Using the CEFR in Taiwan: 
the perspective of a local examination board

Jessica R. W. Wu & Rachel Y. F. Wu
The Language Training and Testing Center

Abstract

English learning and assessment in Taiwan is undergoing a critical change, particularly in relation to the establishment of a common standard of English proficiency through the adoption of the CEFR framework by the Ministry of Education starting 2005, and in the meantime, there is a pressing need to solve the problems that have occurred as consequences of this change. Firstly, this article will review how English language tests have been applied in such a context. Secondly, it will describe the new challenges and responsibilities faced by a local examination board, primarily improvement of stakeholders’ understanding of this change. Thirdly, the paper will report on the experience of the Examination Board of the GEPT (General English Proficiency Test), a high-stakes testing system established with the support of the government in 2000, concerning the alignment between the GEPT and the CEFR framework by following the Manual for Relating Examinations to the CEFR. As the CEFR framework is a rather new input to the Taiwanese EFL context, by presenting our observation of and novice experience in using the CEFR from the perspective of a local examination board, we hope to provide some input to the revision of the CEFR, and we also look forward to insights from the colleagues working in the European nations where the CEFR is more commonly applied.

I Taiwanese context

1. The use of English language tests

With an upgrade in Taiwan’s economic development and the shift to a more international outlook, there has been a strong identification in recent years of the need to acquire competency in English, and this interest is supported by government policies concerning the use of English language assessment.

In May 2002, as part of its efforts to aid in enhancing economic competitiveness for Taiwan, the government launched the Challenge 2008 National Development Plan, which emphasized such goals as strengthening international competitiveness, upgrading the quality of life, and promoting sustainable development. Under Challenge 2008, one of the projects is the cultivation of talent for the ‘e-generation’, recognizing that a skilled workforce is the basis of all development.

To meet the project’s goal, the government has been actively seeking to expand the use of English as a part of daily life. In 2005, the Ministry of Education (MOE) adopted the
CEFR, Common European Framework of Reference for Languages: learning, teaching, assessment (Council of Europe, 2001) as its source for the establishment of target levels of English ability for EFL learners in Taiwan. Following that move, the government created a list of English proficiency tests available in Taiwan which were mapped against the CEFR levels to assist score users in choosing a test that they considered appropriate for their needs (Central Personnel Administration, 2006). In the list of the tests, there are general English proficiency tests (e.g., Cambridge Main Suite and GEPT); tests for academic purposes (e.g., IELTS and TOEFL); and tests for workplace English (e.g., BULATS and TOEIC).

Currently in Taiwan, a score on an external English test, an achievement on an internal test of English, and an achievement under conditions of classroom assessment are influential in qualifying one for school graduation or for a job promotion. The following are some examples.

- **English ability of college and university students**
The MOE has encouraged universities and colleges to establish a regulation requiring that all students attending their institutions achieve a pass in a test of English prior to graduation. The number of students who achieve a passing level in a test will be taken into account in evaluating the quality of a college or university. According to the MOE’s most recent projection (MOE, 2006), it is expected that 50% of university graduates will achieve scores equivalent to CEFR-B1 Level (Threshold) or above in an English language test before the end of 2007; and for technological and vocational colleges, a total of 50% of the graduates are expected to demonstrate a minimum proficiency in English at CEFR-A2 Level (Waystage).

- **English ability of teachers**
Like students, teachers at all levels of the educational system are urged to take an English test. It is anticipated that, by the end of 2007, 30% of all the teachers in elementary and secondary schools will have achieved a score on a test of English equivalent to the CEFR B1 Level and 45% of the teachers of English in elementary and secondary schools will have achieved a score equivalent to the CEFR B2 Level (Vantage).

- **English ability of government employees**
Government employees are also required to demonstrate a minimum proficiency in English at CEFR-A2 Level. Those with a better command of English are given enhanced prospects for promotion. The government estimates that 18% of civil servants under the age of 40 will reach the minimum standard before the end of 2007.

Government support and developing social trends have resulted in an enormous increase in the population of English language test-takers in Taiwan. In 2006, the number of test-takers sitting for the four most popular English tests, GEPT, IELTS, TOEFL, and TOEIC, reached a record high of 600,000. Among them, the GEPT was the most widely used, which accounted for over 80% of the total number of test-takers (approximately 500,000).

