A Comparative Study on the Validity of Proofreading in NMET & CET-6 in China

Yu Shulin and Huang Jianbin

Abstract:

Based on the statistical analysis of proofreading tasks in CET-6 and NMET from 1996-2008, this study aims to examine whether "discourse context" plays a role in proofreading as expected with reference to Li Xiaoju's（1997）four levels of test points. Findings from our exploratory analysis show that although "discourse context" plays a significant role in proofreading, proofreading items with test points at the discourse level from 1996 to 2008 are really not enough to ensure the validity of the test or show the merits of proofreading. The relevant results are important for testing experts when they design proofreading items.

Key terms: proofreading, test points, validity, CET-6, NMET

1. Introduction

Proofreading has been widely used in English proficiency tests in recent years in China. It is a particularly common occurrence in some large-scale English proficiency tests like TEM-8 (Tests for English Majors) and CET-6 (College English Test for Non-English Majors) as well as NMET (National Matriculation English Test). It develops from sentence error-correction but differs from the latter in the discourse context. "Discourse context" distinguishes proofreading from sentence-error correction and gives prominence to the superiority of proofreading with a comprehensive understanding of the whole passage as a precondition. In proofreading, test takers are required to identify and correct mistakes in a certain passage. In NMET, there are about 100 words in the passage where there are 9 mistakes that need to be identified among 10 items while there are about 250-280 words in CET-6 and 10 mistakes are distributed across all the 10 items. Proofreading therefore involves discourse comprehension which requires test takers' high English proficiency.

Oller's Unitary Competence Hypothesis has claimed that proofreading is a valid measure of students' comprehensive English proficiency. What Oller (1979) maintained, in brief, was that language proficiency is indivisible, that tests only differ in their effectiveness at measuring this one factor, and that the elaborate apparatus and dimensions of tests used by psychometrists could be replaced by one test which would directly tap the single indivisible faculty of overall proficiency. Proofreading is such a test that can reflect this unitary ability of test takers, considered as the most accurate and valid test type (Li Xiaoju, 1997).

However, the past few years witnessed inadequacy in proofreading despite its popularity and reputation in some large-scale English proficiency tests. This paper attempts to contribute more to proofreading research from the perspective of test point validity.

2. Literature Review

2. 1 Test Point Validity

Test point validity is proposed by Li Xiaoju (1997), a Chinese scholar in language testing, for cloze and proofreading in papers testing the participants' comprehensive proficiencies. For proofreading, test point validity includes the layers and the focuses of test points (Figure 1).

The layers of test points could be subdivided into four levels: word level (W), phrase level (P), sentence level (S) and discourse level (D). A test point at the level of discourse can test both the discourse comprehension and the knowledge at the levels of sentences, phrases and words, whereas low-level test points cannot involve high-level ones because there is no context involved in these test points. As a result, the validity of the test depends a great deal on the level of test points.（Li Xiaoju, 1997）

