



# RISTAL 6/2023

## Research in Subject-matter Teaching and Learning

Citation:

Keller, S. D., Trüb, R., Raubach, E., Meyer, J., Jansen, T. & Fleckenstein, J. (2023). Designing and validating an assessment rubric for writing emails in English as a foreign language. *RISTAL*, 6, 16–48.

DOI: <https://doi.org/10.2478/rystal-2023-0002>

ISSN 1863-0502



© (2023) Stefan D. Keller, Ruth Trüb, Emily Raubach, Jennifer Meyer, Thorben Jansen & Johanna Fleckenstein. This work is licensed under a Creative Commons Attribution 4.0 International License. (CC BY 4.0)

# Designing and validating an assessment rubric for writing emails in English as a foreign language

*Stefan D. Keller\**, *Ruth Trüb\*\**, *Emily Raubach\*\*\**,  
*Jennifer Meyer\*\*\*\**, *Thorben Jansen\*\*\*\*\**  
 & *Johanna Fleckenstein\*\*\*\*\**

## Abstract

Rubrics are defined as coherent sets of criteria for students' work that include detailed descriptions of quality for a specific learning task. They are used in different subjects to guide student learning and assess its results. This study demonstrates the design and development of an assessment rubric for writing emails in English as a foreign language (EFL). As English emails are a key form of communication in the globalised world, it is essential to have a rubric which is both fit for classroom use and empirically validated. In our study, six raters were trained to assess a sample of N = 1017 emails from learners at lower secondary level in Switzerland. We evaluate the reliability of the ratings by assessing inter-rater agreement after a period of training and by comparing their scores to those of an expert rater. Further, we analyse the linguistic quality of learner texts by focusing on four key markers of emails and compare this analysis to our rubric based scores to analyse its face validity. Our results show high reliability and validity of the rubric. We discuss the potential of such rubrics to improve teaching quality in various subjects and learning domains.

## Keywords

English as a foreign language, Writing, Assessment, Rubrics, Emails

---

\* **Stefan D. Keller** is professor of subject-specific education at University of Teacher Education in Zürich. He specializes in English teaching and learning studies.  
 Zürich University of Teacher Education, Lagerstr. 2, 8090 Zürich, stefandaniel.keller@phzh.ch

\*\* **Ruth Trüb** is a lecturer and researcher at University of Applied Sciences and Arts, Northwestern Switzerland (School of Education). She works at the Institute for Primary Education and wrote her PhD thesis on English writing at primary school.  
 University of Applied Sciences and Arts Northwestern Switzerland, School of Teacher Education, Bahnhofstrasse 6, 5210 Windisch, ruth.trueb@fhnw.ch

\*\*\* **Emily Raubach** is a doctoral student at the Department of Social Neuroscience and Social Psychology at the University of Berne, Switzerland. Her research areas include second language acquisition, text analysis, and political psychology.  
 University of Berne, Fabrikstrasse 8, 3012 Bern, emily.raubach@unibe.ch

\*\*\*\* **Jennifer Meyer** is a researcher at IPN in Kiel and specializes on questions of assessment and feedback, as well as on motivational aspects of learning. She is currently engaged in the project FORMAT, which investigates automated feedback for English writing.  
IPN Kiel, Olshausenstraße 62, D-24118 Kiel, [jmeyer@leibniz-ipn.de](mailto:jmeyer@leibniz-ipn.de)

\*\*\*\* **Thorben Jansen** is a researcher at IPN in Kiel and specializes on questions of assessment and feedback, both in English and in the Natural Sciences. He is currently engaged in the project DARIUS, which studies digital argumentation instruction for science.  
IPN Kiel, Olshausenstraße 62, D-24118 Kiel, [tjansen@leibniz-ipn.de](mailto:tjansen@leibniz-ipn.de)

\*\*\*\*\* **Johanna Fleckenstein** is junior professor of Educational Sciences at University of Hildesheim. She has worked extensively on assessment and feedback in English language learning.  
University of Hildesheim, Universitätsplatz 1, D-31141 Hildesheim, [fleckenstein@uni-hildesheim.de](mailto:fleckenstein@uni-hildesheim.de)

## 1 Introduction

Emails in English are an important form of communication in the modern, globalised world. This genre therefore holds a central place in the writing instruction for English as a foreign language (EFL) at lower secondary level in most European countries (EDK, 2011; KMK, 2004). Writing emails is an authentic and relevant task (Gains, 1999) that is appropriate for the target level of lower secondary education (Hallet, 2016). However, writing emails in a foreign language is challenging for learners as it requires specialised, genre-specific knowledge. For example, learners should formulate an expressive subject line, provide background information, state a request politely, convince the recipient to comply, and choose adequate formulae of greeting and closing (Campbell, 2013). Furthermore, EFL learners need to have an adequate awareness of the register (or level of formality) in which they write, particularly in emails that contain a specific request to the addressee (Al-Ali & Sahawneh, 2008).

One way of helping learners overcome such challenges is by using assessment rubrics. Rubrics are defined as coherent sets of criteria for students' work that include detailed descriptions of quality for those criteria (Brookhart, 2013). They consist of scaled levels of performance for one or more quality dimensions with regard to a particular task or competence. Rubrics are used in many different subjects to evaluate student performances (Allen & Tanner, 2006; Reddy & Andrade 2010). They thus resemble evaluation grids ("grille d'évaluation") that are used in different subjects to assess the outcome of complex learning tasks where students' performances cannot be simply classified as right or wrong, or good or bad (Scallan, 2004).

This paper describes the conceptualisation and validation of an assessment rubric for email writing tasks at lower secondary level. The rubric was designed as a basis for a valid and reliable assessment of student texts, to allow for fast and reliable scoring of learner texts both in the classroom and in writing research. Its second function is to serve as a basis for formative assessment. That is to provide students with information on linguistic structures and organisational aspects of emails in the process of writing.

There is an extensive body of research on the use of rubrics to assess EFL writing competences (Andrade, 2001; Dawson, 2015; Jonsson & Svingby, 2007; Scallan, 2004). Yet little is known about developing and validating an assessment rubric for a specific learning context in a particular subject. In particular, there are few studies on how to validate rubrics for specific genres and levels of schooling. Our study therefore addresses an important research gap for subject-specific teaching and learning which is relevant in many contexts.

To address this research gap, we first discuss the subject-specific learning theory of process-oriented writing which represents the background for the development process. We explain how the rubric was designed and implemented in an on-line learning scenario ( $N = 373$ ) at lower secondary level. We focus especially on the rationale behind the different criteria and quality levels included in the rubric. We then evaluate the reliability of the rubric by presenting data from the rater training phase, looking at how consistent the evaluations of different raters are after a period of training. Next, we focus on the

validity of the rubric in a more comprehensive sense by comparing the rubric-based text scores to the linguistic quality of the texts. To do this, we performed an in-depth analysis of four key linguistic markers of writing quality (salutation, closing, information about the writer, and matter of concern) and checked whether texts which were rated high on our rubric were also of high linguistic quality.

This study thus compares three different approaches towards validating a rubric (East, 2009): a theoretical approach inquiring whether the rubric is based on a convincing theory of learning in the domain; a consistency approach testing whether the rubric can be used to grade learner texts reliably by different raters; and an evaluative approach testing whether the rubric-based scores are a valid measurement of quality in a particular domain. We suggest that this is a procedure which can also be applied in other subjects and domains to validate rubrics for learning and assessment.

