designed in 2005 rather than the ones proposed in *ÖSZ 2007* are still being used here.

To convince, persuade	Others	<ul style="list-style-type: none"> • Statement of personal views, opinions • Letter of application • Letter of advice • Notes 	<ul style="list-style-type: none"> • Kann in Form verbundener Sätze etwas über alltägliche Aspekte des eigenen Umfelds schreiben, wie z.B. über Familie, andere Menschen, Orte, Lebensumstände. • Kann kurze, einfache Notizen und Mitteilungen schreiben, die sich auf unmittelbare Bedürfnisse beziehen. 	<ul style="list-style-type: none"> • <i>Can write about everyday aspects of his/her environment, e.g. people, places, a job or study experience in linked sentences. – A2</i> • <i>Can write short, simple formulaic notes relating to matters in areas of immediate need. – A2</i>
To entertain, please	Others	<ul style="list-style-type: none"> • Given an ending – create a story • Create an ending • Use a visual impulse to create a story 	<ul style="list-style-type: none"> • Kann eine Geschichte erzählen. • Kann kurze, einfache fiktive Biographien und einfache Gedichte über Menschen schreiben. • Kann einfache Texte zu Bildimpulsen, Satzanfängen und Schlüsselwörtern (key words) schreiben. 	<ul style="list-style-type: none"> • <i>Can narrate a story. – B1</i> • <i>Can write short, simple imaginary biographies and simple poems about people. – B1</i> • <i>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. – B1</i>
To keep in touch	Others	<ul style="list-style-type: none"> • Postcards • Letters • E-Mails • Notes 	<ul style="list-style-type: none"> • Kann Karten, persönliche Briefe und E-Mails schreiben und darin auch über Erfahrungen, Gefühle, Ereignisse berichten. • Kann kurze, einfache Notizen und Mitteilungen schreiben, die sich auf unmittelbare Bedürfnisse beziehen. 	<ul style="list-style-type: none"> • <i>Can write very short, basic descriptions of events, past activities and personal experiences. – A2</i> • <i>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. – B1</i> • <i>Can write personal letters describing experiences, feelings and events in some detail. – B1</i> • <i>Can write short, simple formulaic notes relating to matters in areas of immediate need. – A2</i>

Figure 6: Partly based on Vähäpääsi, 1982. Quoted in Weigle 2002, 8-9

5. Structure of the test

The test contains 2 sections. Section 1 consists of a short writing task with an expected response of 40 to 60 words. Section 2 consists of a long writing task with an expected response of 120 to 150 words.

The two tasks will be assessed separately on the basis of the four dimensions of the Writing Scale (pp. 24-28).

6. Time allocation

Total testing time available: 45 minutes.

Time for administration at the beginning (handing out test booklets): 5 minutes.

Time for administration at the end (collecting test booklets): 5 minutes.

Working time: 35 minutes. The short task should take about 10 minutes, the long task about 20, with 5 minutes for revision.

7. Item formats

The candidates' scripts will be handwritten on two separate A4 sheets. The writing task is guided by prompts that ensure that the candidates produce enough language that makes reliable and valid assessment possible.

The prompts may contain black and white pictures or drawings. They need to be appropriate for the age and at a language level no higher than A2. Input texts should be authentic, if at all possible, and as long as necessary to contextualize the task. Ideally, they should not be longer than 50 words.

Prompts are developed to be free of stereotypes. They offer the opportunity to write from experience, but are designed not to intrude on the students' personal feelings.

8. Rubrics

All rubrics are in English. However, they must be formulated in language that is well within reach of the candidates' expected language level and therefore easily understandable for all test takers. Test takers must not be put at a disadvantage because they have difficulty understanding the rubrics.

Rubrics need to indicate the reason for writing, the audience and the required text type. The required length of the text will be indicated in number of words. All expected content points are to be clearly mentioned in the prompts.

9. Writing Rating Scale

The following pages (24-28) include the four dimensions of the analytic rating scale for writing and more detailed scales for each of the four dimensions, where the links to the *Bildungsstandards Englisch* (Gassner et al. 2005) and the *Common European Framework of Reference for Languages* (2001) are made more visible.