Layers of Test points　　Focuses of Test points

D = discourse S = sentence P = phrase W =word

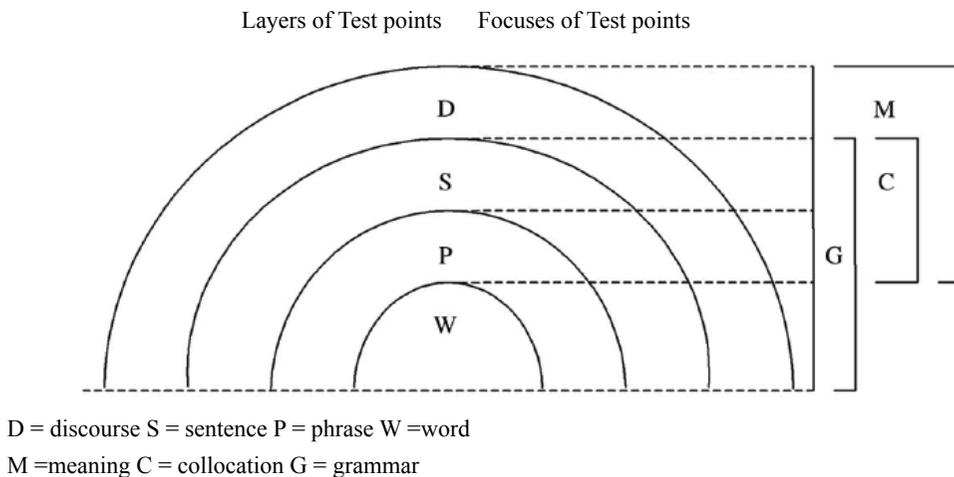M =meaning C = collocation G = grammar

Figure 1 The Layers and The Focuses of Test Points in Proofreading

The focuses of test points in proofreading are mainly classified into factors of grammar (G), collation (C) and meaning (M). It is generally considered that test points at the discourse level focus on meaning and test points at word level focus on the factor of grammar, while test points at the sentence and phrase level may focus on grammar, collation and meaning. Thus, test points of higher layers are usually relevant to meaning whereas those of lower layers focus on grammatical knowledge. The fundamental difference between proofreading and sentence error-correction lies in the fact that successful achievement in the part of proofreading must be based on the understanding of the whole discourse rather than single sentences. Therefore, much emphasis should be laid on test points of the discourse level in order to test participants' comprehensive proficiencies. Only in this way could the merits of proofreading be brought into play in language testing.

2. 2 Previous Researches into Test Point Validity and Proofreading

Previous research into test point validity and proofreading mainly focused on whether proofreading is a more valid and reliable integrative instrument than sentence error-correction in measuring overall proficiency. Statistical and theoretical analyses were done of proofreading in respect of test point validity. Yang Zhonglin (1994) concluded that items designed in sentence error-correction mainly reflect test takers' grammatical ability while items designed in proofreading can examine test takers' various abilities beyond grammatical ability. Liu Jianda and Yang Manzhen (2001) found that the context indeed plays an important role in the proofreading

test. In their studies, quantitative results demonstrated that the proofreading and error correction test correlated with the reading comprehension test and the writing test, but it did not correlate with the grammar test. Besides, it demonstrated that the proofreading test evaluated test-takers' language abilities beyond their grammar knowledge and showed that the proofreading test could test knowledge on the levels of words, phrases, sentences, and discourses.

Nevertheless, some research has revealed that proofreading has not been appropriately employed in test papers due to the frustrations of test item design. Han Jingfeng (2007) conducted qualitative research on the validity of proofreading in TEM8 and the results showed that grammatical mistakes were the most frequent accounting for more than half of the total number of errors. The analysis indicated that proofreading in TEM-8 focused on grammar and attached little importance to discourse comprehension. Moreover, discourse comprehension is restricted within small groups of sentences instead of the whole passage. Wen Hui and Nie Jianzhong (2008) made a study of test point validity analysis of proofreading in CET-6. It is finally proposed that the layers of test points in proofreading should be upgraded to improve the validity of the proofreading test in CET-6.

Although much work has been done on the validity of proofreading in CET-6 and TEM-8, few studies have been conducted to analyze the proofreading in NMET. Moreover, Wen Hui and Nie Jianzhong merely selected a few papers of CET-6 from 1989-2004 so that their results were not so persuasive. In the present study we attempted to conduct a comparative study of the test point validity of proofreading in NMET & CET-6 in China in order to present the role of discourse in proofreading. It touched on three specific questions: 1) To what extent does context play a role in the proofreading test, especially in NMET and CET-6? 2) Can proofreading be used to assess test-takers' language abilities on each language level, particularly on the discourse level? 3) Is the test point validity of proofreading in NMET different from that in CET-6? It is hoped that the results of this study may be useful to the design of proofreading assessments in test papers.

3. Material and Procedure

Materials to be analyzed were taken from the test papers of NMET and CET-6 from 1996 to 2008. They were analyzed according to four levels of test points.

The following is the passage for CET-6 in 2003:

It must reflect that diversity with their news coverage or risk (S1) losing their readers' interest and their advertisers' support.

Operating within Seattle, which has 20 percents racial (S2) minorities, the paper has put into place policies and procedures for hiring and maintain a diverse workforce. The (S3) underlying reason for the change is that for information to be fair, appropriate, and subjective, it should be reported by the (S4) same kind of population that reads it.

A diversity committee composed of reporters, editors, and photographers meets regularly to value the Seattle Times' (S5) content and to educate the rest of the newsroom staff about diversity issues. In an addition, the paper instituted a content (S6) audit that evaluates the frequency and manner of representation of woman and people of color in photographs. (S7)

Early audits showed that minorities were pictured far too infrequently and were pictured with a disproportion

ate number of negative articles. The audit results from (S8) improvement in the frequency of majority representation and (S9) their portrayal in neutral or positive situations. And, with a (S10) result, the Seattle Times has improved as a newspaper.

The diversity training and content audits helped the Seattle Times Company to win the Personnel Journal Optimal Award for excellence in managing change.