## 2 Background

### *2.1 Rubrics as instruments of assessment in research and the classroom*

In educational contexts, rubrics are applied both as summative and formative instruments (Jonsson & Svingby, 2007). As tools of summative assessment, rubrics can be used to improve the accuracy and efficiency grading, making it more transparent, objective and reliable (Andrade, 2005). As tools of formative assessment, they can guide and foster student learning by making apparent what is to be learnt, and by helping students generate self- and peer-assessment (Andrade et al., 2009; Council of Europe 2001).

By making teachers' expectations clear, and by showing students how to meet these expectations, rubrics can also improve student performance. Based on 21 empirical studies, Panadero and Jonsson (2013) showed that rubrics improved student performance by (a) increasing transparency about what is to be learnt, (b) reducing learner anxiety, (c) aiding the feedback process, (d) improving student self-efficacy, and (e) supporting learning motivation. Specifically for English writing, Sundeen (2014) showed that both explicit instructions in teaching students about the elements contained in a rubric and simply providing them with that rubric were equally effective in improving their writing performance. Working with rubrics while writing, as well as getting rubric-based assessment, is thus a significant factor determining the quality of EFL writing.

Rubrics are also an essential tool in empirical studies of writing. When aiming to assess students' writing competences, it is important to ensure that the inferences made about these competences are appropriate (Weigle, 2002). Without such a reliable assessment tool, it is impossible to measure the construct with sufficient accuracy (East, 2009). The availability of a validated rubric thus benefits both student learning and assessment accuracy.

## 2.2 Rubrics in genre-based writing pedagogy

EFL writing curricula in European countries put considerable emphasis on students mastering key genres of the target language. Emails are particularly important in lower secondary education, as educational standards specify in many European countries (EDK, 2011; KMK, 2004). The increasing focus on knowledge of genres as educational outcomes is due to changing views both of discourse and of learning to write. Genres incorporate a specific understanding of how language is structured to achieve social purposes in particular contexts of use. Genre-based pedagogies “offer principled ways of assisting both pre- and in-service writing teachers to provide their students with targeted, relevant, and supportive instruction” (Hyland, 2007, p. 148).

Rubrics play an important role in a genre-oriented teaching and learning cycle. They can encode key aspects of the target genre that are instrumental to its communicative purpose. Additionally, they provide clear formulations for salient features of a particular text-type and make them transparent to teachers and learners alike. In doing so, rubrics help learners to build an understanding of how and why texts in the target language “are written in the way they are” (Hyland, 2007, p. 151). This typically includes mastering appropriate text structures (schemata), addressing different types of readers (socio-linguistic and pragmatic competences), and characteristic language forms (lexis and grammar). As educational curricula often require learners to master the same genre in different languages, rubrics can also be used to express differences between languages at syntactic, semantic and pragmatic levels – for example between students’ school language and other foreign languages they are studying.

## 2.3 The linguistic structure of emails

Emails as a genre have been well studied and consist of a series of key elements that students need to master. The following elements are relevant for all types of emails, but particularly for email requests in English, which are at the centre of our study:

- a. *Subject line*: This is a typical feature of electronic communication and contains important information about the text that is to follow to the addressee;
- b. *Appropriate forms of greeting (salutation) and closing*: Formulae of opening and closing a text are generic and frequently used in a wide range of written (electronic) communication. In semi-formal email requests, it is especially important that the salutation and closing are appropriate in terms of register, style, and correct level of formality;
- c. *Establishing an interpersonal dimension*: In the type of email request envisioned in our project, it is important for the writers to express who they are, why exactly they are writing the email (the matter of concern), and what type of answer is expected from the addressee;
- d. *Content (completeness)*: An email needs to contain all the information of the relevant situational context in order to fulfil its communicative function;

- e. *Appropriate register and style of language, including linguistic correctness:* To fulfil its communicative purpose, the email needs to be pragmatically and stylistically adequate to the situation and the addressee. Also, it should be free from formal mistakes that make understanding difficult (Al-Ali & Sahawneh, 2008; Hallet, 2016).

When writing semi-formal emails to make requests, the level of formality is particularly important. For example, learners need to state their request clearly and unambiguously while also being polite and avoiding going into too much personal detail. This requires knowledge of genre-specific phrases such as "I am writing to tell you" or "I would like to ask about..." (British Council, 2022). To be a useful tool for learning and assessment in this domain, a rubric should refer to the main aspects of genre-specific communication and express them in a way that is meaningful and understandable for learners.

### 3 The study

The rubric presented in this study was developed as part of an on-line writing unit in which students in 8<sup>th</sup> and 9<sup>th</sup> grade in Switzerland learned how to write semi-formal emails. This level of schooling is referred to as lower-secondary education in Switzerland, with students between 14 and 15 years old (ISCED levels 243 and 244; UNESCO, 2011). In a computer-based setting, the students worked on three consecutive writing tasks of comparable difficulty and received instant, rubrics-based assessment for each text. These assessments were provided by trained raters and were formative in nature. They were given with the intention of helping learners improve their texts by directing their attention towards key features of the email genre. It was important therefore that the rubric should contain transparent and unambiguous criteria that raters could use to score texts reliably in a limited period of time and that the rubric was comprehensible for learners and could help them to improve their texts.

The following research questions guided us in our research:

1. Do measures of inter-rater consistency indicate that the scoring rubric can be used with an acceptable degree of reliability?
2. To what extent does a comparison of scores with certain textual and discourse features of the scripts further indicate the validity of the rubric?

To investigate research question 1, we describe the rater training undertaken in the project. We organised these trainings according to best practice as recommended by General Certificate of Secondary Education (GCSE), which includes intense familiarisation with the rubric, providing raters with exemplars (model texts) for each level of quality, and comparing their rating to that of a "principal examiner" (Greatorex & Bell, 2008). We then compared the reliability from the training phase with scores provided during the operational part of the study, and we also compared them to the scores of two expert raters.

To investigate research question 2, we undertook a linguistic analysis of the entire sample of learner texts, focusing on features which are highly indicative of quality writing in the context of email requests. This analysis was independent of the rubric-

based scoring and relied on a detailed coding of all texts in the software MAXQDA (VERBI Software, 2021). The different elements or paragraphs of the texts were coded according to detailed linguistic criteria (see Sect. 4.3 for details). We then checked whether the texts rated high on our rubric were also linguistically more complex or better formulated than texts rated low. We checked, in other words, whether rubric-based scores were associated with higher levels of genre-specific writing quality, providing an indication of the validity of the rubric (East, 2009).

### 3.1 Developing the rubric

To develop the rubric, the research team first compiled a checklist of the central characteristics of the genre of email requests (Campbell, 2013). This was developed based on an analysis of sample email texts taken from course books for secondary school education, as well as linguistic descriptions of this genre (Al-Ali & Sahawneh 2008; British Council, 2022; Hallet, 2016; see also Sect. 2.3, above). We also analysed different types of rubrics found on-line and in the pedagogical literature, comparing both the form and the different categories contained in them. In this process, we identified the linguistic key elements to operationalise the genre-specific construct of email requests. We then adapted these elements into a rubric with five steps or levels, as is shown in Fig. 1.



Fig. 1 Genre-based rubric for writing emails in English

As this project was implemented at lower secondary level in Switzerland, the individual steps in the rubric were described in German to make sure learners had a clear understanding of each individual aspect of quality. This also made certain that they could use this information in their writing without simply copying certain elements directly. The descriptors were arranged in a stepwise manner to indicate a clear process in which the quality of an email might be improved. These descriptors can be translated as follows:

- Step 1. *Waystage*: Email contains all the information required in the task.
- Step 2. *Partly successful*: Salutation and closing are adequate to the situation described in the task.
- Step 3. *Mostly successful*: Subject line successfully communicates the intention of the email.
- Step 4. *Successful*: In the beginning, the author explains who they are together with the purpose of the email. At the end, the author politely describes what kind of response they expect.
- Step 5. *Mastery*: The author uses clear, detailed and adequate language (grammar and vocabulary); the email is free from mistakes which inhibit understanding.