### 2. The General English Proficiency Test (GEPT)
The GEPT is a five-level criterion-referenced EFL testing system implemented in Taiwan to
assess the general English proficiency of EFL learners. In 1999, the Ministry of Education lent its support to The Language Training and Testing Center (LTTC) in its development of the GEPT. The aim of the GEPT is to promote the concept of life-long learning and to encourage the use of the communicative approach in English teaching and learning. The test was created in response to comments from educators and employers about the general lack of ability to communicate in English among Taiwanese English learners, partly due to the previous ‘old-fashioned’ approach to English teaching, which over-emphasized the importance of grammatical accuracy.

Each level of the GEPT consists of four components of assessment: listening, reading, writing, and speaking. The GEPT is being used by various government institutions and schools for entry, classroom achievement, and graduation requirements. So far, more than two million EFL learners in Taiwan have taken the GEPT since its first administration in 2000.

A number of studies related to the GEPT have been conducted by the LTTC on parallel-form reliability (Weir and Wu, 2002), concurrent validity of the GEPT Intermediate and High-Intermediate (LTTC, 2003), mapping the GEPT to the Common English Yardstick for English Education in Taiwan (LTTC, 2005), written language of test-takers (Kuo, 2005), on test impact (Wu and Chin, 2006; Wu, 2007), and test form and individual task comparability (Weir and Wu, 2006).

3. The use of the CEFR in Taiwan
As noted earlier, the MOE adopted the CEFR with the aim of using it as a yardstick to inform assessment of learners’ proficiency in English. This move has indeed created a new context for English language learning and assessment in Taiwan.

As background to explain the idea of locating different tests in relation to the CEFR framework, Taiwan is a very competitive free-market society and people expect to be offered choices. In addition, the government, for various reasons, does not feel it is appropriate for them to choose one test or kind of test over another. Under such circumstances, a positive aspect is that Taiwanese EFL learners and/or score users are free, to some degree, to make their own choices as to which test to take; however, as a natural consequence, this leads to a pressing need for score equivalence, i.e., knowledge of how the scores from two different tests relate to one another and to what extent they can be considered equivalent. Therefore, the MOE thought that the CEFR could function as a common basis on which the comparison between different tests of English could be made and could thus meet the demand for score comparability. Also, by means of the CEFR, the MOE intended to require test developers to relate their examinations to a single scheme so that more interpretable and meaningful score results can be offered.

Despite the good intentions, the reaction to the use of the CEFR is mixed due to problems that have arisen, primarily in the area of test comparability. However, the introduction of the CEFR may stimulate positive developments in EFL education in Taiwan, which is definitely one of the new challenges and responsibilities facing the LTTC and other local professionals in language testing and teaching.
Test comparability
A number of language testers have addressed concerns about the issue of test comparability. For example, Bachman (1995) remarks that any comparability study needs to take account of more than just score equivalences, and other aspects such as test content and performance must also be investigated. In line with this, Davies et al. (1999) and Taylor (2004) also suggest that the concept of test comparison is problematic, because each test is designed for a different purpose and a different population.

Acknowledging the concern noted above, at the time the MOE was creating the ‘table of approximate score comparability’ for the purpose of comparing the English language tests, the LTTC alerted them to the potential misuses of the ‘table’. One of the misuses we think most likely is that test users may focus on the notion of ‘score equivalence’ only, rather than also paying adequate attention to the variations in the features and constructs between different tests. When the test construct is ignored, test users may choose to use a less difficult test (e.g., no speaking or writing component; using all multiple-choice items) to achieve a ‘passing’ score more easily. Our warning messages were accepted by the MOE and later were included in the ‘table’. However, since the launch of the ‘table’, the warning messages have been disregarded by the test users and the problems associated with the misuses of the ‘table’ have arisen as predicted.

Currently, in the context of Taiwan, comparing test scores via the CEFR framework is not only a political mandate issue, but also a marketing issue. As long as an exam provider can make a claim for alignment to the CEFR, even without providing theoretical and empirical evidence to build such a claim, their test can still survive and even prosper in the market. Sadly, we have to note that the concept of a ‘CEFR-aligned’ test has been mistakenly interpreted in our country.