S1. it → they S   S2. percents → percent W
S3. maintain → maintaining P   S4. subjective → objective D
S5. meets → meet P   S6. "an" should be deleted P
S7. woman → women P   S8.   from → in D
S9. majority → minority D   S10.with → as P

As in the above example of CET6 in 2003: S2 is related to W; S3, S5, S6, S7 and S10 are related to P; S1 is related to S; S4, S8 and S9 are related to D.

4. Results and Analysis

The results of the analysis of the proofreading items in CET-6 and NMET from 1996-2008 are presented in Tables 1 and 2. A Chi-Square test revealed that the four levels of test points of proofreading items are different from each other in both CET-6 and NMET. (P=.000<.01) In CET-6, the frequencies of phrase (35.71%) and sentence (34.29%) levels are obviously larger than those of word (8.57%) and discourse (21.43%) levels. In NMET, the frequencies of phrase (30.51%) and sentence (30.51%) levels are the same and they are also apparently higher than those of word (19.49%) and discourse (19.49%) levels.

Table1. Frequencies of Four Levels of Test Points of Proofreading Items
in CET-6 from 1996 to 2008.

Paper	Word	Phrase	Sentence	Discourse
1996.1	0	3	4	3
1996.6	1	4	4	1
2000.1	1	4	2	3
2000.6	0	3	3	4
2001.6	1	2	2	5
2002.1	0	3	4	3
2002.6	1	1	5	3
2003.6	1	5	1	3
2003.12	3	3	3	1
2004.6	0	4	5	1
2005.1	1	5	4	0
2005.12	1	5	4	0
2006.6	1	5	2	2
2006.12.	1	3	5	1
Total	12	50	48	30
Percentage	8.57%	35.71%	34.29%	21.43%

Table2. Frequencies of Four Levels of Test Points of Proofreading Items
in NMET from 1996 to 2008.

Paper	Word	Phrase	Sentence	Discourse
1996	3	3	2	1
1997	0	4	4	1
1998	5	1	2	1
1999	3	1	2	3
2000	1	3	4	1
2001	1	4	3	2
2002	1	3	3	2
2003	2	3	3	1
2004	2	4	2	1
2005	0	3	3	3
2006	1	1	4	3
2007	2	2	3	2
2008	2	4	1	2
Total	23	36	36	23
Percentage	19.49%	30.51%	30.51%	19.49%

Figure 1 shows the comparison of the total percentages of test points of proofreading items from 1996 to 2008 in CET-6 and NMET. The frequencies of test points at the phrase, sentence and discourse levels in CET-6 are high relative to those in NMET, while the frequencies of test points at the word level is much lower than those in NMET. Moreover, test points at the discourse level only account for a small percentage of all the test points in both CET-6 and NMET.
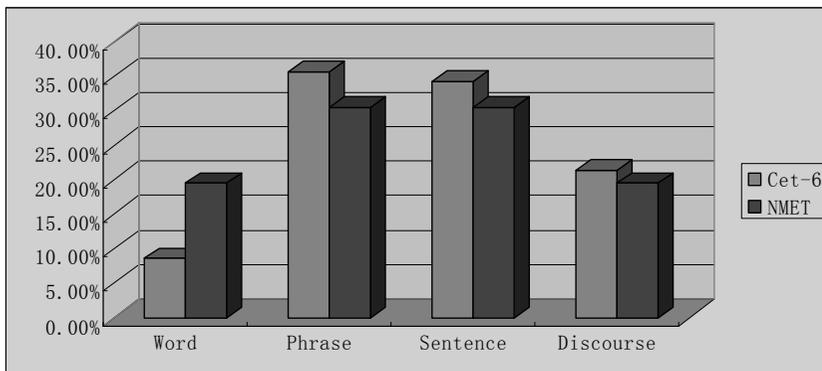

Figure1. Total Percentages of Each Level of Test Points of Proofreading Items from 1996 to 2008 in CET-6 and NMET.

The frequencies of word level and phrase level are relatively stable and balanced, but the sentence and discourse levels varied a lot from 1996-2008 in both CET-6 and NMET. Figure 2 shows the trends and changes of test points of the discourse level of proofreading items in CET-6.The proportion of test points of the discourse