The theoretical concept underlying the sequence of steps is communicative language teaching (CLT). Broadly speaking, CLT bases language learning on relevant and realistic tasks and holds communicative success to be more important than avoiding mistakes (Savignon, 1991). The aim was to guide students (and assess textual quality) by the principle of increasing communicative value. The level at which a particular feature was placed in the rubric thus depended on communicative importance. The basic and most important element of writing an email, from a communicative point of view, is that writers inquire about all the aspects indicated in the task. Therefore, this element appears as Step 1 in the rubric. In subsequent steps, writers are required to contextualise their inquiries by finding appropriate formulas of salutation and closing (Step 2), formulating a clear and precise subject line (Step 3), and providing a frame that expresses who they are and why they are inquiring about a certain subject (Step 4).

This gradation moves along the key elements of an email and sequences them in a way which allows students to progress from one element to another in a pedagogically meaningful way. For that reason, elements that depend upon formulaic language (such as salutation or closing) appear before elements which require learners to use language more freely (e.g., expressing who they are and what they are writing about). Formulaic elements are well defined, shorter and can be learnt more easily, given the right support.

The decision to place linguistic (formal) aspects of writing as the final step (rather than the first) was equally rooted in a communicative view of writing, which holds that meaningful language is a more powerful concept than correct language (Widdowson, 1978). Further, it is easier for learners to master the specific elements of a genre (which can be explicitly taught and learnt) than to make progress in the general and overarching aspects of foreign language proficiency, such as syntax or lexical quality. While progress in these areas takes months, often years, even lower proficiency learners can progress quickly when being given clear instructions on how to improve their texts by adding or replacing certain semi-standardised elements.

By arranging these elements in sequential order, learners are presented with clear and unambiguous strategies of how to improve their texts. For example, a text that does not

contain all the information required in the task would be graded as 0, directing learners to solve that issue before all others. From a communicative point of view, the reason is obvious: writers who inquire about a summer job but forget to ask about flexible working hours might find themselves unable to take the job even if it is offered to them. Fixing this issue is thus the most important step towards producing a well-functioning text. This design is also based upon empirical research showing that feedback will support student learning and motivation only if it is clear and unambiguous, demonstrating to learners the difference between their current and the intended level of performance without cognitive overload (Kluger & DeNisi, 1996; Shute, 2008). The rubric thus provides learners with an overview of the most important elements of emails, and rubric-based formative assessment provides them with concrete cues for revising their texts.

### *3.2 Implementation of the rubric*

In total,  $N = 373$  eighth grade students at ten lower secondary schools in North-Western Switzerland participated in the study. After the exclusion of those who did not respond to all three writing tasks and those whose texts were not rated (due to technical or human error), a sample of  $n = 338$  students remained. To account for judgement error on the part of the raters, we also excluded 33 students (9.8%) who had received incorrect ratings for one of their texts during the intervention. The final sample included  $n = 305$  students, 160 (52.5%) of which indicated female, 132 (43.3%) indicated male, and 6 (2%) indicated non-binary gender. The mean age was  $M = 14.12$  ( $SD = .87$ ). On average, they had studied English at school for 3 years ( $SD = .66$ ). Data collection was conducted between September and November 2020.

Learners worked in an on-line writing environment on three consecutive email writing tasks. The tasks were based on communicative situations relevant for the email request genre: (a) inquiring about details for a vacation at a campground, (b) inquiring about attending an English course at a language school, and (c) inquiring about a summer job at a restaurant. The research team designed the tasks to be of comparable difficulty. The order of the tasks was randomly assigned, and the writing instructions were similar for each task. The instructions for the restaurant task ("Burger Palace"), for example, read as follows: "You want to make some money during your school holidays and are looking for a job. Read the advertisement you found on the internet and look at the notes you took (in red). Write an email to the store manager in which you introduce yourself and say what you are looking for. Inquire about the information in detail by using your notes in red." The prompt was presented in German so as not to interfere with the writing process in English.



Fig. 2 Writing task Burger Palace as implemented in the study

To support learners in their writing process, the rubric was shown after the first writing task together with the following instructions: “Look at the rubric and learn about key elements of a good email. Reflect what stage you have achieved in your text and how you can use the information in the rubric to improve your text.” This procedure was repeated a second time between writing tasks two and three. One group of learners was shown live ratings from expert raters after each writing task. This variation was introduced to test the effect of rubric-based feedback. However, it is not discussed in detail in this article. The focus here is on the reliability, namely the degree of inter-rater agreement achieved by raters in this study.

Since the learners wrote their texts in English, they could not be expected to produce the language structures to improve their texts without further guidance. For that reason, learners in all groups were shown learning material in the form of a genre-specific framework containing linguistic and structural elements which are relevant for email requests. This gave participants the necessary building blocks to express their requests, possibly improving their subsequent texts.

Schreibrahmen: Wie schreibe ich eine E-Mail auf Englisch?	
Struktur	Satzanfänge/Beispiele
<b>Thema/Inhalt</b> Im <b>Betreff</b> schreibe ich, worum es geht.	Subject: <i>Your visit to Switzerland</i> (Thema: Schüleraustausch)
<b>Anrede</b> Ich benutze <ul style="list-style-type: none"> <li>- für eine Person, zu der ich keine persönliche Beziehung habe, eine <b>formelle</b> Anrede.</li> <li>- für Freunde/Familie eine <b>informelle</b> Anrede.</li> </ul>	<i>Dear Mr/Mrs/Ms ...,</i> (für mit Namen bekannte Person) <i>Dear Sir or Madam</i> (für unbekannte Person)  <i>Hi/Hello/Hey ...</i>
<b>Anfang/Erster Teil</b> Ich <b>stelle mich vor</b> , wenn mich die Person, der ich schreibe, nicht kennt.  Ich erkläre, <b>warum</b> ich schreibe.	<i>My name is ... / I am ...</i>  <i>I am writing to tell you ...</i> <i>I would like to ask you about ...</i>

Fig. 3 Email framework with key information about the target genre (see Appendix B for the full framework)

This input material was provided so that learners would see concrete ways of improving their texts based on the assessment which they received on the basis of the rubric.

## 4 Results

### 4.1 Inter-rater reliability

Texts were rated by six trained research assistants based on the rubric described above. The raters were students of education who all had a good command of the English language. This was tested in a questionnaire and a preliminary interview. They underwent specific training which was administered virtually due to contact restrictions during the Covid-19 pandemic. The rater training was supervised by one of the principal researchers of the project. The training involved three consecutive rounds. In the first round, the raters were instructed to familiarise themselves with the rubric, the three tasks (see Appendix A) and the rating manual (see Appendix C). They were then asked to rate  $n = 26$  texts from a pilot study. The raters were instructed to strictly follow the rubric sequence and to assign a certain level only if the text also fulfilled all preceding criteria. In cases of uncertainty, the raters could note down their questions in a comment column. The raters performed the rating individually and without communicating with the other raters. Kendall's  $W$  was calculated as an indicator for pre-training interrater reliability ( $W = .41$ ).

