ÜNİVERSİTE ÖĞRENCİLERİNİN YABANCI DİLDE YAZILMIŞ METİNLERİ ANLAMA BECERİLERİNİN SINANMASINDA ÖZET YAZMA TEKNİĞİNİN KULLANILMASI

THE USE OF SUMMARY TASKS IN TESTING UNIVERSITY STUDENTS' COMPREHENSION OF FOREIGN LANGUAGE TEXTS

Yrd. Doç. Dr. Berrin UÇKUN
Gaziantep Üniversitesi
Fen-Edebiyat Fakültesi

Okutman A. Pınar AKSOY
Gaziantep Üniversitesi
Fen-Edebiyat Fakültesi

ABSTRACT

The use of multiple choice questions in testing students' reading comprehension of foreign language texts has become a widespread practice. However, not only does this technique ignore and/or override most of the reading skills that we train our students in but it also establishes in them the habit of reading not for total comprehension but only enough to answer the questions in the test. The validity of any test will be achieved only to the degree that it tests our students' ability in the manner it has been taught in the classroom, not to mention that it will reflect on what is taught and how it is taught in the classroom (backwash effect). Considering the fact that reading ability consists of certain strategies and sub-skills which interact with each other, we need to use a testing technique which can capture the complexity of the process. As an example of such a technique, summarization was used with university students following their reading of two texts - expository and narrative - and results from this technique were correlated with scores gained from a multiple choice test constructed for the same two texts. Based on the results of the study, we conclude that summarization can be a reliable way of testing reading comprehension, and to use this technique more effectively it is suggested that students receive prior training in summarization.

ÖZET

Yabancı dilde okuma-anlama becerisinin sınanmasında kullanılan çoktan seçmeli tekniğinin kullanımı fazlasıyla yaygınlaşmasına karşın, bu teknik okuma-anlama sürecinde öğrencide geliştirilmesi amaçlanan becerilerin birçoğunu ölçmemekle birlikte, öğrenciyi okuduğunu tümüyle anlamaya değil de soruyu cevaplamaya yönelten bir okuma alışkanlığını yerleştiriyor. Kullandığımız test tekniğinin geçerliliği öğrenciye kazandırmak istediğimiz becerileri ölçmesi oranında sağlanacaktır ve aynı zamanda bu teknik öğretim programlarının içeriğini belirlemede etkili olacaktır. Okuma-anlama becerisinin birçok strateji ve alt becerilerin karşılıklı iletişiminden oluşması nedeniyle, bunu sınavlara yansıtabilecek bir sınama tekniği olarak özet çıkarma üniversite öğrencileri üzerinde iki ayrı türde "öykü ve düzyazı" okuma parçası kullanılarak denendi ve sonuçlar aynı okuma parçaları için geliştirilmiş çoktan seçmeli sınav sonuçları ile karşılaştırıldı. Toplanan verilerden elde edilen sonuçlar özet yazmanın okuma-anlamayı ölçmeye yönelik kullanılabilirliğine ve bu tekniğin daha verimli kullanılabilmesi için öğrencilerin bu yönde eğitilmelerinin gerekliliğine dikkat çekmektedir.

Introduction

Summarizing tasks in reading comprehension tests are rightfully appealing to teachers and test constructors in this era of communicative language testing because such tests simulate real-world tasks in which non-native readers have to read and write a summary of the main idea of the text. This measure is more compatible with the comprehension requirements of university students who have to read academic material. Kintsch and van Dijk (1978) have argued that the macrorules involved in successful summarization are similar to those underlying successful reading comprehension. These researchers postulate that during comprehension readers abstract the macrostructure (i.e. the gist) of the text from the available text microstructure, which comprises of the individual propositions and their relationships. Readers accomplish this by using a set of comprehension macrorules. These are basically reduction and abstraction rules:

1. Selection rules which consist of either keeping some nonredundant or relevant text propositions unchanged or deleting irrelevant propositions;

2. Generalizing propositions by inferring a suitable superordinate proposition; and

3. Constructing a proposition that represents several text propositions given some knowledge-base that can be used to reconstruct them.

Brown and Day (1983) showed that there is a

developmental pattern in the order that these rules are acquired by readers of all ages. Constructing a proposition is the most difficult for readers, since it involves invention of missing topical information, while a copy-delete strategy is the easiest. Once comprehension macroprocesses have been completed, production is straightforward and simply consists of copying out and (re)constructing the macrostructure created during encoding.

