RESEARCH ARTICLE

Approaches to Scoring Translation in the PROFEX EMP Exam

Hegedűs Anita

Department of Languages for Specific Purposes, Faculty of Medicine, University of Pécs, Pécs, Hungary

Introduction: This paper undertakes to investigate the connection between two approaches to scoring translation by examining 115 test papers of the translation component of the C1 level PROFEX English for Medical Purposes (EMP) exam. The main objective of the study is to reveal whether, and to what extent, the method of assessment influences the score.
Material and method: The test papers were scored independently by two experienced raters according to the marking scale of the PROFEX EMP exam, then holistic scoring was carried out by a third rater, who was uninformed of the official scores for the test papers. Correlations were calculated, first between the holistic scores and the official scores based on the combined holistic and discrete point approach, then between other components of the written part of PROFEX EMP exam (reading comprehension and writing) and the holistic scores and the scores reached with the combined method, respectively.
Results: A strong correlation has been revealed between the scores achieved by the purely holistic method and those assessed with the combined holistic and discrete point approach. The holistic method was shown to be slightly more reliable than the combined approach.
Conclusion: The study has revealed that the method of assessment does not significantly influence the score in the evaluation of translation in the PROFEX EMP exam.

Keywords: assessing translation, holistic method, correlation, discrete point testing, reliability

Introduction

This paper undertakes to investigate the connection between two approaches to scoring translation by examining 115 test papers of the translation component of the advanced (C1) level PROFEX English for Medical Purposes (EMP) exam. Specific purpose language tests, including EMP tests, must be preceded by an analysis of the target language situation, so that test tasks and content represent tasks examinees are exposed to in target language situations [1]. Examinees who take the PROFEX EMP exam are doctors, pharmacists, researchers and other health professionals who are inevitably faced with tasks requiring translation, either in their research or their work, e.g. when they translate articles into English before publishing it in a foreign journal or when they translate documents before taking up a job abroad. In the PROFEX EMP exam, tasks for translation may include abstracts of medical research articles, descriptions of research projects or descriptions of diseases. The length of the source language text can range between 250–300 words.

A key issue and a question giving rise to arguments in specific purpose language testing is the development of evaluation criteria along which test papers are assessed [2]. Assessment of translation is an unexplored field in the literature on language testing and evaluation: very few studies have been conducted in this field compared to other areas of the discipline.

The literature on language assessment is replete with arguments for or against the holistic approach to scoring [3,4,5]. Most of them agree that scoring with the holistic method is very rapid and easy, allowing for large-scale and multiple assessment. However, by assigning one single score for a test paper, the holistic method may not be as reliable as other methods of scoring, e.g. analytic or discrete point scoring. In the PROFEX EMP exam, translation is evaluated based on a method combining discrete point testing and holistic scoring. In this paper, this approach is compared to a method involving purely holistic scoring. The main objective of the study is to reveal whether, and to what extent, the method of assessment influences the score.

Material and method

In order to investigate the connection between the two approaches to scoring translation (the purely holistic approach and the one involving an element of discrete point evaluation), a sample consisting of 115 test papers of the translation component of the advanced (C1) level PROFEX English for Medical Purposes (EMP) exam were examined. The test papers were collected after the PROFEX EMP exam conducted in November, 2011. First, the test papers were scored independently by two experienced raters according to the regularly used marking scale of the PROFEX EMP exam, then holistic scoring was carried out by a third rater, who was uninformed of the official scores for the test papers.

According to the Rules and Regulations of Organisation and Operation of PROFEX, in the assessment of transla-

Correspondence to: Anita Hegedűs
E-mail: anita.hegedus@aok.pte.hu

tion in the PROFEX EMP exam, five sentences or clauses containing one or two grammatical structures characteristic of the English language are designated in the text. The selection of the five items takes place simultaneously with test construction, and examinees are not informed of the location of these items in the text. A maximum of two scores can be awarded for each item. Two scores are assigned for an item if the information carried by the sentence or clause is completely conveyed by the examinee and there are only minor grammatical mistakes in the item. One score can be awarded if there are serious or several grammatical mistakes in the item, but the information transfer has been accomplished, while no score is given if the examinee partially or completely fails to convey information carried by the item. After scoring the five, designated items, test papers are assessed holistically, for which a maximum of five scores can be awarded. The holistic score and the scores for the designated items aggregate to a maximum score of 15. Double marking is carried out in the PROFEX EMP exam, which means that each test paper is assessed by two trained raters independently, i.e. raters are not informed of the score assigned by the other rater. The arithmetic mean of the two scores yields the final score, however, if there is a deviation of three or more points between the scores assigned by the two raters, the final score is decided upon by the chief examiner.