Having noted the problems, however, we believe that it does not mean that the CEFR may not play a useful role in Taiwan. We believe the adoption of the CEFR has the potential to improve language teaching and testing in Taiwan but misuse, such as using the Framework ‘to compare tests by mediation’ as Fulcher notes (2004) must be avoided. Therefore, as a local exam board that develops and administers high-stakes language tests, the LTTC recognizes an obligation to educate our stakeholders about the correct meaning of the term ‘CEFR-aligned’ tests. For this purpose, in July 2005 the LTTC decided to launch a research project: Mapping the GEPT onto the CEFR. Furthermore, in order to carry out the mapping project appropriately, the LTTC officially registered with the Council of Europe (CoE) to participate in the project for piloting the manual for relating language examinations to the CEFR.

New challenges and responsibilities
As the full title of the CEFR shows, it is a framework not only for assessment but also for learning and teaching. Although for the present the government in Taiwan seems to have adopted the CEFR only as a new standard for assessment, its potential to affect the way
learning and teaching are viewed and implemented in Taiwan should not be ignored. As in many Asian countries, there is a very exam-oriented culture in Taiwan, and exams, particularly high-stakes ones, such as the GEPT, can have a significant impact on teaching and learning (Wu & Chin, 2006). Therefore, by relating the GEPT to the CEFR, which may make GEPT scores more interpretable and transparent, we will be able to increase teachers’ and learners’ understanding of what ‘communicative language ability’ really means. In other words, provided teachers can understand how a GEPT test-taker with a particular score is able to do in terms of the criteria specified in the CEFR framework, they will gradually consolidate the theoretical criteria with their teaching and assessment practices employed in the classroom.

The article will now turn to the report on the details of the GEPT-CEFR mapping project undertaken by the LTTC.

II Mapping the GEPT onto the CEFR
This project to relate the GEPT to the CEFR followed the ‘internal validation’ procedure, including Familiarisation, Specification and Standardisation (Judgment session only), proposed by the Manual with the intention of re-examining the General English Proficiency Test (GEPT) to ensure that its test development and administrative procedures conform to the internationally accepted code of practice, and also to present various features of the GEPT to its stakeholders in a more comprehensive way.

The LTTC participated in the Council of Europe’s project for Piloting the Manual for Relating Language Examinations to the CEFR, and the feedback report of the study was sent to the CoE in September 2006. During the study, the LTTC focused on piloting the Specification process, and it was found that the Specification provides a well covered outline which includes major features of language examinations, and the Forms which the Manual provides not only are useful tools for re-examination of language tests, but they also facilitate detailed reports on the quality of the tests. The LTTC, therefore, decided to employ the procedure in the Manual to make further attempt to link the GEPT to the CEFR. Due to time and logistics constraints, this project covered only the GEPT Reading Comprehension Tests for the present stage.

- **Familiarisation**
To familiarize those who would be involved in the Specification and Standardisation procedures with the CEFR, the Familiarisation process was conducted in January 2007. A total of 20 people who are experienced professionals in English teaching or testing participated in the familiarisation session. The scheme included the following procedures:
1. Chapters 1, 2 and 3 were distributed to the participants for self-study before they attended the two-stage familiarisation session.
2. The first-stage familiarisation session began with a 30-minute presentation on the CEFR,
followed by small group discussions focusing on the questions presented at the end of the subsections ‘How to use scales of descriptors of language proficiency’ (p. 40), ‘Conditions and constraints’ (p. 50), ‘Communication themes/communicative tasks and purposes’ (pp. 53 and 54), ‘Productive activities’ (pp. 61 and 63) and ‘Receptive activities’ (pp. 68 and 71), in the CEFR. After the discussions, the participants tried to relate the GEPT to the CEFR using Table 1, the global descriptors of the CEFR. Then, the participants self-assessed their English proficiency level using Table 2 (Self-assessment grid) in the CEFR as a wrap-up activity.

3. In the second-stage familiarisation session, the participants sorted slips of individual proficiency statements from lower level to higher level and then compared their results to the CEFR level descriptors; various scales of reading comprehension descriptors in the CEFR, DIALANG self-assessment statements, and self-assessment statements in CEFR Table 2 were reconstructed in the activities.

