The Connection between Analytic and Holistic Approaches to Scoring in the Writing Component of the PROFEX EMP Exam

Hegedűs Anita

Department of Medical Linguistics, Faculty of Medicine, University of Pécs

Introduction: This paper aims at investigating the connection between analytic and holistic approaches to scoring writing in the PROFEX English for Medical Purposes (EMP) exam. The sample for the research consisted of 60 test papers taken at the PROFEX EMP exam in April, 2010.
Material and method: First, the test papers were scored analytically by two raters independently of each other, then they were scored holistically by a third rater. After the scoring had been completed, correlations were calculated between the two methods of scoring and also between writing and other components of the language exam.
Results: The study has detected no significant differences in the correlations between the analytic and holistic scores.
Conclusion: The holistic method was found to be more reliable than the analytic method, but the deviation was not sufficient for discarding the analytic method in the assessment of writing in the PROFEX EMP exam.

Keywords: scoring writing, analytic method, holistic method, correlation, reliability

Introduction

This paper undertakes to investigate the connection between holistic and analytic methods of scoring writing by examining 60 test papers of the writing component of the intermediate level PROFEX English for Medical Purposes (EMP) exam. The two main approaches to assess writing is the holistic and analytic method. The holistic method is applied when one score is assigned to the test, based on overall impression, whereas the analytic method involves giving a separate score for a number of features of the test (e.g. grammar, vocabulary, content). The main focus of the study is to reveal whether the method of assessing writing influences the score.

Material and method

In an attempt to investigate the connection between the analytic and the holistic approach 60 test papers of the writing component of the intermediate (B2) level PROFEX English for Medical Purposes (EMP) exam were collected. The language exam was conducted in April, 2010 and the test papers were selected at random. First, the test papers were assessed analytically by two raters, then holistic scoring was carried out by a third rater, who was unaware of the analytic result.

According to the Rules and Regulations of Organisation and Operation, in the writing component of the PROFEX medical language exam, test papers are assessed along three traits: (1) form and style, (2) grammar and (3) content and coherence. A separate marking scale is employed for each feature. Maximum five points can be awarded for each of the three features, and these aggregate to a maximum of 15 points. Double scoring is applied in the PROFEX EMP exam: each test paper is scored by two trained raters who are unaware of the score assigned by the other rater. The final score is the arithmetic mean of the two scores, however, if there is a difference of three or more points between the two raters' scores, the final score is determined by the chief examiner. For the holistic scoring a modified version of the marking scale of the British Council's IELTS test adjusted to the specific features of the EMP test was applied.

In the writing component of the IELTS test, candidates receive scores on a band scale from 1 to 9, which takes into account four criteria: task achievement, coherence and cohesion, lexical resource and grammatical range and accuracy. These four criteria have equal weight in the scoring process [1] . For the purposes of the study, I have converted the 9 bands into a scale of 5: IELTS bands 0 and 1 correlate to score 1 in this study, IELTS bands 2-3 correlate to score 2, IELTS bands 4-5 correlate to score 3, IELTS bands 6-7 correlate to score 4 and IELTS bands 8-9 correlate to score 5 in this study.

The second phase of the research involved calculating the correlation between the holistic and analytic scores by applying the Correl programme of Microsoft Excel.

The third phase of the study involved calculating the correlation between the writing scores and those of the other components of the PROFEX B2 level EMP exam (speaking, listening comprehension, reading comprehension and summarising) and the holistic and analytic scores, respectively. Since the maximum score is 30 points for speaking and 20 points both for reading comprehension and listening, all scores were converted into a scale of 15 by dividing reading comprehension and listening scores by 1.33 and speaking scores by two, so that scores could be comparable (since a maximum of 15 points can be assigned for the writing component) and correlation be-