The quality of a written summary depends on the extent to which the original material to be summarized is comprehended. Therefore, we could expect the task demands of summarization to be closely related to the characteristics of the original text. Three text characteristics need to be examined with respect to this: length, genre and complexity.

The length of the original material seems to play an especially important role in determining what one must do to produce a good summary. The shorter the text, the more likely that the ideas are closely related and can be expressed by a single topic sentence. With longer texts, however, summarizing becomes more difficult because of the increase in the processing load as more evaluation and decisions have to be made (Hidi and Anderson, 1986).

The genre of the original material also seems to affect summarization. It has been argued that summarization of narratives are easier than expositions (Hidi and Baird, 1985). One explanation for this could be that we tend to have more experience with narratives than with expositions, which makes it easier to judge importance, notice inconsistencies, and condense ideas when working with such texts. Secondly, expositions usually deal with ideas which are more complex and less familiar to the summarizer. Thirdly, narratives tend to follow a temporal-causal course whereby the information is organized linearly. Expositions, on the other hand, have a less-organized and non-linear structure, which is more difficult to process. Finally, in narratives, the same part of the text tends to be both important and interesting, whereas in expositions importance and interestingness do not always correspond. Therefore, the text factor must be considered as an important influence on a subject's ability to summarize.

Text complexity is another factor which deserves full consideration. Text complexity involves such aspects of the text as low-frequency vocabulary, elaborate sentence structure, abstractness, unfamiliarity of concepts and ideas, and inappropriate or vague organization. Hidi (1984) has observed that, when handling complex texts, the majority of adult subjects adopt a paragraph -by-paragraph strategy to produce a summary rather than a whole-text synthesis of ideas.

One aspect of the task procedure which appears to affect the cognitive operations of the summarizer is the presence versus the absence of the target material. Having access to the material allows the summarizer to scan it repeatedly, which in turn reduces the direct memory load of the task and allows him/her to make further discriminations regarding the importance of ideas in the text and to chunk larger text units. When a writer has to summarize in the absence of the text, all propositions have to be retrieved from memory. The increased memory load could result in the reduction of the text for the wrong reason: simple forgetting rather than deliberate deletion, condensation, and integration of ideas.

A summarizing task elicits a wide range of reading strategies which are the focus of most reading instruction and testing in language learning programs. Therefore, it promotes a richer and more interactive approach to reading than other measures of comprehension. Cohen (1994), in a summarization study he carried out with Brazilian students, reported that his subjects underlined words they did not know, paid attention to cohesive devices and graphic cues, made generalizations, and drew on their world knowledge to connect the details with each other. He added further that respondents had little difficulty in identifying and selecting the topical information, but they found it much harder to distinguish between superordinate and trivial or redundant material.

This study proposes to investigate the possibility of using summaries as a means for assessing comprehension of foreign-language texts. This is to be achieved by comparing the statistical results from a summarizing task with results obtained from a multiple-choice type of test, which is a more established and traditional method of assessment. There is a need for the examination of university students' summarizing abilities since the abilities that a summary task promotes are the use of higher order reading skills, such as identification of main ideas and condensation of text while maintaining the focus of the original one. Within the framework of this study, four experimental hypotheses were posited:

1. There will be no statistically significant relationship between subjects' scores on the summarizing task and the multiple-choice test based on the narrative text.

2. There will be no statistically significant relationship between subjects' scores on the summarizing task and the multiple-choice test based on the expository text.

3. There will be no statistically significant relationship between subjects' scores on the summarizing task as determined by text type.

4. There will be no statistically significant relationship between subjects' scores on the multiple-choice test according to text type.

Method

Subjects

The subjects consisted of 25 students who were attending a one-year preparatory English program during the academic year of 1998-1999 in the Department of Foreign Languages of Gaziantep University, Turkey. Enrollment in the preparatory English program was compulsory for students of the Department of English Language and Literature if they scored less than 60 on the department's Exemption Test administered at the beginning of the year. The actual number of students registered in the program was 33; however, four students had dropped out of the program earlier and four students were absent on the days the study was conducted.

The subjects were more or less a homogenous group with respect to their linguistic ability since their language scores on the University Entrance Exam were within a narrow range (450-470). Secondly, being students of the English Language and Literature department, they were assumed to be instrumentally motivated. A third reason for keeping the number of subjects relatively low is due to the in-depth nature of this study. Each student's scores in the two texts and the two tasks had to the analyzed and evaluated in their cross relationship with each other. The data to be processed had to be kept at a reasonable amount. Finally, taking up four classroom hours of more than one group would have been impossible for administrative reasons.