The discussion results showed that based on the participants’ estimation, there was global agreement on the rationale for relating the four levels of the GEPT to the CEFR levels: the GEPT Elementary Level is roughly equivalent to CEFR A2, Intermediate Level to CEFR B1, High-Intermediate Level to CEFR B2, and Advanced Level to CEFR C1-C2.

- **Specification**

CEFR is a useful instrument for re-examination of language tests and the Manual facilitates a detailed report on the quality of the tests. However, based on our experience gained from the CoE’s Project of Piloting the Manual, the qualitative analysis procedures provided in the Manual seem to include only administrative procedures and text-level specifications; specifications of item-level comprehension operations, which should be equally important when test constructs are examined and compared, are overlooked (Alderson et al, 2004, p. 44). Therefore, we applied the Dutch Grid to analyze different levels of the GEPT Reading Comprehension Tests, in the hope of reflecting and differentiating the difficulty levels of the GEPT Reading Comprehension Tests in terms of the CEFR levels.

1. **Method**

The first four levels of the 2006 GEPT Reading Comprehension Tests (excepting the Superior Level, since the Superior Level GEPT was not administered in 2006) have been analyzed mainly with the help of the Dutch Grid. Although the content analyses only covered the most recent year of the test forms and therefore may not present a complete picture of the GEPT Reading Comprehension Tests, we believe the results should be able to exemplify the test constructs of the different levels of the GEPT Reading Comprehension Tests.
For each level of the GEPT Reading Comprehension Tests, the analysis has been carried out by three judges; each level was analyzed by a different pair of judges, and the other person double-checked the results of the analyses for all four levels.

Table 1: Number of texts and items analyzed per GEPT level

<table>
<thead>
<tr>
<th>GEPT</th>
<th>No. of Texts</th>
<th>No. of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary Level</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>Intermediate Level</td>
<td>14</td>
<td>50</td>
</tr>
<tr>
<td>High-Intermediate Level</td>
<td>16</td>
<td>70</td>
</tr>
<tr>
<td>Advanced Level</td>
<td>7</td>
<td>40</td>
</tr>
</tbody>
</table>

2. Text dimensions

In the Dutch Grid, text dimensions include four categories: text sources, text types, communication themes and domains. The present analyses applied the categorization of the Dutch Grid.

◆ Text sources

Figure 1 shows text sources in the GEPT Reading Comprehension Tests. Through all levels, most texts were based on information from newspapers and magazines. It is important to note that in order to compose texts of suitable difficulty level for the target examinees, all texts in the Elementary, Intermediate, and High-Intermediate Level GEPT Reading Comprehension Tests are developed and written by the GEPT item writers, and authentic passages only appear in the Advanced Level GEPT. Therefore, the variety of text sources in the CEFR can not typically represent the breadth of text genres used in the GEPT Reading Comprehension Tests.
The classification of the text types in the Dutch Grid includes five text types: descriptive, narrative, expository, argumentative, and instructive. Since a non-verbal task, i.e. graphs or figures, is typically included in the High-Intermediate Level GEPT Reading Tests, and the text type can not be found in the classification of the Dutch Grid, we therefore supplement the graph/figure as the sixth category of text types.

Figure 2 shows the result of classification of the GEPT reading text types; from the Elementary Level to the Advanced Level, the distribution of text types shifts from a greater number of descriptive and narrative texts types to a greater number of expository and argumentative text types, which corresponds to the rationale of the CEFR.

Figure 2 Text types of different levels of the GEPT Reading Comprehension (%)

---

NEED TO SPELL ELEMENTARY CORRECTLY ON ALL THE FIGURES

◆ Text types

The classification of the text types in the Dutch Grid includes five text types: descriptive, narrative, expository, argumentative, and instructive. Since a non-verbal task, i.e. graphs or figures, is typically included in the High-Intermediate Level GEPT Reading Tests, and the text type can not be found in the classification of the Dutch Grid, we therefore supplement the graph/figure as the sixth category of text types.

Figure 2 shows the result of classification of the GEPT reading text types; from the Elementary Level to the Advanced Level, the distribution of text types shifts from a greater number of descriptive and narrative texts types to a greater number of expository and argumentative text types, which corresponds to the rationale of the CEFR.

