

ON THE RATING STRATEGIES USED BY RATERS OF THE JCEE ENGLISH COMPOSITIONS: A PILOT STUDY

Hsueh-ying Yu

尤 雪 瑛 *

摘 要

從民國七十年開始，大學日間部聯招英文科開始加考作文，由於英文作文採人工閱卷的主觀評量方式，其可靠性一直為衆人所關切。本研究旨在探討評分者閱卷時，各評分項目比重如何？使用那些策略來評量？有沒有個別差異？本研究以問卷方式，請八十四年參與閱卷的六十五位評分委員提供資料。問卷內容分三部份：個人資料，對五項評分項目比重之看法，及三十六種評分策略使用情形。調查結果分別以平均頻率，及變異數分析等方式，進行分析討論。結果發現，儘管評分者專長及教學經驗不同，他們在評分項目比重分配，及策略使用上並無顯著差異。

I. Introduction

English writing ability has been evaluated directly in the Joint College Entrance Examination (JCEE hereafter) since 1981. The main reason for this design was the consideration of the effectiveness of the test itself and of its backwash in teaching. It has received more and more support that writing ability can only be successfully evaluated by asking the students to perform the writing skill. And it is also believed that such a direct test will bring beneficial backwash effect on language teaching and learning, since it requires actual practice of the skill, and in order to get better grades both teachers and students will pay more attention to training writing ability (Heaton, 1990; Hughes, 1989; Weir, 1990). However, because this kind of task tests requires subjective evaluation of writing ability, the reliability in the scoring procedure has been

* 作者為本校英語系專任講師

questioned frequently by high school teachers, students, and even raters themselves, although it is widely accepted to be the right direction for the test of English writing ability in the JCEE (Chen et al., 1992; Chen et al., 1993; Huang et al., 1994).

Two limits of our current situation give rise to the public doubt of this subjective evaluation. First, there is little knowledge about skills and criteria which raters use for the writing assessments in the JCEE. Second, no empirical evidence has been provided to account for the inter-rater reliability of the English composition scoring in the JCEE. Researches have been done in other countries, most of them published in the United States, on the skills involved in evaluating writing ability of English as second language (Cumming, 1990; Homburg, 1984).¹ And there are also studies on the reliability of the subject scoring of English compositions. Among them, Mullen (1980), in his analysis of composition scoring in assessing second-language writing ability, reports that some pairs of raters achieved high reliability, but there were also pairs of raters that did not perform reliable grading. Diederich et al. (1961) finds that even English teachers showed diversity of deciding writing quality. And yet, Kaczmarek (1980) says that the results of her analysis indicated substantial reliability among individual raters. Despite of all the pros and cons of the subjective scoring proposed by the researches in other countries, we are curious about what would happen to local raters of the JCEE English compositions. What are their strategies or behaviors? What criteria do they use? And most important of all, are they consistent in the rating process?

The goal of this study is to investigate (1) the strategies used by our raters of the JCEE English compositions; and (2) the inter-rater reliability of our raters with regard to the strategies and criteria they used. Two factors will be considered in our analysis to see if they had effects on the raters' behaviors. They are the raters' majors and their composition teaching experiences.

II. Methods

1. Subjects

The subjects for this study were 65 college English teachers who participated in the English composition and translation scoring of the JCEE in 1995. Twenty-three (35.38%) majored in English language teaching, twenty-three (35.38%) in English literature, and nineteen (29.24%) in linguistics. Fifty-five (84.62%) of

On The Rating Strategies Used By Raters Of The JCEE English Compositions: A Pilot Study

them had the experience of teaching college writing. The number of the raters who had taught compositions for more than four years amounted to 21 (32.31%). The majority of the raters were female (61.54% female and 38.46% male). Fifty-four of them (83.08%) have participated in the JCEE scoring before, and only 11 (16.92%) of them had no such prior scoring experience.² The average time for scoring fifty pieces of paper (including both composition and translation) was about 2.22 hours at first, but when they got familiar with the job, they needed only 1.62 hours.³

2. Materials

This study investigated empirically the thinking process involved in the composition evaluation. Since the rating process is complex and the psychological activities cannot be observed directly, little has been known about the raters' thinking behaviors thus far. In order to elicit possible strategies, we first interviewed seven raters who had participated in the JCEE English composition scoring. They were selected to cover contrasts between different majors of English language teaching, English literature, and linguistics, and between different prior experiences of teaching composition and scoring the JCEE papers. By doing so, we expected to collect strategies of raters with different backgrounds. In our final draft of the questionnaire, we identified 36 strategies in total.