Writing Rating Scale (May 2008)

	Task Achievement	Coherence and Cohesion	Grammar	Vocabulary
7	<ul style="list-style-type: none"> complete task achievement meets text type requirements 	<ul style="list-style-type: none"> cohesive on both sentence and paragraph level clear, coherent text 	<ul style="list-style-type: none"> good range of structures few inaccuracies 	<ul style="list-style-type: none"> good range of vocabulary generally accurate with some incorrect words
6				
5	<ul style="list-style-type: none"> good task achievement few inconsistencies in text type requirements 	<ul style="list-style-type: none"> good sentence level cohesion some paragraph level coherence and cohesion 	<ul style="list-style-type: none"> generally sufficient range of structures for familiar contexts occasional inaccuracies message clear 	<ul style="list-style-type: none"> sufficient range of vocabulary, communicating clear ideas occasionally inaccurate
4				
3	<ul style="list-style-type: none"> sufficient task achievement some inconsistencies in text type requirements 	<ul style="list-style-type: none"> some simple sentence level cohesion frequent lack of paragraph level coherence and cohesion 	<ul style="list-style-type: none"> limited range of simple structures frequently inaccurate generally without causing breakdown 	<ul style="list-style-type: none"> limited range of vocabulary, mostly communicating clear ideas frequently inaccurate vocabulary tendency to lift phrases from prompt
2	•	•	•	•
1	<ul style="list-style-type: none"> some task achievement does not meet text type requirements 	<ul style="list-style-type: none"> extremely limited cohesion on sentence and paragraph level text not coherent 	<ul style="list-style-type: none"> extremely limited range of structures mostly inaccurate frequently breakdown of communication 	<ul style="list-style-type: none"> extremely limited range of vocabulary, communicating few clear ideas mostly inaccurate vocabulary several chunks lifted from prompt
0	no task achievement	no assessable language	no assessable language	no assessable language

Task Achievement

	Scales*	Extended Scales*	CEFR
7	<ul style="list-style-type: none"> complete task achievement meets text type requirements 	<ul style="list-style-type: none"> Shows complete task achievement with all content points mentioned and most of them elaborated. Meets text type requirements. 	No descriptors available
6			
5	<ul style="list-style-type: none"> good task achievement few inconsistencies in text type requirements 	<ul style="list-style-type: none"> Shows good task achievement with nearly** all content points mentioned and some elaborated. Generally meets text type requirements with a few inconsistencies. 	No descriptors available
4			
3	<ul style="list-style-type: none"> sufficient task achievement some inconsistencies in text type requirements 	<ul style="list-style-type: none"> Shows sufficient task achievement with most content points mentioned and some elaborated. Meets text type requirements with some inconsistencies. 	No descriptors available
2			
1	<ul style="list-style-type: none"> some task achievement does not meet text type requirements 	<ul style="list-style-type: none"> Shows some task achievement with some content points to be mentioned. Does not meet text type requirements. 	No descriptors available
0	no task achievement	no task achievement	

*** Word limit**

40-60 words: anything below 37 or above 80 words will be penalized (downgrading by one band).

120-150 words: anything below 110 or above 180 words will be penalized (downgrading by one band). Fewer than 80 words - downgrading by two bands.

** In this descriptor “nearly” only applies to the long task, meaning, for instance, 5 out of 6 content points; in short tasks all content points must be mentioned.

Coherence and Cohesion

	Scales	Extended Scales	CEFR
7	<ul style="list-style-type: none"> cohesive on both sentence and paragraph level clear, coherent text 	<ul style="list-style-type: none"> Shows cohesion on both sentence and paragraph level using a limited number of cohesive devices. Text is clear and coherent. There may be some occasional 'jumpiness' with complex ideas. 	<i>Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some 'jumpiness' in a long contribution.</i>
6			
5	<ul style="list-style-type: none"> good sentence level cohesion some paragraph level coherence and cohesion 	<ul style="list-style-type: none"> Shows good sentence level cohesion as a linear sequence on a simple level. Shows some paragraph level coherence and cohesion. 	<i>Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.</i>
4			<i>Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.</i>
3	<ul style="list-style-type: none"> some simple sentence level cohesion frequent lack of paragraph level coherence and cohesion 	<ul style="list-style-type: none"> Shows some simple sentence level cohesion using <i>simple connectors</i> like 'and', 'but' and 'because'. On paragraph level there is a frequent lack of coherence and cohesion. 	<i>Can link groups of words with simple connectors like 'and', 'but' and 'because'.</i>
2			
1	<ul style="list-style-type: none"> extremely limited cohesion on sentence and paragraph level text not coherent 	<ul style="list-style-type: none"> Shows basic linear connectors ('and', 'then') on word or word group level. Text is not coherent. 	<i>Can link words or groups of words with very basic linear connectors like 'and' or 'then'.</i>
0	no assessable language	no assessable language	