Table I.   Scores for writing

	Analytical method	Holistic method		Analytical method	Holistic method
1	6	4	31	5	5
2	0	0	32	15	14
3	10	10	33	6	6
4	13	12	34	14	12
5	8	7	35	14	11
6	13	12	36	10	9
7	13	12	37	13	13
8	14	13	38	14	12
9	8	8	39	10	8
10	11	11	40	3	3
11	14	13	41	12	12
12	9	8	42	8	6
13	12	10	43	11	10
14	8	7	44	14	14
15	13	13	45	10	9
16	0	0	46	3	3
17	14	13	47	9	6
18	5	3	48	13	10
19	15	13	49	6	6
20	5	3	50	10	6
21	11	11	51	6	8
22	15	14	52	11	11
23	11	11	53	10	10
24	11	10	54	11	11
25	12	10	55	5	4
26	12	10	56	7	7
27	14	12	57	12	12
28	11	9	58	8	8
29	9	9	59	10	9
30	3	3	60	3	3

Table II.   Scores for other components of the PROFEX exam

Skill	Listening comprehension	Speaking	Reading comprehension	Summarising
1	2	6	2	3
2	5	8.5	6	4
3	5	10	7	7
4	11	11	8	10
5	8	7	6	5
6	13	12	11	14
7	9	13.5	11	12
8	15	14.5	14	14
9	11	10	10	9
10	11	13.5	12	11
11	11	12.5	12	13
12	5	6.5	5	7
13	9	12.5	11	11
14	7	9	8	8
15	12	12	14	12
16	0	0	6	0
17	5	7.5	10	14
18	5	5	3	5
19	13	12	9	11
20	3	3.5	4	5
21	13	14.5	9	12
22	13	14	11	14
23	6	9.5	8	8
24	4	7	8	9
25	11	8	8	9
26	8	9.5	8	10
27	12	11.5	9	11
28	6	12	7	10
29	5	6	8	11
30	2	1	4	2
31	2	6.5	4	4
32	14	13.5	14	14
33	0	0	4	5
34	8	14	6	12
35	11	9.5	8	11
36	10	13.5	14	9
37	10	12	11	6
38	10	12.5	14	10
39	9	12.5	14	8
40	9	11	12	2
41	7	12	8	9
42	7	11.5	4	6
43	8	11	5	11
44	8	12.5	11	12
45	9	10	5	10
46	8	11	10	2
47	7	12.5	7	6
48	7	12	8	7
49	8	11.5	6	5
50	9	12.5	11	6
51	12	7.5	14	10
52	12	11	14	12
53	10	8.5	10	10
54	13	10.5	14	12
55	5	0	13	5
56	14	11.5	14	11
57	14	10	14	13
58	11	11	10	9
59	11	10	13	10
60	2	2.5	3	6

tween the different skills could be calculated. The general rules for rounding were applied, except for values ending in 0.5, which were not changed.

Results

The results of the analytic and holistic method of scoring are displayed in Table I. The comparison of results of the analytic and holistic scoring yielded a strong correlation of 0.9607.

Table II shows the scores for other components of the PROFEX EMP exam.

Table III shows the correlation between the other components of the PROFEX language exam (listening comprehension, speaking, reading comprehension and summarising) and the writing component.

Table III.   Correlation results

Skills	Correlation
Writing scored analytically – listening	0.6072
Writing scored analytically – speaking	0.6617
Writing scored analytically – reading	0.4257
Writing scored analytically – making an abstract	0.8314
Writing scored holistically – listening	0.6469
Writing scored holistically – speaking	0.6405
Writing scored holistically – reading	0.5017
Writing scored holistically – making an abstract	0.8726

The correlations were calculated both for the analytic and holistic scores. A slightly higher correlation was found

between the scores for writing assessed with the holistic method and the scores of the other components of PRO-FEX EMP exam. The only exception was found in the speaking component, where the correlation for the analytic scoring was slightly higher.

The most significant difference can be observed in the results of the reading component: the correlation between reading and writing scored holistically was by 0.8 more than that between reading and writing scored analytically.