The questionnaire consisted of two sections containing 44 items. The first section reported the subjects' personal data including sex, teaching experiences, major, and prior experience of scoring the JCEE papers. In addition, the subjects were asked to indicate how they weighed the five scoring categories of content, organization, grammar, vocabulary and mechanics. A five-point scale (very important, important, neither important nor unimportant, not very important, and not important) was provided.

The second section contained 36 strategies used in the composition scoring. Again, a five-point scale of "all the time," "most of the time," "half of the time," "seldom," and "never" was provided for the subjects to indicate the frequency of adopting each strategy. The strategies covered six topics according to their purposes. First, Strategy 1 to Strategy 5 (S-1 to S-5) dealt with the topic of the holistic vs. analytic scoring methods. Then, for the strategies approaching the five scoring categories, we had S-6, S-7, and S-13 concerning the content category, S-8 the organization category, S-18 to S-21 and S-24 to S-27 the grammar category, S-22 and S-23 the vocabulary and the mechanics categories. S-

16 and S-17 were strategies about choosing the global categories (content and organization) or the local categories (grammar and vocabulary) as the primary consideration. Third, S-9 to S-15 and S-31, and S-32 included strategies of establishing personal grading scale. Then, S-28 to S-29 were strategies of making judgment by the appearance of the composition. S-33 was an interpretation strategy. And finally, S-34 to S-36 reported social strategies used by raters.

3. Procedures

The questionnaires were delivered to the subjects on the second and the third days of the one-week scoring period. The reason of collecting data at that time was out of the consideration that the raters should have got some ideas of the general quality of the compositions, have been familiar with the grading guidelines, and most important of all, they should have developed their own strategies.

The raters' responses were scored by counting 5 for "very important" or "all the time" to 1 for "not important" or "never" across all items. Measures ANOVAs were used for the analysis. The scores for each of the subjects were treated as the dependent variable and the scores of the factors of the raters' majors and their prior teaching composition experiences as independent variables. Means were calculated for the weights of the five scoring categories assigned by the raters, and for the frequencies of the 36 strategies used by the raters.

III. Results

1. The Rating Strategies and Their Frequencies

The mean frequencies of each strategy were presented in Table 1. From the data, we wanted to investigate the following questions:⁴

1. Which of the two scoring methods was more frequently used? The holistic method or the analytic method?
2. How did the subjects make judgments on the five scoring categories? What strategies did they tend to apply?
3. What strategies were adopted to establish the raters' own rating systems?

On The Rating Strategies Used By Raters Of The JCEE English Compositions:
A Pilot Study

4. Were social strategies popular among the raters?

Table 1 Mean Frequencies of the 36 Strategies

Strategies	Means	Strategies	Means	Strategies	Means	Strategies	Means
1	2.66	11	3.52	21	2.80	31	2.15
2	3.20	12	3.17	22	3.22	32	2.72
3	1.89	13	3.51	23	2.94	33	2.54
4	1.60	14	3.45	24	2.74	34	2.66
5	2.63	15	4.40	25	2.92	35	1.89
6	2.05	16	3.57	26	2.43	36	2.47
7	3.11	17	2.45	27	3.31		
8	3.06	18	2.40	28	3.09		
9	2.29	19	2.26	29	1.63		
10	3.54	20	3.00	30	1.98		

Before we start to discuss the four questions, let us survey the mean frequencies of all the strategies first. As displayed in Table 1, most of the mean frequencies fell on the range from 2.05 to 3.57, which represented an area between the points of "seldom" and "at half of the time." Only one strategy was reported to be used "most of the time" (S-15, mean=4.40). And five were reported to be nearly "never" used (S-3, mean=1.89; S-4, mean=1.60; S-29, mean=1.63; S-30, mean=1.98; S-35, mean=1.89).

For the choice between the holistic and the analytic methods, according to the mean frequencies indicated in Table 1, S-1, an analytic strategy, was used less frequently than S-2, a holistic strategy. The frequency mean of S-1 was 2.66, and 3.20 for S-2. In other words, when being asked if they gave scores separately for each category and then added them together as a whole, the raters reported an average frequency as "seldom" on the scale. And when they were asked if they gave a single score, based on the overall impression, the frequency mean fell on the "half of the time" level on the scale. The mean of S-5 provided supporting evidence of this finding. A strategy of reading compositions more than one time for decision-making, as stated in S-5, was more likely to be adopted by analytic raters. And it was found that the mean of S-5 (2.63) was close to that of S-1.