Grammar

	Scales	Extended Scales	CEFR
7	<ul style="list-style-type: none"> good range of structures few inaccuracies 	<ul style="list-style-type: none"> Shows a good range of structures for most communicative needs. Communicates with few inaccuracies and a relatively high degree of grammatical control. 	<i>Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.</i>
6			<i>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.</i>
5	<ul style="list-style-type: none"> generally sufficient range of structures for familiar contexts occasional inaccuracies message clear 	<ul style="list-style-type: none"> Shows a generally sufficient range of structures for familiar contexts; mostly uses a repertoire of 'routines' for predictable situations. Communicates with occasional inaccuracies which can impair communication, but message is clear. L1 influence is noticeable. 	<i>Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.</i>
4			
3	<ul style="list-style-type: none"> limited range of simple structures frequently inaccurate generally without causing breakdown 	<ul style="list-style-type: none"> Shows a limited range of simple structures. Communicates frequently inaccurately with basic mistakes, e.g. tenses or verb-noun agreement, but generally without causing breakdown; message is usually clear. 	<i>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.</i>
2			
1	<ul style="list-style-type: none"> extremely limited range of structures mostly inaccurate frequent breakdown of communication 	<ul style="list-style-type: none"> Shows an extremely limited range of simple structures or patterns within a learnt repertoire. Communicates with limited control, frequently causing breakdown of communication. 	<i>Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.</i>
0	no assessable language	no assessable language	

Vocabulary

	Scales	Extended Scales	CEFR
7	<ul style="list-style-type: none"> • good range of vocabulary • generally accurate with some incorrect words 	<ul style="list-style-type: none"> • Shows a good range of vocabulary dealing with most general topics. Can vary formulation to avoid repetition. • Shows vocabulary that is generally accurate with some incorrect words. 	<p><i>Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution. Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.</i></p>
6			
5	<ul style="list-style-type: none"> • sufficient range of vocabulary, communicating clear ideas • occasionally inaccurate 	<ul style="list-style-type: none"> • Shows a sufficient range of vocabulary dealing with familiar topics and everyday life, communicating clear ideas. • Occasionally uses inaccurate vocabulary; major errors occur when expressing more complex ideas or dealing with less familiar topics. 	<p><i>Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events. Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts</i></p>
4			<p><i>Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.</i></p>
3	<ul style="list-style-type: none"> • limited range of vocabulary, mostly communicating clear ideas • frequently inaccurate vocabulary • tendency to use phrases from prompt 	<ul style="list-style-type: none"> • Shows a limited range of vocabulary dealing with concrete everyday needs or familiar topics, communicating clear ideas. • Frequently uses inaccurate vocabulary, controlling a narrow repertoire dealing with concrete everyday needs. • Shows a tendency to use phrases from the prompt. 	<p><i>Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs. Can control a narrow repertoire dealing with concrete everyday needs.</i></p>
2			
1	<ul style="list-style-type: none"> • extremely limited range of vocabulary, communicating few clear ideas • mostly inaccurate vocabulary several chunks lifted from prompt 	<ul style="list-style-type: none"> • Shows an extremely limited range of vocabulary communicating few clear ideas. • Uses mostly inaccurate vocabulary, frequently causing breakdown of communication. • Several chunks are lifted from the prompt. 	<p><i>Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.</i></p>
0	no assessable language	no assessable language	

10. Prompts and performance samples with justifications

The task prototypes below are taken from the 2007 E8 Writing Test. It is important that the tasks are structured in a way that Task Achievement can be measured. The expected length of the scripts, however, needs to be adapted according to point 5 in the specifications.