In round 2, raters were provided with benchmark texts at the different levels which had been compiled by the research team. They worked in groups of three and compared their own ratings with the expert ratings of the benchmark texts. Raters were instructed to pay special attention to differences in the ratings, to discuss possible reasons and to try to understand the ratings of the benchmark texts. They were allowed to take notes

in the rating manual to ensure a common understanding of the rubric and they were given the opportunity to contact the supervisor if necessary. Then they rated six more texts (two per task) for the supervisor to see if the rating reliability had improved and to provide feedback before the final round.

In round 3, raters were instructed to rate 21 texts independently. These ratings were used as the basis to calculate the interrater reliability. Kendall’s W for the six raters in the final round was  $W = .83$ . These results suggest that the raters were able to assess the texts reliably after the training they had received.

During the study, raters assessed student texts and provided live ratings for the learners in the on-line learning environment. After students had completed each writing task, the raters quickly scored the texts according to the rubric. These scores were then given to learners as feedback to guide them in the subsequent writing tasks. As these ratings were done under time pressure, their quality was checked by comparing them to the scores of two expert raters from the research team, who re-scored all texts without time pressure after the intervention at the schools had been completed (post-hoc). Ratings were adjudicated by the team leader where differences occurred to produce a single score for each text. The relationship between live ratings and expert ratings is expressed in Table 1:

**Tab.1:** Live-rating \* expert rating (contingency table)

		Live rating						Total	
		0.00	1.00	2.00	3.00	4.00	5.00		
<b>Expert Rating (ER)</b>	0.00	<i>N</i>	111	23	0	1	1	0	136
		<i>% in ER</i>	81.6%	16.9%	0.0%	0.7%	0.7%	0.0%	100%
	1.00	<i>N</i>	6	523	17	29	14	9	598
		<i>% in ER</i>	1.0%	87.5%	2.8%	4.8%	2.3%	1.5%	100%
	2.00	<i>N</i>	0	5	38	4	7	8	62
		<i>% in ER</i>	0.0%	8.1%	61.3%	6.5%	11.3%	12.9%	100%
	3.00	<i>N</i>	2	1	0	26	3	0	32
		<i>% in ER</i>	6.3%	3.1%	0.0%	81.3%	9.4%	0.0%	100%
	4.00	<i>N</i>	2	5	5	18	52	38	120
		<i>% in ER</i>	1.7%	4.2%	4.2%	15.0%	43.3%	31.7%	100%
	5.00	<i>N</i>	1	4	4	1	14	42	68
		<i>% in ER</i>	1.5%	6.1%	6.1%	1.5%	21.2%	63.6%	100%
<b>Total</b>		<i>N</i>	122	561	64	79	91	97	1014
		<i>% in ER</i>	12.0%	55.3%	6.3%	7.8%	9.0%	9.6%	100%

Note: N = number of learners; % in ER refers to the percentage of ratings that matched the expert ratings.

The contingency table shows the agreement of the live ratings with the (adjudicated) post-hoc ratings ( $\kappa = .66$ ). The extent of agreement differs for the six levels: 80 to 90 percent of all texts on level 0, level 1, and level 3 were identified correctly, about 60 percent on level 2 and level 5, and about 40 percent on level 4. This indicates that some levels are more difficult to judge than others. The differentiation between level 4 (where raters assess whether all the genre-specific elements required in the task are present) and level 5 (where raters assess whether a text is adequate in terms of grammatical, lexical or grammatical quality) seems to be particularly difficult to make when judging texts in a synchronous context under time pressure.

#### 4.2 Comparison of rubric-based scores with linguistic analysis

This second analysis, which relates to research question 2, was undertaken to investigate the validity of the rubric, that is whether the ratings based on the rubric correspond to actual differences of quality in student texts. We conducted a linguistic analysis of four key linguistic qualities of email requests. We focused on four core qualities of emails that learners need to master to be communicatively successful: “salutation”, “information about the writer”, “matter of concern”, and “closing”. “Salutation” and “closing” are relatively general aspects of the email genre, consisting of formulaic language found in a wide range of contexts. These formulae are usually graded as appropriate or inappropriate according to level of formality or status of the addressee (“Dear Mr Black” vs. “Hi James”). They were chosen for detailed linguistic analysis because, on the one hand, mastering them was an important goal of the study, and, on the other hand, they are a key component of any email.

Mastering the features “information about the writer” and “matter of concern” was also an important goal of the learning environment implemented in this study. The writing prompts required learners to introduce themselves and to briefly contextualise the situation provided in task. They were also asked to state the intention of their email (see prompt in Fig. 2 for details). These elements are typical features of emails where the writer and the addressee do not know each other and where the topic of the request is unknown to the addressee. They are especially challenging in terms of register, as they need to be clear and unambiguous without appearing to be pushy or overbearing.

The linguistic analysis of these elements was conducted on the basis of the descriptors shown in Appendix D. All texts of the sample were analysed using the software MAXQDA, by highlighting the relevant passages in each text and coding them according to the descriptors. This procedure is fundamentally different from the rubric-based ratings described in Sect. 4.1, which provides one score (0-5) for the entire text. The linguistic analysis focuses on specific parts of the texts and rates them on a Likert scale from *fully appropriate* to *inappropriate* or *missing*. The resulting scores, thus, link directly to qualities that are particular to a section of text. If these detailed ratings correspond to the scores of the rubric, it would be an indication of validity for the rubric-based ratings.

Raters in this linguistic analysis were four experts in English language teaching from the research team conducting this study. We first developed the codes to describe key aspects of writing quality, and then practised their use on selected texts. The interrater reliability was calculated based on  $n = 40$  texts that were rated by all four raters.

**Tab. 2:** Inter-rater reliability of the linguistic coding of learner texts

Dimension	Kendall's <i>W</i>
Salutation	0.92
Information about the writer	0.96
Matter of concern	0.87
Closing	0.97

Note: 4 raters, *n* = 40 texts

The results in Table 2 show that the reliability of our linguistic rating was high. Having completed the linguistic analysis, we divided our sample into three parts based on the scores they received in the rubric-based ratings. Texts which received overall scores of 0-1 were assigned to the category of low writing quality (*n* = 109), texts graded 2-3 were assigned to the category of medium quality (*n* = 539) and texts grades 4-5 were assigned to the category of high writing quality (*n* = 225). We then compared the linguistic quality of the texts to the rubric-based scores. This procedure is similar to the one proposed by East (2009) to determine the face validity of rubric-based scoring.

**Tab. 3:** Linguistic qualities in texts rated low, medium, and high

		Overall rating (0-5)		
		Low (0-1) ( <i>n</i> = 109)	Medium (2-3) ( <i>n</i> = 539)	High (4-5) ( <i>n</i> = 225)
<b>Salutation</b> (0-3)	M	1.97	2.34	2.90
	SD	0.84	0.79	0.34
	SEM	0.081	0.034	0.023
<b>Information about the writer</b> (0-3)	M	1.94	1.96	2.10
	SD	1.01	0.95	0.85
	SEM	0.097	0.041	0.056
<b>Matter of concern</b> (0-3)	M	1.94	2.29	2.57
	SD	0.85	0.80	0.56
	SEM	0.082	0.034	0.038
<b>Closing</b> (0-2)	M	0.60	0.88	1.88
	SD	0.73	0.82	0.46
	SEM	0.070	0.035	0.030

Note: M: mean; SD: standard deviation; SEM: standard error of the mean; *n*: number of emails.