Below we will present the results of the use of the strategies related to assessing the five scoring categories. For the content and the organization categories, we proposed one strategy for each of them. As a strategy to evaluate content, S-7 stated that raters would summarize key points of the composition content; and we got a mean of 3.11. S-8 suggested an organization evaluation

strategy of tracing the beginning, the development and the conclusion of the compositions to have a picture of the whole organization. The frequency mean of this strategy was 3.06. These results seemed to go counter to our findings of the weights of the categories as we will see in the later analysis, which indicated that the raters considered content and organization as "very important" criteria for rating. One possible explanation might be the limit of the type of writing task. In this year, the JCEE English composition task was to require the students to write two paragraphs on a topic, with a given topic sentence placed at the beginning of each paragraph. On several occasions, raters said that such a design had predetermined the content and the organization of the compositions. Therefore, the students' compositions showed few varieties in these categories, and the raters did not have to labor in the assessments of these two categories, even though they still believed that the categories were very important.

Eight strategies were suggested for the assessment of the grammar category. The results show that the raters seldom counted the number or the ratio of grammatical errors (S-18, mean=2.40; S-19, mean=2.28; S-21, mean=2.80), but they would pay special attention to syntactic complexity at half of the time (S-20, mean=3.00). Furthermore, they tended not to deduct scores for grammatical errors of the use of articles, plural forms, and subject-verb agreement if they did not hinder comprehensibility (not to deduct, S-24, mean=2.74; and to deduct, S-25, mean=2.92). However, they showed less tolerance of the errors of verb tenses: not to deduct, S-26, mean=2.43; and to deduct, S-27, mean=3.31. For the mechanics category, the raters were likely not to punish students for errors of punctuation and spelling at half of the time (mean=3.22).

Over all, in the rating process, the raters at more than half of the time tended to give a score on the basis of the content and the organization of the compositions, then adjusted the scores on the basis of the grammar and the vocabulary (S-16, mean=3.57). It was seldom that they would give a score at first on the basis of the grammar and the vocabulary categories and adjusted the score according to their performances on the content and the organization categories (S-17, mean=2.45).

For the third issue of the establishment of personal rating system, seven strategies were proposed. Interestingly, the mean frequencies of six of them were relatively high. The highest mean was that of S-15 (mean=4.40), which was the highest mean of all of the 36 strategies. It stated that before the actual scoring, the raters would first rate representative compositions for each level provided by the JCEE Committee, to compare their own scales with that proposed by the committee. The other related strategies were:

On The Rating Strategies Used By Raters Of The JCEE English Compositions: A Pilot Study

- S-9 to compare their scores with the equal scores of the percentage system (mean=2.29)
- S-10 to compare compositions with the representative compositions (mean=3.54)
- S-11 to compare compositions with those that have been rated as a regular check on intra-rater reliability (mean=3.52)
- S-12 to set norms of several levels before actual scoring (mean=3.17)
- S-13 to think about possible content or points of the topic beforehand (mean=3.51)
- S-14 to set standard levels of high school graduates' writing ability (mean=3.45)

The results suggest that more than half of the time, the raters would decide their own rating scales or systems and tried to adhere to the systems with several strategies.

The final question was how popular the social strategies were among the raters. According to Table 1, the three suggested strategies were seldom used. They did not tend to discuss with or consult people when they faced difficulties of not knowing how to give a score (S-34, mean=2.66), or of uncertain grammatical or language usages (S-36, mean=2.47). The other social strategy was to consult the dictionary. Again, the frequency was pretty low (mean=1.89). There was a dilemma of using social strategies: peer socialization was a good way to reach agreement, but it was time-consuming and disturbing at the same time. This could be a main reason that prevented the raters from using the social strategies.

2. Inter-rater Reliability of the Use of Strategies and Criteria

Two independent factors of academic backgrounds and prior composition teaching experiences were tested to see if they had effects on the use of the strategies. Interestingly, the results indicate that the two factors had neither main effects nor interaction effects significantly on the use of almost all the strategies, except S-32. This strategy was a way of rating the two paragraphs of the compositions separately and then adding the scores together. The ANOVA results show that the interaction effects between the raters' majors and their composition teaching experiences on S-32 was significant ($F=5.39$, $p<.05$). Table 2 shows the mean frequencies for the six groups.