10.1 Long task

Long prompt from E8 Testing 2007

You will have 20 minutes to do this task.
Read the instructions carefully and then write your text on the separate sheet.
Your text must be 120-150 words. If you write more than 180 words, you will get minus points.
Do not forget to make paragraphs!
Please count your words and write the number at the end of your text.

You have just moved to another town/village.

Write a **letter** to your American/English friend in which you tell him/her about your new situation.

Inform him/her about

- your new place of living
- the reason for moving

Describe

- the town/village you're living in now (buildings, people,...)
- your new home

Tell about

- the first days of your 'new life' (new school, teachers,...)
- how you feel about your new situation

10.1.1 Script 1

Dear Bill,

how are you? Now I'm living in Vienna. That is in Austria. It is a very big city with nice people. There is also a fun fair called "Prater". My parents got divorce and so I'm living here with my mother. There are wonderful buildings in this city like the animal park "Schönbrunn" and many castles. I like the river "Donau" very much, because I often go swimming there. My new house is very big and next to it is a forest. I like that. The first day of my "new life" was not so good. When I came into my class most of the pupils laughed at me but the teacher was nice. I hope you will write back.

Yours, Raphael

(124 words)

Justifications

Task Achievement 6

The text meets the text type requirements, is within the word limit set, and uses an informal register. The candidate works his way through the content points, only just touching on the last one. All the others are mentioned, but elaboration is often on a limited scale. Although there is no elaboration on how the writer feels about the new situation and the salutation is rather short, there is enough elaboration on the new place (city and house) to justify band 6.

Coherence and Cohesion 4

There are no paragraphs and text organisation is informed by the sequence of the content points provided. The “Prater” sentence, which belongs to content point 3, is displaced. The text is certainly cohesive at sentence level but not at paragraph level. Moving from one idea to the other – telling about the new place – the reason for moving and then describing the new town may be implied by the order of the content points; nevertheless, the writer could have made paragraphs to make the text less jumpy. While band 5 could be considered for this performance, the lack of paragraph organization leads to downgrading by one band to band 4.

Grammar 5

The sentence structures used are generally very simple, but show good command within the range chosen. There are hardly any mistakes (e.g. *got divorce*). Two “complex” sentences are linked by a subordinating conjunction. Whereas the range has clear limitations, it seems generally sufficient displaying “[reasonably accurate use of] a repertoire of frequently used ‘routines’ and patterns.”

Vocabulary 6

The candidate has a sufficient range of vocabulary to express himself. On the one hand, there is rather unexpected vocabulary used (nearly) correctly (*get divorce, fun fair, laugh at*) as evidence of a good range of vocabulary, on the other, there is some L1 influence (*animal park*). But the ideas communicated are always clear. The range of vocabulary stays within topics pertinent to everyday life (band 5), hence band 6.

10.1.2 Script 2

Hi Steven!

In Salzburg it is very cool and I'm living in a small flat with 5 rooms two bedrooms a kitchen a livingroom and a bathroom. Our garden is not so big, but big enough for us. The building very beautiful and it give no skyscrapers and it is very hot. I always go in the garden and I lie in the sun. The people are very funny and they accept that I speak english. I have got two new friends and they speak very good english. The teachers are very good and we have a lot of english and I'm the best one, but in Deutsch I'm very bad. I feel very good with my new situation and I wish all my old friends and the teachers a good luck for the next time and I hope you always wish me a good next time.

Yours Olav!

(149 words)

Justifications

Task Achievement 4

The text meets the text type requirements, is within the word limit set, and uses an informal register. The text briefly mentions Salzburg, but leaves out the reasons for moving. The other four content points are all mentioned, but mixed up in the presentation. Content point 5 is not directly mentioned, but two statements could refer to it, content point 6 is lifting from the prompt adding the word 'good'. The text is not good enough for band 5, but better than band 3.

Coherence and Cohesion 3

The simplest connector "and" occurs with undue frequency, simply stringing ideas together without expressing logical relations. There are no paragraphs, so text organisation is quite low with simple addition as the dominating structuring principle. The text does not read well because some sentences that follow each other have no connection at content level. Sometimes things are joined in one sentence that have little in common (*The building very beautiful and it give no skyscrapers and it is very hot.*).