For each of the four text features, we see differences in text quality between the three groups. These differences are most pronounced for the features salutation and closing, which rely on the use of formulaic expressions. These features are highly important in the rubric, and the students’ attention is directed to them at an early stage. They also feature prominently in the learning material provided in our study. By attending closely to the rubric and the material, learners had the chance to improve their texts in this area in particular. It is therefore an indication of validity that the highly rated texts also show higher linguistic scores. Detailed differences in quality are presented in the examples in Table 4:

**Tab. 4:** Examples of salutations and closings from the two groups with a high and low overall rating, respectively

Overall rating	High (4-5)	Low (0-1)
<b>Salutation</b>	“Dear Mr Black”	“Hi Black”
	“Dear Ms. Robinson”	“Hello CaroLine”
		“Hello my friend Sullivan”
<b>Closing</b>	“Sincerely, Kim Weber”	“Bye bye, Kim”
	“With my best wishes, Kim Weber”	“Big hug, Kim Weber”
		“Love, Kim Weber”

As can be seen in Table 4, in the group of highly rated texts, we can find formulations of salutation and closing that correspond closely to the target genre. In the group of low texts, salutations and closings often appear to be inappropriate in terms of register, incomplete, or missing.

For the dimensions “information about the writer” and “matter of concern” the differences between the texts in the two groups are less pronounced. These features require students to use language more freely, finding their own formulations to express who they are and why they are writing the email. However, when looking at the concrete formulations used in the texts, the groups in our sample show clear differences, suggesting that the rubric-based ratings correspond to actual differences in the quality of genre-based writing. Typical examples for such formulations can be seen in Table 5:

**Tab. 5:** Examples for information about the writer and matter of concern in high vs. low texts

Overall rating	High (4-5)	Low (0-1)
<b>Matter of concern</b>	<p>“I am interested for studying at the Central School, but I have some questions.”</p> <p>“I would like to do a summer job at your place. I saw that you are looking for someone to help out and I would like to. But I have a few questions about the job.”</p> <p>“Yesterday i saw a post in the internet, about a Job at Burger palace. I am interested about the job and have some questions.”</p>	<p>“I saw your Flyer in the Internet.”</p> <p>“I have a few Questions.”</p> <p>“I hope i can go tu this shool.</p>
<b>Information about the writer</b>	<p>“My Name is Kim Weber, I'm 14 years old and i'm from switzerland. I love teamwork and i wish i can have a better english then in my future job i'm going to have it.”</p> <p>“But first I want to introduce myself: My name is Kim Weber and I am fourteen years old.”</p> <p>“My name is Kim Weber. I am 14 years old and am currently living in Switzerland.”</p>	<p>“Hi im Kim.”</p> <p>“Im 14 Years old”</p> <p>“I am very flexible and like to learn new things, also to get better in English.”</p>

In Table 5, we can see that texts with a high overall rating introduce the matter of concern by specifying what is of interest to the writer (e.g., job), their motivation, as well as that the writers have questions they want to pose. Furthermore, in terms of information about the writer, they provide relevant information, such as the name (pseudonym), age, and another piece of information (e.g., location or motivation). Further, texts at this level contain more phrases of appropriate formality such as “I am interested \*for...” or “I would like to...”, which express linguistic politeness in English (Brown & Levinson, 1987).

By contrast, we see that texts with a low overall rating are shorter, less specific, and lack relevant information. In terms of matter of concern, the extracts are statements that lack key information relevant for conveying the intended message to the addressee. As for information about the writer, these extracts show how the writers introduced only one piece of information, such as their name, age, or motivation, without including other relevant aspects. Thus, we can see substantial differences in linguistic quality between the groups with a high and low overall rating, again indicating the validity of the rubric.

## 5 Discussion

The aim of this paper was to document the process of designing and validating an assessment rubric for semi-formal emails in EFL writing. We started with an analysis of the target genre and designed a rubric with high face validity to guide students in the process of writing a good email. Taking a consistency approach, we then showed that raters were able to reach sufficient inter-rater reliability after a period of training. We also showed that ratings by operational raters (provided live while students were working on the tasks) correlated highly with ratings of two expert raters. Our study showed, however, that not all ability levels were equally difficult for raters to judge. Assessing whether a certain element that was required by the genre (e.g., an expressive subject line) is present or not seems to be feasible for raters even under time pressure. By contrast, making a highly inferential assessment concerning the formal quality of a text in terms of grammar, vocabulary, and spelling appears to be much more difficult under time pressure. This indicates that raters should be trained carefully to make qualitative judgments of formal language proficiency. In general, the reliability of our rubric in terms of two trained raters agreeing on a score was high, even under time pressure when decisions had to be made quickly.

Taking an evaluative approach, we further demonstrated that texts rated high, middle, and low according to the rubric were found to be substantially different in terms of linguistic quality. Thus, we conclude that the rubric-based rating measured real linguistic differences in the texts. It quickly and reliably captures the genre-specific qualities and deficits of learner texts that it was designed for, which is an indication of its face validity. Taking a broader look at the concept of validity, as used in assessment research, we could say that the rubric offers a reasonable balance of different types of validity: it focuses on a key element of EFL curricula at lower-secondary level in Europe (curricular validity); it operationalizes the construct of the email request in the details that are relevant for EFL learning (construct validity); and it can be used consistently by raters after a period of specific training (scoring validity; Grotjahn & Kleppin, 2017).

We believe that the process of validating a rubric as shown in this study is relevant for different educational contexts where the outcome of learning is clearly defined. Examples might be doing a presentation, interpreting a historical source (in History) or conducting an experiment in Chemistry or Physics. The approach could also be generalised between different languages across the curriculum: using rubrics could assist learners to extend their knowledge of their first language into a second or third (foreign) language by pointing out similarities or differences between them. When developing and implementing a rubric for a new domain, however, one must consider that a certain amount of training is necessary for sufficient reliability to be achieved. Further, it became clear in our study that not all aspects of a student's performance are equally easy to grade.

As a limitation of this approach, one could argue that teachers rarely participate in rater training, and that reliable or objective feedback is difficult to attain in a classroom context. From the perspective of formative assessment, however, usefulness and timeliness of feedback are more important than objectivity in the psychometric sense

(Moss, 1994). Feedback is one of the most powerful factors influencing student learning. In their meta-analysis, Hattie (2009) found the effect size of different forms of feedback to be  $d = 0.73$ , which shows that it is one of the most powerful factors for successful learning in school contexts. According to the model by Hattie and Timperley (2007), feedback is effective when it addresses three feedback questions: Where am I going? (*feed up*), How am I going? (*feed back*), and Where to next? (*feed forward*). By focusing students' attention closely on key elements of an email, *feed forward* could be implemented with this rubric in the classroom. Working with such a rubric will be a good way of addressing these feedback questions as it is very closely connected to a specific domain of learning.

Further, research has shown that formative rubric use can be a feasible way of harnessing the power of feedback while addressing some of its difficulties – especially constraints on teachers' energy and time (Brookhart & Chen, 2015; Lee, 2005). Feedback is rare in actual classrooms because it is challenging for teachers in terms of time, energy, and motivation (Carless, 2006; Hyland & Hyland, 2006; Lee, 2005). A rubric such as the one presented here might help teachers to provide high-quality feedback to their students while they work, without overtaxing them in terms of time. It can also help teachers to move from mistake correction towards more content-related types of feedback, for example judging a text's effectiveness in a communicative situation. This is particularly important in modern (foreign) language education where language is primarily regarded as a tool of communication, and where formal language skills are seen as necessary to perform communicative tasks (Council of Europe, 2001).