Figure 2 Text types of different levels of the GEPT Reading Comprehension (%)

---
Domains

Figure 3 shows the result of classification of the GEPT reading text domains. It appears that for all levels of the GEPT, most reading texts are within the public and educational domains, and the percentage of the texts in the personal domain decreases as the GEPT level advances.

Communication topics

Figure 4 shows the range of communication topics which the GEPT Reading Comprehension Tests cover.

The CEFR list of communication topics includes only more concrete subjects, e.g., daily life, education, weather, and topics related to more conceptual matters, such as culture/customs, science, history, literature, and fine arts, which typically appear in the High-Intermediate and Advanced Level GEPT Reading Comprehension Tests, are therefore added to the categories. Figure 4 shows there is a wide variety of communication topics in the GEPT Reading Comprehension Tests.
The linguistic and cognitive complexity of texts
The linguistic and cognitive complexity in this analysis is characterized in terms of the degree of abstraction, vocabulary, and grammar.

**Degree of abstraction**
Figure 5 shows the degree of abstraction of the texts in the GEPT Reading Comprehension Tests; as the GEPT level advances, more abstract texts are included.
To more clearly display the tendency for increasing difficulty level of the GEPT Reading Comprehension Tests, an abstraction score for each level is computed using the Dutch approach; values for the degree of abstraction are assigned to each text: 1 for ‘only concrete’, 2 for ‘mostly concrete’, 3 for ‘fairly extensive abstract’ and 4 for ‘mainly abstract.’ Figure 6 shows the average abstraction scores for different levels of the GEPT Reading Comprehension Tests. Figure 6 suggests the GEPT Reading Comprehension Tests reflect the tendency described in the CEFR. The Elementary Level contains only concrete content, whereas higher levels of the GEPT tend to contain more abstract texts. For Intermediate and High-Intermediate Levels the difference is less noticeable.

Figure 6 Abstraction scores of the GEPT Reading Comprehension Tests

- **Vocabulary**

  Figure 7 shows the analysis of range of vocabulary in the different levels of GEPT reading texts. As the level of GEPT advances, range of vocabulary expands, which reflects the expected tendency described in the CEFR.
As the analysis of the abstraction dimension does, an average vocabulary score for each level of the reading texts is computed; values for the range of vocabulary used are assigned to each text: 1 for ‘only frequent vocabulary’, 2 for ‘mostly frequent vocabulary’, 3 for ‘rather extended vocabulary’ and 4 for ‘extended vocabulary.’ Figure 8 shows the average vocabulary scores for different levels of the GEPT Reading Comprehension Tests.

Figure 8 suggests the GEPT Reading Comprehension Tests reflects the tendency described in the CEFR. The Elementary Level contains only frequently used vocabulary, whereas higher levels contain more extended vocabulary. Again, for Intermediate and High-Intermediate Levels the difference is less evident. Since the GEPT High-Intermediate Level Reading Comprehension Test normally contains a non-verbal task, the result is expected.
Figure 8 Vocabulary scores of the GEPT reading texts

Figure 9 shows the analysis of grammatical complexity of the reading texts in the GEPT. The GEPT appears to reflect the expected tendency described in the CEFR; the Elementary Level texts contain only simple grammatical structures, while the Intermediate, High-Intermediate and Advanced Level reading texts tend to contain a greater number of more complex grammatical structures (e.g., long and complex sentences).

Figure 9 Grammatical complexity of the GEPT texts (%)

A grammar score for each level is computed; values for the range of grammar used are assigned to each text: 1 for ‘only simple structures’, 2 for ‘mainly simple structures’, 3 for ‘limited range of complex structures’ and 4 for ‘wide range of complex structures.’ Figure 10 shows the average grammar scores for each level of the GEPT reading texts and reveals a relatively small difference between the
Intermediate Level and the High-Intermediate Level, and the High-Intermediate Level and the Advanced Level. Of note is the apparent drop in grammatical complexity in the High-Intermediate Level. Since the GEPT High-Intermediate Level Reading Comprehension Test normally contains a non-verbal task, and the answers to the multiple-choice questions are usually written with simple structures, the drop in the grammar score is expected.