Grammar 4

The candidate uses simple structures correctly, repeating the same basic pattern with little variation. Whereas the range of structures chosen from "a repertoire of frequently used 'routines' and patterns" seems generally sufficient, the general simplicity of the structures and the breakdown in a basic situation (*it give*) indicate that band 5 is not applicable. On the other hand, the relatively high degree of "correctness" rules out band 3.

Vocabulary 3

At first sight the range seems sufficient and control of elementary vocabulary is given. However, sometimes the meaning is not clear: "*The building very beautiful and it give no skyscrapers [...]*." – Here we do not know whether the speaker refers to the new house they are living in or the buildings of the town. In the last sentence the speaker seems to be attempting too much, leaving the safe area, and it results in several breakdowns (*I wish them a good luck; L1: for the next time; I hope you always wish me a good next time*) demonstrating the limitations. Whereas range seems to point towards band 4, it is the lack of control leading to unclear messages that supports band 3.

10.2 Short Task

Short prompt from E8 Testing 2007

You will have 10 minutes to do this task.
Read the instructions carefully and then write your text on the separate sheet.
Your text must be 40-60 words. If you write more than 80 words, you will get minus points.
Please count your words and write the number at the end of your text.

Your friend's birthday party was a few days ago. Write **an e-mail** to tell him/her that you liked the party.

- Tell him/her why you liked the party.
- Tell him/her what you liked best.
- Ask your friend when you are going to meet again.
- Suggest something for the next weekend.

10.2.1 Script 3

Dear Daisy, how are you? Let's talk about your party. It was so great! I liked the party best, because Lukas was there. And I liked the games we played. Oh and tell you mum, that the food was excellent! What are you going to do on Sunday? Maybe we can go to cinema or swimming. Tell me please if you have time. Okay I have to help my mum with dinner.

Love you big kiss

Yours,

Aida

(75 words)

Justifications

Task Achievement 7

Text type requirements are perfectly met and all content points are mentioned and (within the format given) elaborated, so we have complete task achievement.

Coherence and Cohesion 7

The text flows well as the individual parts are well-connected and the presentation of the content points follows an internal logical order. The text is cohesive on sentence and text level using a number of cohesive devices.

Grammar 7

There is a high degree of grammatical control with only one slight slip in accuracy (*go to cinema*). In addition, there is a good range of structures showing variation and flexibility.

Vocabulary **6**

The text shows a good range of vocabulary displaying even a certain “chattiness” although it stays within a safe zone without any attempts at aiming higher. The whole text is without errors in vocabulary. However, there are no lexical elements displaying the ability to “vary formulation”.

10.2.2 Script 4

Hey Kevin!!! I've liked your party because it was very cool. What I liked most were all the nice girls and all the nice waterpipes. Can we do such a party again tomorrow? It would be very nice. But please, buy more waterpipes, and more grass! And next weekend, we can do it again, or? I have a better idea, we can go to the city and chillout at a concert.
(65 words)

Justifications**Task Achievement** **4**

The problems with content organisation are obvious, but there is still “sufficient task achievement with most content points mentioned and some elaborated”. The first content point is handled very briefly. Content point 3 (Ask your friend when you are going to meet again) is implied rather than addressed explicitly. But although there is no question for a new meeting, there is a suggestion for the next day and the next weekend. Text type requirements are only partly met (missing closing formula).

Coherence and Cohesion **3**

There is some simple sentence level cohesion using simple connectors. At text level there is noticeable “jumpiness”, some links are awkward, others missing (*And next weekend, we can do it again, or? I have a better idea, we can go to the city...*). Given the simplicity of the sentence structure, there is little opportunity for sentence-level cohesion to manifest itself. However, the last sentence is incoherent at sentence level. Therefore, we can ascertain ‘some sentence level cohesion’. Coherence across sentences, which is better in the first half of the text, declines considerably towards the end.

Grammar **4**

Simple structures are used correctly, but there are basic mistakes (*I've liked*) or lack of structural knowledge (*we can do it again, or?*). Nevertheless, the candidate “uses reasonably accurately a repertoire of frequently used ‘routines’ and patterns” albeit within a strictly limited range.

Vocabulary **5**

The candidate has sufficient knowledge to express himself and communicate his (rather unconventional) ideas. There are occasional inaccuracies, but also nice phrases like “*What I liked most*” or “*all the nice girls*”.