In the purely holistic scoring, besides grammatical accuracy and style, a main factor in the assessment was the extent to which the examinee conveyed information contained in the source language text, since specific purpose tests are — by definition — communicative [1].

The second stage of the study involved calculating the correlation between the holistic scores and the official scores based on the combined holistic and discrete point approach, using the Correl programme of Microsoft Excel.

In the third stage of the research, correlations were calculated between other components of the written part of PROFEX EMP exam (reading comprehension and writing) and the holistic scores and the scores reached with the combined - holistic and discrete point - method, respectively. Since the maximum score for reading comprehension is 20, scores for this skill have been converted into a 15-point-scale by dividing scores by 1.33, in order to make them compatible with scores for translation and writing, both of which having a maximum score of 15. During the conversion, the general rules for rounding were applied.

Results

In the study, a strong correlation of 0.972 was revealed between the scores achieved by the purely holistic method and those assessed with the combined holistic and discrete point approach. Correlations between the scores for the two other sub-skills of the written part of the C1 level PROFEX EMP exam (reading comprehension and writing) and the scores for translation achieved by the two approaches investigated in the study are displayed in Table I.

Table I.   Correlation results

Skills	Correlation
Translation scored holistically – reading	0.6079
Translation scored holistically – writing	0.6549
Translation scored with the combined method – reading	0.5652
Translation scored with the combined method – writing	0.6219

As the results show, the correlations between the scores for translation with the holistic method and the scores for the two other sub-skills were slightly higher than those between the scores for translation assessed with the combined method and the scores for the two other sub-skills.

Discussions

In evaluating translation in the PROFEX EMP exam, reliability is designed to be provided by assigning two-thirds of the scores (10 out of 15) based on five discrete point elements (sentences or clauses) selected in the text in advance. The remaining five points are allotted based on holistic scoring, so that the overall effect of the translation could be evaluated as well. Although the proportion of the scores allotted for each method (two-thirds of the scores based on discrete point evaluation, while one-third of the scores based on overall impression) might give rise to debates, the results of this study does not call for the elimination of this approach. Nevertheless, careful consideration is required to determine the five designated items, so that reliable scores could be achieved. Also, the selection of five items may seem insufficient compared to the length of the text (250–300 words), but, still, this problem can be dealt with by meticulous selection of the five assessed items.

One might argue that by allotting two-thirds of the obtainable scores based on five designated items containing grammatical structures characteristic of English, too much emphasis is laid on using grammar accurately, resulting in lower scores in a language exam which is intended to be communicative. However, calculations of mean scores contradict this hypothesis. Test papers evaluated with the purely holistic approach had a mean score of 8.76 (out of the maximum 15 scores), which was lower than the mean of those assessed with the combined approach, which was 9.16. That is, examinees would have received a lower score, had their pieces of work been evaluated based on the purely holistic method.

By exhibiting a stronger correlation to the other two sub-skills of the written part of the PROFEX EMP exam, the purely holistic method seems to be more reliable than the combined, holistic and discrete point approach. The deviation, however, is not significant enough to discard the application of the combined method in the PROFEX EMP exam. Furthermore, other factors, such as differences in rater characteristics or the location of the five pre-determined items in the text may have had an impact on the correlation results.

Conclusions

This study involving a sample consisting of 115 test papers has revealed that the method of assessment does not significantly influence the score in the evaluation of translation in the PROFEX EMP exam. The holistic method of scoring was found to be slightly more reliable than the combined approach involving holistic and discrete point evaluation, still, the difference was not big enough for the elimination of discrete point testing from the assessment of translation in the PROFEX EMP exam.

References

1. Douglas D. Assessing Languages for Specific Purposes. Cambridge University Press. Cambridge, 2000.
2. McNamara T. Measuring second language performance. Longman. London, 1996.
3. Hughes A. Testing for Language Teachers. Cambridge University Press. Cambridge, 1997.
4. Shaw SD, Weir C. Examining Writing: Research and Practice in Assessing Second Language Writing. Cambridge University Press. Cambridge, 2007.
5. Weigle SC. Assessing Writing. Cambridge University Press. Cambridge, 2002.