Table 2 Numbers, and Mean Frequencies
for the Interaction Groups

Groups	No.	Means
EFL*-TEA	6	2.83
EFL*+TEA	17	2.53
LIT*-TEA	2	5*
LIT*+TEA	21	2.76
LIN*-TEA	2	1*
LIN*+TEA	17	2.71

Notes: EFL, LIT, and LIN represent the academic backgrounds of English language teaching, English literature and linguistics. +TEA and -TEA mean with or without composition teaching experience. The combination of "EFL*-TEA" stands for the group with the major of EFL but with no composition teaching experience.

As indicated in Table 2, the mean frequencies of four of the groups were close, but the raters of the groups with English literature or linguistics backgrounds and without prior composition teaching experience had distinctive attitudes toward S-32. Those who majored in English literature reported to use the strategy "all the time" and the majors of linguistics reported "never" using the strategy. Although the interaction effect was significant here, it involved only four raters. The results might not be representative. Therefore, we might infer that, over all, for most raters, the factors of academic backgrounds and composition teaching experiences did not significantly affect the raters' behaviors of using the rating strategies.

The same factors were tested again to see their effects on the raters' criteria. ANOVA results indicate that the raters shared common value of the scoring categories, in spite of the differences of academic backgrounds and prior composition teaching experiences among them. The two factors showed neither significant main effects nor interaction effects on the raters' focus on the five categories. This fact suggests that the raters, whether they majored in English language teaching, English literature or in linguistics, and whether they had taught college writing or not, assigned similar weights on the five categories, when they rated the papers.

On The Rating Strategies Used By Raters Of The JCEE English Compositions:
A Pilot Study

Table 3 Mean Weights of the Scoring Categories

Items	Means
Content	4.67
Grammar	4.28
Organization	4.23
Vocabulary	3.78
Mechanics	3.17

As Table 3 shows, the categories of content, organization, and grammar were almost regarded as very important criteria for rating. As for the other two categories of vocabulary and mechanics, they were considered to be neither important nor unimportant. One thing worth noticing was that the mean of the content category (mean=4.67) was much higher than those of the other categories, and it was pretty close to the score of being "very important." On the other hand, the mean of the mechanics category was the lowest (mean=3.17), which was ranged closely to the "neither important nor unimportant" point. The grammar and the organization categories were almost equally important with the means of 4.28 and 4.23 respectively. In other words, in the raters' opinions, content was the most important criterion of the rating. What came after were grammar, organization, and vocabulary. Mechanics was the least important category.

IV. Discussion and Conclusions

The analysis of the mean frequencies of the strategies demonstrates that most of the strategies (30 among 36) were seldom used or were used at half of the time. One major reason of this phenomenon, as some raters explained, was that the raters used various strategies for compositions with different qualities. Therefore, it was almost impossible to have one strategy suitable for a majority of the compositions. However, for the strategies not applied to grading individual compositions, we found higher frequencies. For example, the mean frequencies of the strategies used to establish personal grading systems like those from S-9 to S-15, almost reached the point of "most of the time." In particular, the mean frequency of S-15, stating a behavior of rating the representative compositions of each level to know if their scoring systems were close to the provided standards, amounted to the highest 4.40. The high frequencies of the

strategies suggest that the grading guides provided by the JCEE Committee were not only necessary but also helpful to the raters to decide their scoring systems.

The frequency analysis also demonstrates that the time pressures the raters were experiencing in the rating process had effects on the use of the strategies. First, under time pressures, the raters tended to adopt the holistic method. According to their reports, they would give a mark based on the overall impression they got after the first reading, unless they encountered unusually good or bad compositions. In such cases, they would reread the compositions before they made judgment. The other effect of the time pressures was that the raters rarely turned to social resources in their rating process, because it might disturb their concentrating on their ratings. Despite the possible disadvantage, some raters still strongly recommended to achieve inter-rater reliability through peer socialization.

The analysis of the strategies used to evaluate the five scoring categories indicates that the communicative competence has become the focus of assessing writing ability in the JCEE English compositions. According to the mean frequencies, the raters tended to consider first the categories of content and organization rather than those of grammar and vocabulary to decide the overall writing quality. Furthermore, they tended to tolerate minor errors of punctuation, plural forms and the use of articles, if they did not hinder communication. This communication-oriented tendency will certainly bring some backwash effects on the English language teaching.