10.2.3 Script 5

My friend have at 7.5. birthday. The birthday-party was very good. We had a lot of fun on the party. We play playstation and we evenb play football but the best was that we are ate pizza. We go at the weekend to a football match.
(46 words)

Justifications

Task Achievement 1

Text type requirements are not met. There is no salutation and no ending to the message which reads like a report rather than an e-mail message. Some content points are mentioned. Content point 1 is unsuccessfully dealt with, CP 2 is ok, CP 3 is missing completely, and CP 4 is not recognizable as a suggestion. So there is some task achievement with some content points mentioned.

Coherence and Cohesion 2

Text organisation is very low. There is a string of unconnected sentences with groups of words linked by basic connectors. However, in one sentence a number of connectors are used fairly successfully, so it is clearly better than band 1, but not good enough for band 3.

Grammar 1

The text shows an extremely limited range of simple structures with a number of basic mistakes (*my friend have; we are ate; on the party; at 7.5.; we go at the weekend to a match; plus tense errors*). It is not always clear what the writer wants to say. In the last sentence the message is not clear.

Vocabulary 2

The vocabulary used is basic and the range is very limited. However, within these limitations the candidate is able to communicate some ideas successfully. This makes the performance better than band 1, but it is not "sufficient [...] for the expression of basic communicative needs" (band 3).

10.2.4 Script 6

*My friend's birthday party was a few days ago. Write an e-mail to tell ihm that you liked the party. Tell him why you liked the party. Tell him what you liked best. Ask your friend when you are going to meet again.
Suggest something for the next weekend.
My best friend's. Name von my best friend's is ...*
(58 words)

Justifications

Task Achievement 0

There obviously is no task achievement as the candidate has merely copied the rubrics instead of doing what they said. The only addition made by the writer is minimal and bears no relation to the task set.

On the grounds of task achievement being band zero, the other 3 dimensions are not assessable as we could only assess the rubrics, but not any of the candidate's competences.

Coherence and Cohesion 0

No assessable text in terms of coherence.

Grammar 0

Too little independently produced language to allow assessment.

Vocabulary 0

Too little independently produced language to allow assessment.

Literature

- Alderson, J.C. 2004. Washback in language testing. Research contexts and methods. In: Cheng, L., Watanabe, Y. & Curtis, A. (Eds.). *Context and Method in Washback Research: The influence of language testing on teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- Alderson, J.C., Clapham, C. & Wall, D. 1995. *Language Test Construction and Evaluation*. Cambridge: University Press.
- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: University Press.
- KET Handbook. Cambridge UCLES. Retrieved 28 November 2007 from: http://www.cambridgeesol.org/assets/pdf/resources/teacher/ket_handbook.pdf
- KET Vocabulary List. Cambridge UCLES. 2006. Retrieved 17 January 2008 from: <http://www.cambridgeesol.org/teach/ket/Website%20KET%20Vocablist%202006.pdf>
- PET Handbook. Cambridge UCLES. Retrieved 28 November 2007 from: http://www.cambridgeesol.org/assets/pdf/resources/teacher/pet_handbook.pdf
- PET Vocabulary List. Cambridge UCLES. 2006. Retrieved 17 January 2008 from: <http://www.cambridgeesol.org/teach/pet/Website%20PET%20Vocabulary%20List%202006.pdf>
- Canale, M. & Swain, M. 1980. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. In: *Applied Linguistics* 1 (1), 1-47.
- Council of Europe (Ed.). 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: University Press.
- Ek, J. A. van [Council of Europe]. 1979. *The Threshold Level for modern language learning in schools*. Harlow: Longman.
- Ek, J. A. van & Alexander, L. G. 1980. *Waystage English*. An intermediary objective below Threshold Level in a European unit/credit system of modern language learning by adults. Prepared for the Council of Europe in association with M. A. Fitzpatrick. Oxford: Pergamon Press.
- Field, J. 2004. *Psycholinguistics. The Key Concepts*. London: Routledge
- Gassner, O., Horak, A., Mewald, C., Moser, W., Schober, M., Stefan, F. & Valsky, C. 2005. *Bildungsstandards in Österreich. Fremdsprachen. Englisch 8. Schulstufe. Version September '05*. Ed. by bm:bwk. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur.
- Gassner, O., Mewald, C. & Sigott, G. 2007. *Testing Reading. Specifications for the E8-Standards Reading Tests. LTC Technical Report 2*. Wien: bm:ukk. Available as download from: <http://www.uni-klu.ac.at/ltc/inhalt/346.htm>
- Gerngroß, G., Puchta, H., Davis, R. & Holzmann, C. 1997. *The New You & Me, Textbook 4. Basic course*. Wien: Langenscheidt Longman.
- Grabe, W. & Kaplan, R.B. 1996. *Theory and Practice of Writing. An Applied Linguistic Perspective*. London: Longman.
- Hyland, K. 2002. *Teaching and Researching Writing*. London: Longman.