One further concrete application of this rubric lies in automated text evaluation (Deane, 2013). Its categories are well defined and clear-cut. In this study, we created reliable human scores for a corpus of over 1000 texts. Based on these human scores, techniques of natural language processing (NLP) can be used to generate automated assessments of the different email features. This research group is currently developing such an automated assessment system in an interdisciplinary co-operation of EFL methodology specialists and computational linguists (see Meyer et al., 2021). Such technology holds great promise to make assessment of learning outcomes more reliable, and to provide timely feedback to students. Preliminary results suggest that an NLP system can predict (or imitate) the human ratings on our rubric with an 80% accuracy (Meyer et al., 2021). This software can give learners instant feedback on key aspects of emails while they are writing, and it can assist teachers in assessing the outcomes in a fair and objective way. We are confident, therefore, that the reliability and validity of the rubric laid out in this article can lead to new developments in automated feedback and assessment for English emails, supporting teachers and learners in this key domain of the EFL writing and laying the foundation for adapting such tools to an increasing number of learning contexts.

### Acknowledgement

This work was supported by the Jacobs Foundation Switzerland.

## References

- Al-Ali, M., & Sahawneh, M. (2008). An investigation into the generic features of English requestive e-mail messages. *LSP & Professional Communication*, 8(2), 40-64.
- Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sciences Education*, 5(3), 197-203. <https://doi.org/10.1187/cbe.06-06-0168>
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-19.
- Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, 4(4), 1-22. <https://cie.asu.edu/ojs/index.php/cieatasu/article/view/1630/665>
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-31. <https://doi.org/10.3200/CTCH.53.1.27-31>
- Andrade, H., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102(4), 287-302. <https://doi.org/10.3200/JOER.102.4.287-302>
- Bhatia, V. (2020). Genres and styles in world Englishes. In C. Nelson, Z. Proshina & D. Davis (Eds.), *Handbook of World Englishes* (pp. 368-382). Wiley-Blackwell.
- British Council (2022). *LearnEnglish: An email request*. Online: <https://learnenglish.britishcouncil.org/skills/writing/c1-writing/an-email-request>
- Brookhart, S. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brookhart S. & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343-368. <https://doi.org/10.1080/00131911.2014.929565>
- Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge University Press
- Campbell, K. (2013, March 7). *The genre of email requests*. Retrieved from <https://proswrite.com/author/proswrite/>
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education* 31(2), 219-233. <https://doi.org/10.1080/03075070600572132>
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Dawson, P. (2015). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360. <https://doi.org/10.1080/02602938.2015.1111294>.
- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18, 7-24. <https://doi.org/10.1016/j.asw.2012.10.002>
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14(2), 88-115. <https://doi.org/10.1016/j.asw.2009.04.001>
- Ene, E. & Upton, T. (2018). Synchronous and asynchronous teacher electronic feedback and learner uptake in EFL composition. *Journal of Second Language Writing*, 41, 1-13. <https://doi.org/10.1016/j.jslw.2018.05.005>
- Gains, J. (1999). Electronic mail – A new style of communication or just a new medium? An investigation into the text features of e-mail. *English for Specific Purposes*, 18, 81-101. [https://doi.org/10.1016/S0889-4906\(97\)00051-3](https://doi.org/10.1016/S0889-4906(97)00051-3)
- Greatorex, J., & Bell, J. F. (2008). What makes AS marking reliable? An experiment with some stages from the standardisation process. *Research Papers in Education*, 23(3), 333-355. <https://doi.org/10.1080/02671520701692593>
- Grotjahn, R., & Kleppin, K. (2017). Gütekriterien bei der Evaluation von Schreibkompetenzen. In B. Akukwe, R. Grotjahn, & S. Schipolowski (Eds.), *Schreibkompetenzen in der Fremdsprache. Aufgabengestaltung, kriterienorientierte Bewertung und Feedback* (pp. 41-70). Narr Francke Attempto.
- Hallet, W. (2016). *Genres im fremdsprachlichen und bilingualen Unterricht*. Hannover: Klett/Kallmeyer.

- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.  
<https://doi.org/10.3102/003465430298487>
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148-164.  
<https://doi.org/10.1016/j.jslw.2007.07.005>
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83-101.  
<https://doi.org/10.1017/S0261444806003399>
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.  
<https://doi.org/10.1016/j.edurev.2007.05.002>
- KMK (2004). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss*. [Educational Standards for the first foreign language (English/French) for middle school]. Kultusministerkonferenz [Conference of Educational Ministers in Germany].
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.  
<https://doi.org/10.1037/0033-2909.119.2.254>
- Lee, I. (2005). Why burn the midnight oil? Marking student essays. *Modern English Teacher*, 14(1), 33-40.
- Meyer, J., Jansen, T., Fleckenstein, J., Keller, S., & Köller, O. (2021). Machine learning im Bildungskontext: Evidenz für die Genauigkeit der automatisierten Beurteilung von Essays im Fach Englisch. *Zeitschrift für Pädagogische Psychologie*, 12 (2020).  
<https://doi.org/10.1024/1010-0652/a000296>
- Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, 23 (2), 5-12.  
<https://doi.org/10.3102/0013189X023002005>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144.  
<https://doi.org/10.1016/j.edurev.2013.01.002>
- Reddy, Y, & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.  
<https://doi.org/10.1080/02602930902862859>
- Savignon, S. J. (1991). Communicative language teaching: State of the art. *TESOL Quarterly*, 25(2), 261-277.  
<https://doi.org/10.2307/3587463>
- Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88.  
<https://doi.org/10.1016/j.asw.2014.03.003>
- EDK (2011). Lehrplan 21 [Curriculum for compulsory public schools]. Swiss Conference of Cantonal Ministers of Education. Retrieved from: <https://www.lehrplan21.ch/>
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. ERPI.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.  
<https://doi.org/10.3102/0034654307313795>
- UNESCO (2011). *International standard classification of education (ISCED) 2011*. United Nations Educational, Scientific and Cultural Organization. Retrieved from [www.unesco.org](http://www.unesco.org)
- VERBI Software (2021). *MAXQDA 2022* [computer software]. VERBI Software.  
<https://www.maxqda.com>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford University Press.

## Appendix A

### Writing Task: Burger Palace

Du möchtest während der Schulferien gerne etwas dazuverdienen und suchst daher nach einem Ferienjob. Lies das Inserat, das du im Internet gefunden hast und die Notizen, die du dir bereits in rot dazu gemacht hast. Schreibe eine offizielle E-Mail («formal email») an die Leiterin, in der Du Dich kurz vorstellst und Deine Anliegen detailliert schilderst. Verwende dazu **alle gemachten Notizen** (unten). Benutze keine anderen Hilfsmittel.

[You would like to earn some extra money during the school vacations and are therefore looking for a vacation job. Read the ad you found on the Internet and the notes you already made in red. Write an official email ("formal email") to the leader, in which you introduce yourself briefly and describe your concerns in detail. **Use all the notes you made** (below). Do not use any other tools.]