The second part of our analysis was to investigate the inter-rater reliability of our raters. For the use of the rating strategies, the results indicate that there were no significant differences among the raters, although they varied in their majors and experiences of teaching composition. For the 35 strategies, there was no statistic evidence showing that the two factors significantly influenced the way the raters used the strategies. The only affected strategy was influenced by the interaction effect of the two factors and it involved only four raters in two groups. Therefore, we may infer that, over all, there was no significant difference of the use of the strategies among the raters.

As for the criteria analysis, the results also indicate that there was no significant differences among the raters in determining the weights of the five scoring categories. One thing worth noticing was that the mean weights attached to the categories, as displayed in Table 3, were almost equal to the grading standards provided by the JCEE Committee. They reached the agreement that content, organization, and grammar played heavy roles in determining the overall quality of the compositions, and mechanics played the least important role. This

On The Rating Strategies Used By Raters Of The JCEE English Compositions: A Pilot Study

finding suggests that the grading guides have been observed and they are, thus, helpful to enhancing inter-rater reliability. Since there are practical difficulties of training raters for such a large-scale scoring situation, it is desirable to have comprehensive, applicable grading guides for the raters. Furthermore, because different types of writing tasks like free composition and composition with topic sentences, require specific guides, it would be most satisfactory, if the grading guides are to be reinterpreted every year for new topics.

In conclusion, this study has provided empirical accounts for some aspects of the complex rating process and for the inter-rater reliability of our raters. We identified and described the use of 36 strategies. Although this is by no means a comprehensive list, it is hoped that this investigation can cast a light on the complex rating process and raise the interest in this topic. The statistic results of the analysis concerning the reliability of the raters' strategies and criteria do not present a complete description, because only two factors were considered. Many additional factors like prior scoring experience, and their belief about good writing quality may play important roles in the rating process. Therefore, the results of this study are tentative. Further investigation of other relevant factors and their interaction is needed to have better understanding of the rating process.

Acknowledgements

I am indebted to Professor Tsan-sui Huang for his advice on the construction of the questionnaire. I also greatly appreciate helpful suggestions from professors H.Y. Chen, W.P. Lee, Y.M. Yin, Peace Feng, H.E. Ling, P.L. Hiao, and H.F. Cheng on the identification and description of the strategies. Special thanks are due to the raters who patiently filled in the questionnaire under great time pressures and offered valuable comments on this study.

Notes

1. The researches have proposed different descriptions of rating strategies. Homburg (1984) presents a "funnel model," which indicates that raters give scores on the basis of one major feature first, and then further discriminate the compositions on the basis of other features. Cumming (1990) identifies 28 interpretation and judgment strategies that operate interactively in the rating process.

2. Whether the raters had prior experience of scoring the JCEE papers could be a critical factor affecting the raters' behaviors. Although we did not analyze the effect of this factor in the present study, we believe an investigation on it would gain insights from another aspect.
3. The standard deviations show that the time discrepancy among individual raters did not vary greatly, as the raters got familiar with the job (SD=0.602 at first and SD=0.487 later).
4. As mentioned in Section II, the strategies identified in this questionnaire covered six major topics. Because of space limit, this study will merely cover parts of the strategy analysis.

References

- Chen, Kun-tien, T.S. Huang, S.F. Su, C.Y. Lin, S.E. Kang, and S.E. Lin. 1992. A Study of the Construction and Scoring of the JCEE English Writing Tasks. Taipei: The ROC Center of College Entrance Examination.
- Chen, Kun-tien, T.S. Huang, S.E. Lin and C.Y. Lin. 1993. A Study of the Construction and Scoring of the JCEE Composition Tasks. Taipei: The ROC Center of College Entrance Examination.
- Cumming, Alister. 1990. "Expertise in Evaluating Second Language Compositions." *Language Testing* 7(1), 31-51.
- Diederich, Paul, John French, and Sydell Carlton. 1961. "Factors in Judgments of Writing Ability." *Educational Testing Service Research Bulletin* RB-61-15. Princeton, New Jersey: Educational Testing Service.
- Heaton, J.B. 1990. *Writing English Language Tests*. New York: Longman Inc.
- Homburg, T. 1984. "Holistic Evaluation of ESL Compositions: Can it be Validated Objectively?" *TESOL Quarterly* 18(1), 87-107.
- Huang, Tsan-sui. 1994. A Qualitative Analysis of the JCEE English Tests. Taipei: The Crane Publishing Co., Ltd.
- Huang, Tsan-sui, Vincent W. Chang, S.Y. Gu, W.C. Yang, and K.T. Chen. 1994. *The Annual Report of the English Test Research Team*. Taipei: The ROC Center of College Entrance Examination.
- Hughes, Arthur. 1994. *Testing for Language Teachers*. New York: Cambridge University Press.
- Kaczmarek, Celeste M. 1980. "Scoring and Rating Essay Tasks." In *Research in Language Testing*, John W. Oller, Jr. and Kyle Perkins (Eds.), 151-159. Rowley, Massachusetts: Newbury House Publishers, Inc.