- Kellogg, R. T. 1994. *The Psychology of Writing*. New York: Oxford University Press.
- Kellogg, R. T. 1996. A model of working memory in writing. In: Levy, C. M. & Randsdell, S. (Eds.) *The Science of Writing*. Mahwah, NJ: Lawrence Erlbaum, 57-72.
- Lehrplan der Hauptschule: Lebende Fremdsprache (alle Sprachen) 2007. Retrieved 10 May 2008 from: <http://www.bmukk.gv.at/medienpool/887/hs23.pdf>
- Lumley, T. 2005. *Assessing Second Language Writing. The Rater's Perspective*. Frankfurt: Peter Lang.
- ÖSZ (Ed. Wolfgang Moser) 2007. *Bildungsstandards für Fremdsprachen (Englisch), 8. Schulstufe, Praxishandbuch*. Praxisreihe Band 4. Graz: ÖSZ
- Papajohn, D. 1999. The effect of topic variation in performance testing: the case of the chemistry TEACH test for international teaching assistants. In: *Language Testing* 16 (1), 52-81.
- Read, J. 1990. Providing Relevant Content in an EAP Writing Test. In: *English for Specific Purposes* 9 (2), 109-21.
- Scardamalia, M. & Bereiter, C. 1987. Knowledge telling and knowledge transforming in written composition. In: Rosenberg, S. (Ed.) *Advances in Applied Psycholinguistics, Vol. 2: Reading, writing, and language learning*. Cambridge University Press, 142-175.
- Shaw, S.D. & Weir, C.J. 2007. *Examining Writing. Research and practice in assessing second language writing*. Cambridge: University Press.
- Sigott, G., Gassner, O., Mewald, C. & Siller, K. 2007. *E8-Standardstests. Entwicklung der Tests für die rezeptiven Fertigkeiten: Überblick. LTC Technical Report 1*. Language Testing Centre, Alpen-Adria-Universität Klagenfurt.
- Tankó, G. 2005. *Into Europe. The Writing Handbook*. Budapest: Teleki László Foundation.
- Weigle, S.C. 2002. *Assessing Writing*. Cambridge: University Press.
- Weir, C.J. 2005. *Language Testing and Validation. An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.

Appendix

Prompt interpretation: Long prompt

You have just moved to another town/village.

Write a **letter** to your American/English friend in which you tell him/her about your new situation.

Inform him/her about

- your new place of living (1)
- the reason for moving (2)

Describe

- the town/village you're living in now (buildings, people,...) (3)
- your new home (4)

Tell about

- the first days of your 'new life' (new school, teachers,...) (5)
- how you feel about your new situation (6)

Task Achievement

General description

- The text type is clearly marked as a personal letter, which is why we expect an informal opening and closing formula like **Dear Jim / Hi, Sue** and **Yours, / Love** etc.
- The register needs to be **informal**.
- Although the first content point can be dealt with very briefly (new place), the second point (reasons for moving) allows for some elaboration.
- The next four points (describe town/village and new home; first days and feelings) definitely allow for elaboration and we expect somewhat detailed descriptions (points 3, 4, 5).
- The last point about feelings is more difficult because this requires some creative thinking, which is why we do not expect much elaboration. If there is elaboration here, it needs to be specially recognized.

Text type requirements

- Text type requirements are **not met** if there is no opening or closing formula AND if the register is not appropriate (e.g. rude language or mismatch of informal situation and formal language). This will lead to downgrading.
- Missing opening or closing formula alone will lead to downgrading by one band.
- Inappropriate register/tone alone will lead to downgrading by one or two bands depending on text. The same applies to an impolite or offensive tone.