level during 1996-2002 is much higher than that of 2002-2006. It is worth noticing that the test points of the discourse level of proofreading items in CET6 in recent years are quite low.
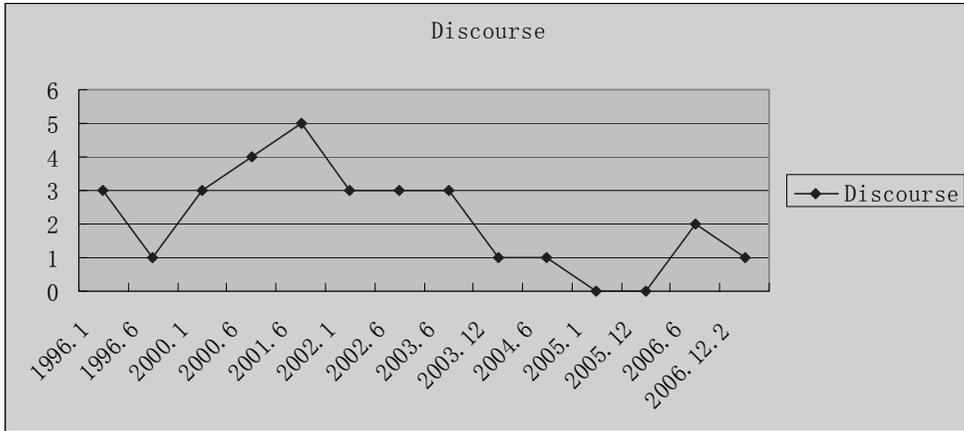

Figure2. The Trend of Test Points of Discourse Level of Proofreading Items from 1996 to 2008 in CET-6

Figure 3 shows the trends and changes of test points of the discourse level of proofreading items in NMET. The proportion of test points of the discourse level during 2004-2008 is much higher than that of 1996-2003.
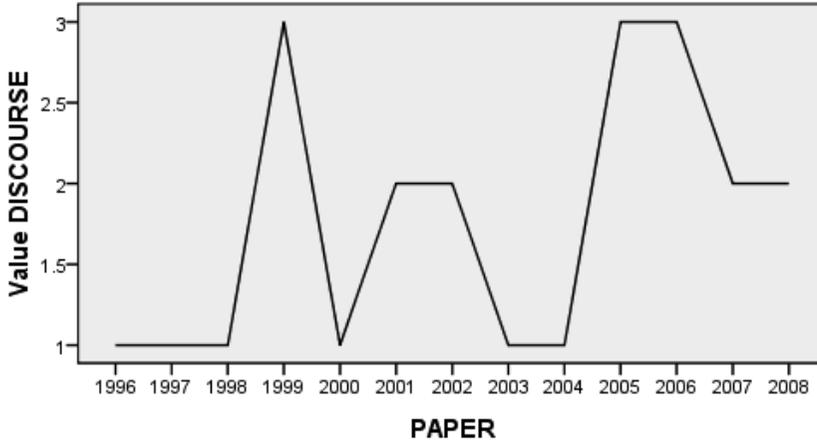

Figure3. The Trend of Test Points of Discourse Level of Proofreading Items from 1996 to 2008 in NMET

5. Discussion

Through the analysis of the four levels of test points of proofreading items both in CET-6 and NMET from 1996-2008, we may come to the conclusion that "discourse context" has played a significant role in proofreading. Some mistakes could not be identified and corrected without a good understanding of the whole passage. The outstanding characteristic of proofreading lies in its capability of testing the comprehensive proficiencies of language learners in a certain context. Therefore, for the setting of test points, test designers should make good use of the context in order to enhance the validity of the test. In addition, a test point at the level of discourse can test both discourse understanding and knowledge at the levels of sentences, phrases

and words, whereas low-level test points cannot involve high-level ones. Therefore, the higher the level of test points, the higher the validity of the test. It is probably impossible to set all the test points at the discourse level, but the key to enhancing the validity of the test is to properly increase the percentage of test points at the discourse level.

However, another significant part lies in the fact that the discourse level only takes up a small percentage of all the levels of test points. It seems that proofreading has not been expected to test the participants' language proficiency through understanding the whole passage. Thus, it is pretty obvious that the number proofreading items with test points at the discourse level is really not sufficient to ensure the validity of the test or show the merits of proofreading.

As for question 2 "Can the proofreading test the test-takers' language abilities at each language level, particularly at the discourse level?", it has been found that proofreading in both CET-6 and NMET could test test-takers' language abilities on each language level such as word, phrase, sentence and discourse levels. Nevertheless, much emphasis has been laid on checking the basic language knowledge, in particular the grammatical knowledge and less attention has been paid to discourse analysis. In CET-6, test points at the discourse level only constitute 21.43% and a smaller percentage-- 19.49% in NMET.