## Appendix B

### Writing guidelines

#### Schreibrahmen: Wie schreibe ich eine E-Mail auf Englisch?

Struktur	Satzanfänge/Beispiele
<b>Thema/Inhalt</b> Im <b>Betreff</b> schreibe ich, worum es geht.	Subject: <i>Your visit to Switzerland</i> (Thema: Schüleraustausch)
<b>Anrede</b> Ich benutze <ul style="list-style-type: none"> <li>- für eine Person, zu der ich keine persönliche Beziehung habe, eine <b>formelle</b> Anrede.</li> <li>- für Freunde/Familie eine <b>informelle</b> Anrede.</li> </ul>	<i>Dear Mr./Mrs/Ms ...,</i> (für mit Namen bekannte Person) <i>Dear Sir or Madam</i> (für unbekannte Person)  <i>Hi/Hello/Hey ...</i>
<b>Anfang/Erster Teil</b> Ich <b>stelle mich vor</b> , wenn mich die Person, der ich schreibe, nicht kennt. Ich erkläre, <b>warum</b> ich schreibe.	<i>My name is ... / I am ...</i>  <i>I am writing to tell you ...</i> <i>I would like to ask you about ...</i>
<b>Hauptteil</b> Ich schreibe alle wichtigen Punkte, die in der Aufgabe gefordert sind. Ich achte darauf, dass die Person, der ich schreibe, meine Sätze gut verstehen kann. Ich verwende <ul style="list-style-type: none"> <li>- <b>formale</b> Sprache für eine Person, die ich nicht (gut) kenne.</li> <li>- <b>Informelle</b> Sprache für Freunde/Familie</li> </ul>	<b>Mögliche Satzanfänge:</b> <i>First I want to ... / Second ...</i> <i>It's interesting/great/sad that ...</i> <i>I (don't) like ...</i> <i>I would like to ask/suggest ...</i> <b>Bindewörter:</b> <i>After ... / Then ... / Finally ... / As ...</i> <i>Besides ... / In addition ...</i> <i>That's why ... / ... because ...</i>
<b>Schlusssatz</b> Ich beende meine E-Mail mit einem Schlusssatz, der ausdrückt, welche Reaktion ich erwarte.	<i>I look forward to hearing from you. /</i> <i>Thanks for answering my questions.</i>
<b>Verabschiedung</b> Ich benutze <ul style="list-style-type: none"> <li>- für eine Person, zu der ich keine persönliche Beziehung habe, eine <b>formelle</b> Schlussformel und meinen <b>Vor- und Nachnamen</b>.</li> <li>- für Freunde/Familie eine <b>informelle</b> Schlussformel und meinen <b>Vornamen</b>.</li> </ul>	Best wishes, ... / (Yours) sincerely, ...  Bye, ... / Love, ... / See you soon, ...

## Appendix C

### Rater manual

**Level 1:** Die E-Mail enthält alle in der Aufgabe geforderten Informationen.

<p>Jede der <b>drei geforderten Informationen</b> muss <b>erkennbar erfragt</b> werden, z.B.:</p>	
<ul style="list-style-type: none"> <li>- Can I...</li> <li>- What is...</li> <li>- How much...</li> <li>- Do you...</li> </ul>	<ul style="list-style-type: none"> <li>- How much is the salary?</li> <li>- Working in the kitchen?</li> <li>- Weekends?</li> </ul>
<p><b>Note:</b> - Zusätzliche Fragen/Informationen sind in Ordnung.</p> <ul style="list-style-type: none"> <li>- Grammatische Fehler oder Zeichensetzung sind nicht relevant.</li> <li>- Mittendrin abgebrochene Sätze werden als fehlende Information gewertet.</li> <li>- Bei Unsicherheit ob Level 0 oder 1 → Level 1</li> </ul>	

**Level 2:** Begrüssung und Verabschiedung sind der in der Aufgabe beschriebenen Situation angemessen.

<b>Angemessene Begrüssungen</b> (mit Ms/Mrs/Miss/(Caroline) Robinson):	<i>Nicht ausreichend wäre:</i>
<ul style="list-style-type: none"> <li>- Dear</li> <li>- Good morning</li> <li>- Good afternoon</li> </ul>	<ul style="list-style-type: none"> <li>- Dear Sir/Madam</li> <li>- Hi/Hello (with or without name)</li> <li>- Good morning/good afternoon (without name)</li> <li>- Caroline (Robinson)</li> <li>- Falsches Geschlecht: Dear Mr. Robinson</li> </ul>
<b>Angemessene Verabschiedungen</b> (mit Vor- und Nachnamen):	<i>Nicht ausreichend wäre:</i>
<ul style="list-style-type: none"> <li>- Sincerely</li> <li>- Yours sincerely</li> <li>- Best (wishes, regards)</li> <li>- Yours/your</li> <li>- Greetings (from)</li> <li>- (I wish you) all the best</li> </ul>	<ul style="list-style-type: none"> <li>- From</li> <li>- Kim (Weber)</li> </ul>
<b>Note:</b> Zeichensetzung ist nicht relevant.	

**Level 3:** Der Betreff beschreibt, worum es dir in der E-Mail geht.

Es kommt mindestens <b>einer der folgenden Begriffe</b> vor:	<i>Nicht ausreichend wäre:</i>
<ul style="list-style-type: none"> <li>- Summer job</li> <li>- Holiday job</li> <li>- Vacation job</li> <li>- Job</li> </ul>	<ul style="list-style-type: none"> <li>- Questions (about the flyer)</li> <li>- Burger Palace</li> </ul>
<b>Note:</b> Gross- und Kleinschreibung ist nicht relevant.	

**Level 4:** Am Anfang der E-Mail erklärst du, wer du bist und warum du schreibst; am Ende der E-Mail erklärst du höflich, welche Reaktion du erwartest.

<p>Der <b>einleitende Satz</b> sollte Informationen zur Person und ihrem Anliegen enthalten:</p>	<p><i>Nicht ausreichend wäre:</i></p>
<ul style="list-style-type: none"> <li>- I am a student from Switzerland and I have some questions about the job at your restaurant...</li> <li>- I saw your flyer and I have some questions...</li> </ul>	<ul style="list-style-type: none"> <li>- I have a question.</li> <li>- I have many/three question(s).</li> </ul>
<p>Der <b>abschliessende Satz</b> sollte ausdrücken, dass eine Antwort erwartet wird:</p>	<p><i>Nicht ausreichend wäre:</i></p>
<ul style="list-style-type: none"> <li>- I'm looking forward to hearing from you.</li> <li>- It would be great if you could answer my questions.</li> <li>- Thank you for your help.</li> <li>- Please send me more information.</li> <li>- Please let me know about ...</li> <li>- I hope you can help me.</li> </ul>	<ul style="list-style-type: none"> <li>- Have a good day. (alone)</li> <li>- Answer me. (alone)</li> <li>- Thanks/thank you (for your time).</li> </ul>
<p><b>Note:</b> - Sprachliche Richtigkeit ist nicht relevant.</p> <ul style="list-style-type: none"> <li>- Formulierung anderer Erwartungen (z.B. Let me know if I get the job) werden als falsch gewertet.</li> </ul>	

**Level 5:** Du verwendest einen Sprachstil, der bezüglich Grammatik und Wortschatz einer offiziellen E-Mail angemessen ist.