Degrees of Task Achievement

In this long text, all content points need to be mentioned for **complete task achievement**. If one of the major points (3, 4, 5) is missing, complete task achievement cannot be awarded. If one of the major points is only mentioned, the other two points must compensate this through ample elaboration. The lack of elaboration in the last point about feelings should not be especially penalized as this is considered to be difficult to achieve.

Good task achievement can still be reached if one minor point and/or one of the major points (3, 4, 5) is missing. 'Nearly all' suggests that 5 content points must be mentioned. Depending on the degree of elaboration, we would expect at least two, but preferably three points to be elaborated for good TA.

For **sufficient task achievement** we would expect four content points to be mentioned and one or two to be elaborated, depending on the degree of elaboration.

Some task achievement would be given if the message was clear enough to convey the information that the person has moved, where to and some brief information about one of the other points.

Coherence and Cohesion

In a long text we definitely expect paragraphing. Since the content points are grouped into three units, the test takers are provided with sufficient hints as to how the text could be structured into paragraphs. A lack of paragraph organization in bands 5 to 7 leads to downgrading by one band.

- Test takers should make use of paragraphs and some cohesive devices to reach band 7. Within the three units the text should be coherent.
- For band 5, cohesion should be achieved at least at sentence level, and longer stretches should show some paragraph level cohesion. For example this could be demonstrated in the description of the new place and/or home.
- For band 3, we would at least expect the test takers to use simple connectors (e.g. *because, but, and, then*) when giving reasons for moving or for the description of places and/or the first days at school.
- For band 1, it would be enough if the sentences were linked with very basic connectors. We do not expect paragraphing or textual coherence for band 1.
- If the letter is written in chunks rather than sentences, we suggest that there is no assessable language for coherence and cohesion.

Prompt interpretation: Short prompt

Your friend's birthday was a few days ago. Write an **e-mail** to tell him/her that you liked the party.

- Tell him/her why you liked the party. (point 1)
- Tell him/her what you liked best. (point 2)
- Ask your friend when you are going to meet again. (point 3)
- Suggest something for the next weekend. (point 4)

Task Achievement

General description

- We can expect an informal opening and closing formula like **Dear Jim / Hi, Sue** and **Yours, / Love** etc.
- Point 1: The writers need to give one or several reasons for liking the party. It is not enough to say that they liked the party but this should include giving reasons.
- Point 2: This can be done in a more or less elaborate way depending on the writer's choices and abilities.
- Point 3: This is simple enough and will be dealt with very briefly. Most probably this will immediately lead on to point 4.
- Point 4 is expected to be the most challenging one as it requires some thinking and a minimal amount of creativity. They need to include at least one suggestion referring to the following weekend.

Text type requirements

- For band 7 there must be an appropriate opening and closing formula and the register must be informal and demonstrate the familiarity indicated by "your friend".
- Problems with the beginning or ending of the text or minor problems with the register would lead to downgrading by one band.
- When the problems become more obvious, TA will drop by 2 bands.
- Text type requirements are not met when there is no opening or closing formula AND when the register is not appropriate. (Downgrade by 2 bands or even 3 if the problems with register are very serious).

Content points

With short tasks, the relevant descriptor needs to be interpreted very strictly. In fact, with 4 content points required it makes little sense to tolerate a missing one. It should be taken for granted that ALL 4 content points must be dealt with for complete or good task achievement. This would normally mean a text length of around 60 words.

- It will be the degree of elaboration that distinguishes between complete and good task achievement.
- With one less important point missing (*what you liked best* – if point 1 compensates for it), good task achievement is still possible. If point 1 or 4 is missing, we can ascertain sufficient task achievement. The degree of elaboration will decide whether it is band 3 or band 4.
- If 2 content points are missing, we cannot expect much elaboration either and it seems to point towards band 1.

Coherence and Cohesion

- We do not expect any paragraphing with short tasks.
- We also found that demonstrating this dimension in short texts is rather difficult.
- We can expect the writers to deal with the four content points in the order given. Implicit and explicit linking (connectors) is not easy to place in this kind of text and should be specially acknowledged. Content points 1 and 2 suggest the use of *because* and/or *and*.