Figure 10 Grammar scores of the GEPT reading texts

3. Dimensions of question types

Figure 11 shows the distribution in percentages of question types in the GEPT Reading Comprehension Tests. In consideration of the large volume of candidates in each test administration of Elementary, Intermediate and High-Intermediate Level GEPT, the first three levels of the GEPT adopt multiple-choice questions only, while the number of the candidates taking the Advanced Level is comparatively small, more variety of question types can be accommodated.

Figure 11 Questions types in the GEPT Reading Comprehension Tests (%)
4. **Operations**

The analyses applied the framework of the Dutch Grid; the reading operations contain three independent dimensions: the task dimension, the explicitness dimension, and the content dimension.

◆ **Task dimension**

Figure 12 shows an overview in the task dimension of the GEPT reading items in terms of reading purposes. Since speed does not appear in the CEFR reading comprehension scales, the expected speed of reading seems not to be considered as a factor that may affect learners’ choice of reading strategies and difficulty level of the reading tasks in the CEFR. While the GEPT Elementary to High-Intermediate Level Reading Comprehension Tests assess careful reading only, in the Advanced Level Reading Comprehension Test, careful reading and expeditious reading are tested separately. It is believed that purposes for reading should be included as an independent variable when analyzing operations of the reading process (Weir 2005, pp. 9-10).

![Figure 12 Task dimensions of the GEPT reading items (%)](image)

Figure 13 shows the item-level operations that the GEPT reading items assess; as the GEPT level advances, variety of items aimed at measuring different operations increases. There is a decrease in items measuring direct recognition of information and at the same time an increase in inference and evaluation items, which reflects the development described in the CEFR. For Advanced level, the pattern is more diffuse because only half of the items aim to assess careful reading and the other half aim to assess expeditious reading, i.e. skimming and scanning.
Figure 13 Operations which the GEPT items intend to measure (%)

- **Explicitness dimension**
  Figure 14 shows the extent to which reading items refer to explicit or implicit information. Through all levels, the pattern seems to be consistent; the information which the items ask for is mostly explicit.

Figure 14 Nature of the information that the GEPT reading items refer to (%)

- **Content dimension**
  Figure 15 shows the content of the operations in the reading items for different levels of the GEPT Reading Comprehension Tests. Through almost all levels, the pattern seems to be consistent; as the GEPT level advances, the items become more diversified and more cognitively challenging, but there is a drop in the Advanced Level. The construct of the GEPT Advanced Level Reading Comprehension Test is different from other levels of the GEPT Reading Comprehension Tests; it consists of careful reading and expeditious
reading, and the operations in the expeditious reading items include skimming and scanning. Therefore, the percentages of main idea and detail questions in the Advanced Level GEPT appear to be significantly higher than other levels.

Figure 15 Content of the operations in the GEPT reading items (%)

- **The Judgment Process**

  To consolidate the claim that the GEPT is related to the CEFR, after the Specification procedure was completed, the judgment process was carried out in April 2007. The judgment process followed the training and standard-setting procedures that the Manual proposed. A total of 15 judges who are experienced professionals in English teaching or testing participated in the one-day session. All participants attended Familiarization before the judgment session. The judgment session included a half-day training session in the morning and a half-day standard-setting session in the afternoon. During the morning session, the participants were trained to relate their interpretation of the CEFR levels to calibrated sample items. The exemplars employed in the session included a total of 13 calibrated samples of reading texts, along with the corresponding items, provided by the CoE, and 12 CEFR scales:

  - Table 1: Common Reference Levels: global scale (CEFR, p. 14),
  - Overall reading comprehension (CEFR 4.4.2.2, p. 69),
  - Reading for correspondence (CEFR 4.4.2.2, p.69),
  - Reading for information and argument (CEFR 4.4.2.2, p. 70),
  - Reading for orientation (CEFR 4.4.2.2, p. 70),
  - Reading instruction (CEFR 4.4.2.2, p. 71),
During Phase I Illustration and Phase II Controlled Practice of the morning session, the participants were trained to assimilate the rationale behind the CEFR levels using the calibrated sample items and the CEFR scales through consultancy with other participants and the coordinators. In the following phase, Individual Assessment, the participants then rated the CEFR levels of the calibrated tasks individually, and the coordinators checked whether they reached a general consensus. During the training session, the descriptors in Reading for information and argument were the most frequently consulted scale.