Procedure

Two measuring tools were used in the study: a summary task of two different texts and multiple choice reading comprehension questions relating to the same texts. The summary task was administered without prior instruction to the students by one of the researchers during a single class hour. The summaries were to be limited to 250-300 words for the expository text and 200-250 words for the narrative text. The expository text was taken from an advanced level course book (Radley and Burke, 1994) with the unknown vocabulary provided in the back, while the narrative text was an unabridged short story by Saki (in Sachs, 1969), again with vocabulary given. The summary protocols were collected together with the texts at the end of the class hour. After a ten-minute break, students were given the multiple choice questions in the absence of the original texts, which took them 20-25 minutes to complete. The multiple choice questions were improvised following feedback from colleagues (several of whom are native speakers) who have had long-time experiences with the teaching and testing of reading skills.

Scoring Summaries

For the purposes of this study, we found it most convenient to use the scoring sheet developed by Johns (1985). Using Kroll's (1977) definition, we coded idea units in subjects' summary protocols rather than punctuated sentences because it was possible for sentences to contain two or more propositions.

The data to be analyzed was grouped under three main headings as

1. Essential idea units, which are idea units that should be included in a summary because the author of the original text probably considers them important (as determined by the judgements of the expert readers);

2. Non-essential idea units, which express either redundant or trivial information; and

3. Personal Comments, which are either comments on the reading itself or general observations created by the reading.

All the idea units that appeared under the Essential idea units category were once again categorized under two sub-headings as "Correct Replications" and "Distortions". A Correct Replication was either an accurate paraphrase of a single idea unit or direct copying of a single idea unit from the passage. Subjects were instructed to reconstruct the text in their own words, therefore not many instances of direct copying were expected. We made no distinction between paraphrasing and direct copying as separate categories, and hence both types of reproduction received 1 point per idea unit. Writer-invented statements, on the other hand, were idea units that expressed the gist of a paragraph or of the entire reading, or else a metastatement relating to the reading. Each reproduction of this sort received 2 points from the raters.

Under the subcategory of Distortions were included idea units whose noun phrase was appropriate to the original, but the verb phrase was deviant, or vice versa. Or these could be idea units from the reading, either copied or paraphrased, from which the essential information had been deleted. In combined idea units, the unit which contained accurate information gained points while the inaccurate unit did not. Distorted idea units were recorded for the benefit of determining idea units which were essential in capturing the main idea of the text but which were erroneous at the grammar level.

The non-essential idea units were determined according to the macrorules operating for action discourse in general and narrative discourse in particular (van Dijk and Kintsch, 1985). Accordingly, in an appropriate action description, the following types of propositions may in general be abstracted from:

1. Desciptions of reasons, purposes, and intentions for actions and the mental consequences of actions.

2. Descriptions of alternative possible courses of events.

3. Descriptions of auxiliary actions which are normal.

4. Descriptions of properties of states which do not condition further action.

5. Metadescriptions: propositions announcing, repeating, resuming or commenting other propositions.

6. Description of dialogue. (van Dijk and Kintsch, 1985:804)

The application of these rules yielded a macrostructure for both the narrative and expository passages. When parsed into idea units, the narrative text yielded 34 idea-units while the expository text yielded 38 idea units. Each summary protocol was coded according to this scale. A certain number of summary protocols were scored by both researchers until one researcher gained enough confidence to score the rest of the summaries.

Results

In this study, correlational research was conducted to evaluate the degree of relationship between students' ability to summarize a text of narrative and expository type and their success in answering reading comprehension questions relating to the same texts in a foreign language. To test the first two of the four hypotheses stated earlier, we need to check whether the subjects' scores on the summarising task correlated with their scores on the multiple choice questions, firstly for the narrative text and secondly for the expository text.

The results for both text types pointed to a statistically significant relationship between scores on the summarizing and multiple choice tasks (see Table 1). The null hypotheses were rejected for both cases. Therefore, the data gave support to the idea that the factors which determine students' performance on one measure are equally valid in determining performance on the other measure, irrespective of text type.

The third hypothesis queries whether subjects' performance on the summarizing task will differ according to text type. The null hypothesis is once again rejected in favor of a relationship existing between the same subject's summary scores on two different text types.