用字，拼字

格式，標點，大小寫

— — — — —
— — — — —

二、評分策略

以下是閱卷過程中可能使用的策略，請您在五個等級中（總是，時常，普通，很少，不會）勾選一項，來說明您使用該策略的情形。

	總是	時常	普通	很少	不會
1. 按聯招會評分項目（如內容，組織，文法等）逐一給分，再加起來。	—	—	—	—	—
2. 大概看完一遍，就可以給分了。	—	—	—	—	—
3. 看完第一段，就可以知道該試卷可得幾分。	—	—	—	—	—
4. 看開頭，和結尾就可以知道該試卷可得幾分。	—	—	—	—	—
5. 先快速看一遍，再細看一遍或數遍。	—	—	—	—	—
6. 計算和主題相關句子的多寡。	—	—	—	—	—
7. 歸納該作文主要論點有那幾點。	—	—	—	—	—
8. 找出該作文的開頭，發展，結尾等結構。	—	—	—	—	—
9. 給分時會換算成，以一百分為滿分的分數。	—	—	—	—	—
10. 將試卷與聯招會提供之樣卷比較，以決定得分。	—	—	—	—	—
11. 和已閱過的作文比較，以檢測自己給分標準是否一致。	—	—	—	—	—
12. 閱卷前，先設定幾個等級（如，上中下三級）的分數。	—	—	—	—	—
	總是	時常	普通	很少	不會
13. 閱卷前，先思考該作文題，主要的論點或內容有那些。	—	—	—	—	—
14. 設定一個高中畢業生應該具備的寫作能力，以決定給分標準。	—	—	—	—	—
15. 先試閱樣卷，以瞭解自己的評分標準和聯招會標準相差多少。	—	—	—	—	—
16. 先依內容，組織給分。再視其文法及用字，斟酌加減分數。	—	—	—	—	—
17. 先依文法，用字給分。再視其內容及組織，斟酌加減分數。	—	—	—	—	—
18. 逐句分析文法是否正確。	—	—	—	—	—
19. 計算文法錯誤有多少。	—	—	—	—	—
20. 注意其句型是否有變化。	—	—	—	—	—

**On The Rating Strategies Used By Raters Of The JCEE English Compositions:
A Pilot Study**

	總是	時常	普通	很少	不會
21. 估計文法及用字均正確的句子多寡或比例。	—	—	—	—	—
22. 標點及拼字錯誤如果不多，且不影響文意，則不扣分。	—	—	—	—	—
23. 標點及拼字錯誤如果不多，且不影響文意，仍酌予扣分。	—	—	—	—	—
24. 名詞單複數變化，和冠詞的使用，如果錯誤不多，且不影響文意，則不扣分。	—	—	—	—	—
25. 名詞單複數變化，和冠詞的使用，如果錯誤不多，且不影響文意，仍酌予扣分。	—	—	—	—	—
26. 動詞時態的使用，如果錯誤不多，且不影響文意，則不扣分。	—	—	—	—	—
27. 動詞時態的使用，如果錯誤不多，且不影響文意，仍酌予扣分。	—	—	—	—	—
28. 長度比規定長度短的作文，酌予扣分。	—	—	—	—	—
29. 超過規定長度時，酌予扣分。	—	—	—	—	—
30. 字跡亂的作文，酌予扣分。	—	—	—	—	—
	總是	時常	普通	很少	不會
31. 給作文分數時，參考中譯英分數。	—	—	—	—	—
32. 兩段分別計分。	—	—	—	—	—
33. 遇到語意不明的句子時，就跳過去。	—	—	—	—	—
34. 遇到不知如何給分的試卷時，會和別人討論。	—	—	—	—	—
35. 遇到不確定的文法或用字時，會去查字典。	—	—	—	—	—
36. 遇到不確定的文法或用字時，會和別人討論。	—	—	—	—	—
* 說明或建議：					