With regard to question 3, "Is the test point validity of proofreading in NMET different from that in CET-6?" shown from the comparison of the total percentages of Test Points of Proofreading Items from 1996 to 2008 in CET-6 and NMET, the frequencies of test points at phrase, sentence and discourse levels in CET-6 are relatively higher than those in NMET, while the frequencies of test points at the word level is much lower than those in NMET. On the other hand, what they have in common is that test points of the discourse level only take a small percentage of test points of all the levels in both CET-6 and NMET.

Another noteworthy aspect of the study lies in the fact that the frequencies of both the word level and phrase level are relatively stable and balanced, but the sentence and discourse levels vary a lot from 1996-2008 in both CET-6 and NMET. It is worth noticing that test points of the discourse level of proofreading items in CET6 in recent years are significantly lower, while the proportion of test points of the discourse level during 2004-2008 is much higher than that of 1996-2003 in NMET. It is safe to conclude that experts and scholars in language testing gradually realize the importance of discourse analysis in proofreading when they design proofreading items in NMET. While in CET-6, from the year of 1997, when Li Xiaoju put forward the theory of four levels of test points, to the year of 2002, the level of test points in proofreading rose slightly and steadily. What confuses us is the sharp drop in the third period from 2003 to 2006. In the course of seeking the reason for this, what enlightens us is the adoption of the new test type in CET-6. The section named Vocabulary and Grammar has not been included in the new CET-6, so the assessment of vocabulary and grammar may be transferred into the section of proofreading or cloze, which may consequently increase the weight of test points at the word level, the phrase level and the sentence level. To our disappointment, it was found that the new CET-6 did not take place until 2006. Thus, the assumption above could not account for the sharp drop of test points at the discourse level in the third period from 2003 to 2006. So we have to leave this question to further investigation and future researchers.

Another point we need to be aware of is that, according to current national English exams, there is a proofreading task in CET-6 but not in CET-4, in TEM-8 but not in TEM-4. It probably shows that proofreading is more appropriately adopted in language tests for learners with high proficiencies. Hence, is it reasonable to employ proofreading in NMET, a test in English for national entrance examinations for universities and colleges? Furthermore, although proofreading is widely used in some large-scale English proficiency tests in China, it is a rare occurrence in language tests in other countries and only some tasks similar to proofreading could be found in language tests abroad such as TEEP (Test of English for Educational Purpose), UETESOL (University Entrance Test in English for Speakers of Other Languages) in England and TE (Test of English) in Australia.（Liu Jianda and Yang Manzhen , 2001）The reasons underlying these differences are expected to be further explored.

6. Conclusion

Proofreading in CET-6 and NMET is regarded as a valid measure of candidates' comprehensive English language ability as introduced in their specifications. The present study revealed that "discourse context" has played a role in proofreading but the proportion of proofreading items with test points at the discourse level from 1996 to 2008 is not high enough to ensure the validity of the test and show the merits of proofreading. Furthermore, distributions of test points of proofreading items from 1996-2008 are unbalanced and unstable, particularly in CET-6, where test points at the discourse level are even lower in recent years, which should be taken into consideration when testing experts design proofreading items in CET-6.

There is no doubt that context is of utmost importance. The real confusion lies in whether higher level test points could validly measure test-takers' comprehensive language ability. But to what extent context plays a role and how to determine its weight in exams is hard to do accurately. As to how much weight the test point at each level of should account for in proofreading, it's an open question that needs to be further researched. What is certain is that a proper share of items, which can be used to test discourse level, should be advocated. It is hoped that more attention will be directed to proofreading research and that the findings obtained from this study could help to improve the design of proofreading items so that proofreading may be adopted by other large-scale language tests both in China and other countries of the world.

References

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 161-170.

Gui Shichun. (1986). Standardized tests–theories, principles and methods. Guangzhou: Guangdong Higher Education Press.

Han Jingfeng. (2007). The analysis of the validity of proofreading in TEM-8. *Journal of Xinzhou Teachers University,* 6, 107-108.

Li Xiaoju. (1997). *The science and art of language tests*. Changsha: Hunan Educational Press.

Liu Jianda & Yang Manzhen. (2001). What does proofreading test? *Journal of Modern Foreign languages*, 2, 170-180.

Oller, J. W. 1979. *Language tests at school: a pragmatic approach.* London: Longman.

Wen Hui & Nie Jianzhong. (2008). The analysis of test points validity of proofreading in CET-6. *Journal of*

Taiyuan University of Science and Technology, 2, 148-150.

Yang Zhoulin. (1994). Proofreading and language tests. *Journal of Foreign language teaching and research,* 3, 37-42.

Zhou Changyin & Li Yuliang. (2000). The analysis of proofreading in language tests. *Theory and Practice of Higher Education*, 2, 80-83.