The fourth hypothesis puts forward a similar argument saying that the subjects' multiple choice test scores will not show any relationship to their scores on the other text. This null hypothesis was accepted because the calculated t-test value was lower than the value given to t. Clearly, the subjects' multiple choice scores for one text were inconsistent with their scores on the second text (see Table 2).

Tables 3 and 4 (see Appendix A) give a clearer picture of the information obtained from subjects' summary protocols for narrative and expository passages, respectively. When we examine the percentage of essential idea units which were correctly replicated by subjects, the mean for the narrative text was higher than that of the expository text (44.7 and 39.57, respectively), although the difference did not prove to be significant at .05. Distortions were slightly higher for the expository text. These values were also supported by the results of the multiple choice tests, whereby the mean score for the expository text was lower than the mean for the narrative text (54 and 72, respectively). Distortions, idea units

Table 1

Relationship between scores on the summarization and multiple choice tasks for two text types

Text Type	Task Type	Correlation	Significance	SD	t-value
Narrative	Summarization Versus Multiple choice Q	.28	.17	2.80	16.85 (.05)
Expository	Summarization Versus Multiple choice Q	.34	.09	7.31	9.76 (.05)

N = 25

Table 2

Relationship between scores on the expository and narrative texts for two task types

Task Type	Text Type	Correlation	Significance	SD	t-value
Summarization	Expository Versus Narrative	.41	.04	6.98	2.43 (0.5)
Multiple choice Q	Expository Versus Narrative	-.00	.99	1.76	-4.09 (0.5)

Table 3. Distribution of Idea Units for Narrative Text

NARRATIVE TEXT		SAMPLE	Adile	Elif	Safiye	Serap	Kübra	Özgül	Berna	Tuğba	Şirin	Hatice	Nazlı	Gülşen	Dilara	Sevilay	Ayşe	Eren	Gül	Emsal	Arda	Zehra	Naide	Ferha	Melte	Asu	Emre
Essential Idea Units	Correct Replications	34	14	17	18	14	19	17	18	15	9	14	17	16	12	13	18	14	18	10	11	17	17	15	17	19	11
		100	41,17	50	52,94	41,17	55,88	50	52,94	44,12	26,47	41,17	50	47,06	35,29	38,24	52,94	41,17	52,94	29,41	32,35	50	50	44,12	50	55,88	32,35
	Distortions	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	2
Personel Comments		0	0	2	4	1	6	2	7	3	3	4	3	4	1	1	11	3	0	6	6	4	14	4	4	5	4
Non-Essential Idea Units		0	11	15	17	8	11	29	30	25	25	9	25	19	28	34	18	10	15	28	10	24	24	24	34	25	32

Correct Replications Mean = 44,7 Distortions Mean = 0,32 Personal Comments Mean = 4,04 Non-Essential IU's Mean = 21,2

Table 4. Distribution of Idea Units for Expository Text

EXPOSITORY TEXT		SAMPLE	Adile	Elif	Safiye	Serap	Kübra	Özgül	Berna	Tuğba	Şirin	Hatice	Nazlı	Gülşen	Dilara	Sevilay	Ayşe	Eren	Gül	Emsal	Arda	Zehra	Naide	Ferhan	Meltem	Asu	Emre
Essential Idea Units	Correct Replications	47	12	11	10	11	35	14	30	14	14	19	19	9	18	22	27	22	16	11	5	24	31	25	25	19	22
		100	25,53	23,40	21,27	23,40	74,47	29,79	63,83	29,79	29,79	40,42	40,42	19,14	38,29	46,80	57,44	46,80	34,04	23,40	10,63	51,06	65,95	53,19	53,19	40,42	46,80
	Distortions	0	1	1	1	0	1	0	0	2	1	1	0	0	2	2	0	0	1	0	0	0	2	0	0	0	0
Personel Comments		0	1	0	0	2	1	0	4	3	2	0	0	0	5	0	0	3	0	5	0	0	0	1	0	3	15
Non-Essential Idea Units		0	24	2	19	18	9	22	18	10	1	12	16	14	14	17	17	16	2	21	29	16	17	25	19	30	20

Correct Replications Mean = 39,57 Distortions Mean = 0,6 Personal Comments Mean = 1,8 Non-Essential IU's Mean = 16,32

which reflect subjects' misunderstanding and/or misinterpretation of the text, were also evident in subjects' answers to multiple choice questions. Subjects tended to put in more personal remarks and more non-essential ideas, such as elaboration and details, into their summaries of the narrative text.