<b>Der Sprachstil ist einer offiziellen E-Mail angemessen:</b>
<ul style="list-style-type: none"> <li>- Einsatz von thematisch angemessenem differenziertem <b>Vokabular</b>, um Details abzufragen</li> <li>- Klarer und verständliche <b>Grammatik</b></li> <li>- Verwendung von <b>idiomatischen Wendungen</b>, z.B.:               <ul style="list-style-type: none"> <li>○ I look forward to hearing from you</li> </ul> </li> <li>- <b>Textstruktierende Elemente</b>, z.B.:               <ul style="list-style-type: none"> <li>○ Aufzählung (first, second, third)</li> </ul> </li> <li>- Einsatz von grammatischen Konstruktionen zur Herstellung von <b>Höflichkeit</b>, z.B.:               <ul style="list-style-type: none"> <li>○ Could you tell me...</li> <li>○ I was wondering...</li> <li>○ I would like to know...</li> <li>○ Can/could I...</li> <li>○ Would it be possible...</li> </ul> </li> </ul>
<p><b>Note:</b> - Rechtschreibung wird honoriert, aber nicht notwendigerweise sanktioniert. Bei ansonsten hohem sprachlichen Niveau (Grammatik, Struktur, Vokabular) werden Rechtschreib-/Tippfehler nicht sanktioniert, sofern sie die Lesbarkeit nicht erheblich stören.</p> <ul style="list-style-type: none"> <li>- Sprachliche Wagnisse werden honoriert, auch bei fehlerhafter Grammatik.</li> </ul>

## Appendix D

### Descriptors for linguistic analysis of text qualities

#### Salutation

	Level	Code	Description	Example
<b>Salutation</b> (0-3)	3	Fully appropriate	A salutation is provided, and it is appropriate and respectful.	“Dear Ms Black”
	2	Partly appropriate	A salutation is provided, but it is inappropriate for a formal email.	“Dear Sir”
	1	Inappropriate	A salutation is provided, but it is inappropriate for a formal email.	“Hi”
	0	Missing	A salutation is missing.	-

#### Information about the writer

	Level	Code	Description	Example
<b>Information about the writer</b> (0-3)	3	Fully appropriate	Information about the writer is provided and appropriately expanded upon. Must include one piece of information in addition to the name that is context relevant.	"My name is Kim Weber and I'm 16 years old."
	2	Partly appropriate	Information about the writer is provided and is sufficient for a formal email. Additional details are superfluous in the introduction.	"My name is Kim Weber."
	1	Inappropriate	Information about the writer is only partially provided, with no name given.	"I am 16 and come from Switzerland."
	0	Missing	No information about the writer is provided.	-

Matter of concern

	Level	Code	Description	Example
<b>Matter of concern</b> (0-3)	3	Fully appropriate	The information about the matter of concern is provided and appropriately expanded upon. Includes all three of the pieces of information:  1) Mentions one's own interest; 2) Is specific;  3) Mentions the existence of questions.	"I think about learning English at the Central School. First I Need some closer information."
	2	Partly appropriate	The information about the matter of concern is provided. Mentions two of the three pieces of information specified.	"I have a question about the Central School."
	1	Inappropriate	The information about the matter of concern is insufficient. Mentions only one of the three pieces of information specified.	"I want to learn English."
	0	Missing	The information about the matter of concern is missing.	-

Closing

	Level	Code	Description	Example
<b>Closing</b> (0-2)	2	Appropriate and respectful	A closing is provided. It is appropriate and respectful.	“Best wishes”
	1	Partly appropriate	A closing is provided, but it is inappropriate for a formal email.	“Love”
	0	Missing	A closing is missing, in German, or just the name.	“Kim Weber”

## Tables

**Table 1**

*Live-rating \* expert rating (contingency table)*

		Live rating							
		0.00	1.00	2.00	3.00	4.00	5.00	Total	
<b>Expert Rating (ER)</b>	0.00	<i>N</i>	111	23	0	1	1	0	136
		<i>% in ER</i>	81.6%	16.9%	0.0%	0.7%	0.7%	0.0%	100%
	1.00	<i>N</i>	6	523	17	29	14	9	598
		<i>% in ER</i>	1.0%	87.5%	2.8%	4.8%	2.3%	1.5%	100%
	2.00	<i>N</i>	0	5	38	4	7	8	62
		<i>% in ER</i>	0.0%	8.1%	61.3%	6.5%	11.3%	12.9%	100%
	3.00	<i>N</i>	2	1	0	26	3	0	32
		<i>% in ER</i>	6.3%	3.1%	0.0%	81.3%	9.4%	0.0%	100%
	4.00	<i>N</i>	2	5	5	18	52	38	120
		<i>% in ER</i>	1.7%	4.2%	4.2%	15.0%	43.3%	31.7%	100%
	5.00	<i>N</i>	1	4	4	1	14	42	68
		<i>% in ER</i>	1.5%	6.1%	6.1%	1.5%	21.2%	63.6%	100%
<b>Total</b>		<i>N</i>	122	561	64	79	91	97	1014
		<i>% in ER</i>	12.0%	55.3%	6.3%	7.8%	9.0%	9.6%	100%

Note: N = number of learners; % in ER refers to the percentage of ratings in fell into the expert ratings.

**Table 2**

*Inter-rater reliability of the linguistic coding of learner texts*

Dimension	Kendall's <i>W</i>
Salutation	0.92
Information about the writer	0.96
Matter of concern	0.87
Closing	0.97

Note: 4 raters, *n* = 40 texts

**Table 3**

*Linguistic qualities in texts rated low, medium and high*

		Overall rating (0-5)		
		Low (0-1) (n = 109)	Medium (2-3) (n = 539)	High (4-5) (n = 225)
<b>Salutation</b> (0-3)	M	1.97	2.34	2.90
	SD	0.84	0.79	0.34
	SEM	0.081	0.034	0.023
<b>Information about the writer</b> (0-3)	M	1.94	1.96	2.10
	SD	1.01	0.95	0.85
	SEM	0.097	0.041	0.056
<b>Matter of concern</b> (0-3)	M	1.94	2.29	2.57
	SD	0.85	0.80	0.56
	SEM	0.082	0.034	0.038
<b>Closing</b> (0-2)	M	0.60	0.88	1.88
	SD	0.73	0.82	0.46
	SEM	0.070	0.035	0.030

Note: M: mean; SD: standard deviation; SEM: standard error of the mean; *n*: number of emails.

**Table 4**

*Examples of salutations and closings from the two groups with a high and low overall rating, respectively*

Overall rating	High (4-5)	Low (0-1)
<b>Salutation</b>	“Dear Mr Black”	“Hi Black”
	“Dear Ms. Robinson”	“Hello CaroLine”
		“Hello my friend Sullivan”
<b>Closing</b>	“Sincerely, Kim Weber”	“Bye bye, Kim”
	“With my best wishes, Kim Weber”	“Big hug, Kim Weber”
		“Love, Kim Weber”

**Table 5**

*Examples for information about the writer and matter of concern in high vs. low texts*

Overall rating	High (4-5)	Low (0-1)
<b>Matter of concern</b>	<p>“I am interested for studying at the Central School, but I have some questions.”</p> <p>“I would like to do a summer job at your place. I saw that you are looking for someone to help out and I would like to. But I have a few questions about the job.”</p> <p>“Yesterday i saw a post in the internet, about a Job at Burger palace. I am interested about the job and have some questions.”</p>	<p>“I saw your Flyer in the Internet.”</p> <p>“I have a few Questions.”</p> <p>“I hope i can go tu this shooh.</p>
<b>Information about the writer</b>	<p>“My Name is Kim Weber, I'm 14 years old and i'm from switzerland. I love teamwork and i wish i can have a better english then in my future job i'm going to have it.”</p> <p>“But first I want to introduce myself: My name is Kim Weber and I am fourteen years old.”</p> <p>“My name is Kim Weber. I am 14 years old and am currently living in Switzerland.”</p>	<p>“Hi im Kim.”</p> <p>“Im 14 Years old”</p> <p>“I am very flexible and like to learn new things, also to get better in English.”</p>

**Figure Captions**

**Figure 1:** *Genre-based rubric for writing emails in English*

**Figure 2:** *Writing task “Burger Palace” as implemented in the study*

**Figure 3:** *Email framework with key information about the target genre (see Appendix B for the full framework)*