Discussions

In the literature many arguments have been put forward for or against the analytic and holistic approaches to scoring [2, 3, 4]. Most of them agree that the main advantage of the holistic method is that it is very rapid and easy, thereby enabling large-scale and multiple assessment, therefore it is preferred by the largest examination boards, like Cambridge ESOL. On the other hand, the main disadvantage is that by assigning one single score the holistic method falls short of providing diagnostic information and distinguishing between different aspects of writing [2, 3].

By assigning separate scores for different features of writing, the analytic method may be considered more reliable. Also, in this way features of performance which might be ignored by using the holistic method are evaluated. Furthermore, the analytic method can provide diagnostic information on the development of individual sub skills. The main disadvantage of the analytic method is that it is time-consuming. In addition, by focusing on sub skills the overall effect of the piece of writing may not receive enough attention, therefore the analytic scoring may be less valid than the holistic [2, 3, 4].

The present study has revealed no significant difference in the correlation between the analytic and holistic scores of the writing component of the PROFEX EMP exam, which provides further evidence for the findings of other research carried out in this field [5]. This result may be explained by the finding that the same sub skills are evaluated during the application of both methods, as raters unconsciously tend to resort to holistic methods while scoring a test paper analytically [4]. That is, some raters may read and assess a test paper holistically and then adjust the analytic scores to the holistic impression. Furthermore, Shaw and Weir argue that marking scales serve as only "mnemonic devices" for scoring, because most raters - especially the experienced raters - have already internalised a level of representation of the different skills, which they apply during the evaluation. Moreover, holistic scoring scales incorporate the same sub skills (e.g. style, accuracy, content) as the ones separated out as different sub skills in the analytic method [3].

This study has also demonstrated that higher correlations could be detected between the holistic scores and the scores for other components of the language exam than that between the analytic scores and scores for the other skills. However, the deviations were not significant enough to prove that the holistic method is more reliable. In the author's view, differences between rater characteristics, such as differences in education, experience, consistency, degree of leniency, account for the differences between the analytic and the holistic scores to some extent, but further research is required to prove that.

Although the main focus of the study was to investigate the connection between analytic and holistic scoring, it is worth noting that the highest correlation between the scores for writing in general (both analytically and holistically scored) and other skills was detected in the case of the skill of summarising. This finding can be explained by the fact that the same sub skills are required to produce both of them. The lowest correlation was observed between reading comprehension and writing, which is accounted for by a few conspicuous differences, e.g. Student 40 scored 3 on writing (for writing the official analytic scores are displayed) and 12 on reading, Student 46 scored 3 on writing and 10 on reading, Student 51 scored 6 on writing and 14 on reading, Student 55 scored 5 on writing and 13 on reading and Student 56 scored 7 on writing and 14 on reading. Examining a larger sample would result in a higher correlation between writing and reading. However, since the main aim of this paper is to investigate the link between the holistic and analytic approaches to scoring, a more profound examination of the correlation between writing and other skills falls outside the scope of this study, but it may give rise to further, larger-scale investigations.

Conclusions

This paper has revealed no significant difference between the results of scoring with the analytic and holistic method, which supports the results of studies carried out in this field. Calculations of correlation between writing assessed with both methods and other components of the PROFEX EMP exam have shown the holistic method to be slightly more reliable than the analytic method. However, in the author's view, this finding can largely be due to differences in rater characteristics, and does not provide ample evidence to discard the analytic method in the scoring of the writing component of the PROFEX EMP exam. Further research is required to investigate the differences between rater characteristics.

References

1. www.ielts.org
2. Hughes A – Testing for Language Teachers. Cambridge University Press. Cambridge, 1997.
3. Shaw SD, Weir C – Examining Writing: Research and Practice in Assessing Second Language Writing. Cambridge University Press. Cambridge, 2007.
4. Weigle SC – Assessing Writing. Cambridge University Press. Cambridge, 2002.
5. Lee Y, Gentile C, Kantor R – Toward Automated Multi-trait scoring of Essays: Investigating Links among Holistic, Analytic, and Text Feature Scores. Applied Linguistics 2009, 3:391-417.