Conclusions

The statistical results obtained from this study may not appear to be highly significant but they do point to a direction which requires us to give more serious consideration to the use summarization as a method of testing subjects' comprehension of written texts and illustrating how the comprehended message is organized in the reader's mind. One of the conclusions suggested by this study is that there is a significant degree of agreement between subjects' scores on the multiple choice comprehension test and the summarization task irrespective of text type. The fact that "well-constructed summary tests promote a richer, more interactive approach to reading than do comprehension tests that focus more on details" has been attested by many studies (Cohen, 1994; Johns; 1985; Rinehart et al.; 1986). Subjects in this study had not received any formal training in summary writing (although they were frequently given oral summary assignments for reading). For the sake of developing tests which are more reliable, it is important that students receive training in writing summaries. The studies mentioned earlier confirm the metacognitive hypothesis that summarization training improves reading skill by heightening awareness of top-level information in texts, and that this kind of testing elicits a wide range of reading strategies.

Another finding of this study is that a subject's success on a summarization task for any text type could be predicted with some reliability whereas his success on a multiple choice test could not be predicted across different text types. The mean score for the multiple choice questions relating to the narrative text was significantly higher than the mean score for the expository text. This means that the same subject's success on such a test could be significantly higher if a narrative text is used rather than an expository one. This difference is insignificant in a summarization activity. We can also conclude from the SD values in Table 3 that multiple choice tests do not discriminate among subjects as well as the summarization task does. Also striking is that subjects tended to provide more personal comments for the narrative text, as they made inferences and personal evaluations of acts and people based on the given information. This could be the result of being taught to make critical evaluation of everything they read and to inject themselves into what they write.

In this study only two raters were involved in the assessment of the summary protocols and consensus was reached regarding the evaluation of responses in relation to the score keys and the ways of dealing with certain problems. In any case, rater reliability needs to be achieved through careful training of raters and a score key that lists the main ideas and connecting schemata need to be developed and followed rigorously. Given the problems of this kind and others, more research needs to be done on how people write summaries and how raters respond to them. It is important to remind ourselves that a test's function is not only to identify individuals as more or less proficient but also to create a positive backwash effect in determining what teachers should teach and what students should learn in the classroom.

References

Brown, A.L. and Day, J.D. (1983) "Macrorules for Summarizing Texts: The Development of Expertise". *Journal of Verbal Learning and Verbal Behaviour,* Vol.22: 1-14.

Cohen, A.D. (1994) "English for Academic Purposes in Brazil: The Use of Summary Tasks", in C. Hill and K. Parry (eds.), *From Testing to Assessment: English as an International Language*, London: Longman.

Hidi, S. (1984) "Summarization of Complex Texts" (Occasional Paper #4). Toronto: Ontario Institute for Studies in Education, Center for Applied Cognitive Studies.

Hidi, S. and Baird, W. (1985) "The Effect of Structural Revisions on Learning from School Texts". Paper presented at the meeting of the American Educational Research Association, Chicago.

Hidi, S. and Anderson, V. (1986) "Producing Written Summaries: Task Demands, Cognitive Operations, and Implications for Instruction". *Review of Educational research,* Vol. 56: 473-493.

Johns, A.M. (1985) "Summary Protocols of 'Underprepared' and 'Adept' University Students: Replications and Distortions of the Original". *Language Learning*, Vol.35: 495- 517.

Kintsch, W. and van Dijk, T.A. (1978) "Toward a Model of Text Comprehension and Production". *Psychological Review,* Vol. 85: 363-394.

Kroll, B. (1977) "Combining Ideas in Written and Spoken English: A Look at Subordination and Coordination" in A. Keenan and R. Bennett (eds.), *Discourse Across Time and Space,* Los Angeles: University of Southern California, S.C.O.P.I.L. #5.

Rinehart, S.D. and Stahl, S.A. and Erickson, L.G. (1986) "Some E Effects of Summarization Training on Reading". *Reading Research Quarterly,* Vol.21: 422-437.

van Dijk, T.A. and Kitsch, W. (1985) "Cognitive Psychology and Discourse: Recalling and Summarizing Stories" in H. Singer and R.B. Ruddell (eds.), *Theoretical Models and Processes of Reading,* 3rd ed., Newark, Delaware: International Reading Association.