After the participants reached consensus on the CEFR levels of calibrated samples in the morning session, the afternoon session, standard-setting, adopted the method applied by the DIALANG project to determine ‘the minimum CEFR-level needed by a candidate to successfully perform (DIALANG, 2002)’ on a given level of the GEPT. The participants used Item Rating Form, adapted from Form B5 provided in the Manual, to estimate the difficulty of the GEPT reading tasks and items in terms of the CEFR levels. The tasks and items from different levels of the GEPT Reading Comprehension Test were sampled according to the following criteria: text types, topics, operations, and content questions. The GEPT items were then arranged in random order for the participants to judge the difficulty of the items in terms of the CEFR levels. The result was summarized in Table 2.

Table 2: Rater reliability, rater agreement and minimum required CEFR-level of each GEPT level

<table>
<thead>
<tr>
<th>GEPT Level</th>
<th>No. of Items</th>
<th>Rater Reliability</th>
<th>Rater Agreement</th>
<th>Mean Required minimum CEFR level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>20</td>
<td>0.75</td>
<td>0.53</td>
<td>2.40 (A2+)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>11</td>
<td>0.91</td>
<td>0.71</td>
<td>3.14 (B1)</td>
</tr>
<tr>
<td>High-Intermediate</td>
<td>22</td>
<td>0.87</td>
<td>0.51</td>
<td>3.75 (B2-)</td>
</tr>
<tr>
<td>Advanced</td>
<td>17</td>
<td>0.91</td>
<td>0.82</td>
<td>5.01 (C1)</td>
</tr>
</tbody>
</table>

The overall rater agreement for the GEPT items was 0.91, which suggested the participants had a generally satisfactory degree of consensus on the minimum CEFR-level required for the GEPT sample items. The mean required minimum CEFR level increases as the GEPT level advances.
III Conclusion

The findings of this GEPT-CEFR mapping project reveals that the GEPT Reading Comprehension Tests correspond in general to the CEFR levels. The content analyses guided by the framework of the Dutch Grid indicate that the GEPT reading texts become more complex as the GEPT level ascends; the degree of abstractness of the texts increases and the range of vocabulary used is more extensive. The analysis of the items showed an increase in the diversity of the operations involved from the Elementary Level to Advanced Level, which corresponds to the CEFR rationale.

However, we have also seen some problems. For instance, expeditious reading and non-verbal task input are not addressed in the CEFR. Therefore, we agree with Fulcher’s comment (2004) that although the Specification provided in the Manual facilitates detailed reports on the quality of the tests, its lack of precision and clear guidelines makes comparison of constructs and difficulty level between different testing systems difficult. In addition, the content and format of the Specification are not easy for non-professionals in the language testing field, i.e. examinees, teachers, and other stakeholders involved in the assessment process, to understand the test in question, since they are more familiar with subtest-oriented categories, e.g. reading comprehension tests and writing tests, rather than the communicative language activity oriented ones, e.g. written production, written interaction and written mediation, that the Manual proposes.

As an exam board, during the process of Specification, we encountered some difficulties when relating our test to the CEFR. In the CEFR Can-do statements, few descriptions of how well learners are expected to perform at a particular CEFR level are included (Fulcher 2004, p. 256; Weir 2005, p. 8). This was a major obstacle because the quality of a learner’s performance is a major indicator of his or her proficiency level. Furthermore, as Weir (2005) noted, some test conditions obviously affect task difficulty, e.g., rate of delivery, speech rate, expected reading speed, passage dependency, response format, and length of output, but these aspects are left out of the CEFR, thus hindering our efforts to link the GEPT to the CEFR levels.

This project related the GEPT Reading Comprehension Tests to the CEFR by following the proposed methods in the Manual and the procedures in the Dutch project. The experience has been a valuable lesson that will benefit our future projects to relate other GEPT tests, namely Listening Comprehension, Writing and Speaking Tests, to the CEFR levels. To strengthen the forthcoming linking project, an empirical validation that follows the procedure the Manual proposed is underway.
IV References


Council of Europe. (2005). *Relating language examinations to the Common European Framework of Reference for Language: Learning, teaching, assessment (CEFR)—Reading and Listening Items and Tasks: Pilot Samples illustrating the common reference levels in English, French, German, Italian and Spanish*.


