

Task types in assessing reading comprehension

Feladattípusok a szövegértés mérésében

Az idegennyelv-tudás és az idegen nyelvi kompetenciák mérésekor számtalan különböző feladattípust használnak a vizsgáztatók és mérési szakemberek. A feladat típusának kiválasztása a nyelvtudáselméletekre alapszik, és különböző szempontok szerint történik: függ a mérés konstrukciójától, vagyis hogy milyen tudáskomponensek vagy részkészségek méréséről van szó, függ a feladattípus mérési alkalmasságától, valamint olyan gyakorlati megfontolásoktól is, mint a vizsgázók létszáma, vagy az értékelésre rendelkezésre álló idő. A szakirodalom számos összehasonlító kutatásról számol be a feladattípusok alkalmasságával kapcsolatban, az eredmények azonban ellentmondóak. A dolgozatban egy főiskolai hallgatók és tanárok körében végzett kérdőíves felmérés eredményeinek bemutatására kerül sor, mely a Budapesti Gazdasági Főiskolán végzett összehasonlító kutatás részét alkotta.

Factors affecting foreign language performance

In foreign language assessment it is common to rely on BACHMAN'S (1991) model of communicative language ability. As abilities and competences cannot be measured in any other way than by observing an individual's performance which is supposed to reflect the underlying competences, it is important to consider all the possible factors that, besides actual language competences, may influence performance. These contaminating factors include personal attributes, such as age and gender, test method facets, such as task type or native language use in testing, and random factors, such as weather and the examinee's emotional state. From these factors it is test method facets only that can be controlled and moderated by the examiner. In testing reading comprehension the test method facets include the reading text, the facets of the rubric and the task, dictionary use, native language use in task completion, and the type of the expected response. Selecting a task type for assessment is often based on competence models, and depends on the construct that is to be measured, on the assessment potential of a task type, and practical constraint such as the number of examinees and the time available for test administration and marking. Each task type may have a different potential, and may be suitable for testing different skills or elements of knowledge.

Comparative studies on task types

Several comparative studies have been conducted to investigate the effects of task types on reading comprehension performance (BACHMAN, 1985; CHAPPELLE & ABRAHAM, 1990; GORDON & HANAUER, 1995; RILEY & LEE, 1996, SHOHAMY, 1984; Wolf, 1993), but findings are controversial. BACHMAN (1985), for example, com-

* BGF Idegen Nyelvi és Kommunikációs Intézet, intézetvezető docens.

pared more than 900 students' performance on cloze tests with fixed-ratio and rational deletions. In the rational deletions test BACHMAN ensured that filling in the gaps require understanding different levels of context. He found that the two tests represented very different levels of difficulty for the examinees. CHAPELLE and ABRAHAM (1990) compared scores on fixed-ratio and rational deletions cloze tests, multiple-choice items, and a C-test. Besides the different test formats representing different difficulty levels, they found that the comprehension tests correlated with a proficiency test in varied ways, which suggests that the four tests did not measure the same construct. RILEY and LEE (1996) compared performance on two global tasks: a free recall task and a summary task. They found that the two task types elicited performance different both in organization and cohesion. They found significant differences in the number of main ideas and the number of details that the participants included in their responses. SHOHAMY (1984) and WOLF (1993) compared short-answer questions and multiple-choice items both in the native and the target language. Based on their findings it is clear that the items or questions in the task provide additional information for the reader that may help comprehension. The amount and the quality of this information may substantially differ in the case of different task types.

A questionnaire survey

A large-scale study was conducted to investigate task type effect on reading comprehension performance at Budapest Business School in 2004. More than 200 students and 50 teachers were involved in the research. Besides collecting and processing test data statistically, a questionnaire was also developed to explore students' and teachers' opinions on the suitability of task types for assessment.

Teachers' and students' opinions on task types

Fifty-one teachers and 207 students completed the questionnaire in which opinions were collected about native language use and eight different reading comprehension tasks. They were asked to rank order eight commonly used task types on a 1-5 suitability-for-testing scale. The list of task types included short-answer questions, multiple-choice items, true and false statement, matching, summary, free recall, gap-filling, and information transfer – all task types that can be applied in either the native or the target language.

The teachers' and the students' opinions varied considerably in this study (*Figure 1*). The SPEARMAN rank order correlation did not show any systematic relationship between their opinions ($\rho = -0.24$).

Students considered summary the best method, which corresponds to HELTAI'S opinion (2001, p. 13), and short-answer questions the second best. No other task types received an average score above 4.0 on the 1 to 5 scale. It is noteworthy that both task types include language production, which is a major source of worry for test developers who would like to exclude all contaminating factors from assessing

the receptive skill of reading. Besides practical constraints, this worry is a most important justification for employing recognition tasks such as multiple-choice items. It seems, however, that students themselves consider these task types authentic and reliable measures. The differences between the rank scores of other task types are usually tiny and insignificant; however, multiple-choice items are sixth only in the rank order. Information transfer has the lowest rank, which seems to indicate that some of the students feel uneasy about this task type. The group interview revealed, for example, that some students do not feel comfortable about filling in charts. This supports ALDERSON'S (2000) claim that information transfer may include a variety of cognitively different tasks, and thus might represent an additional difficulty in a test.

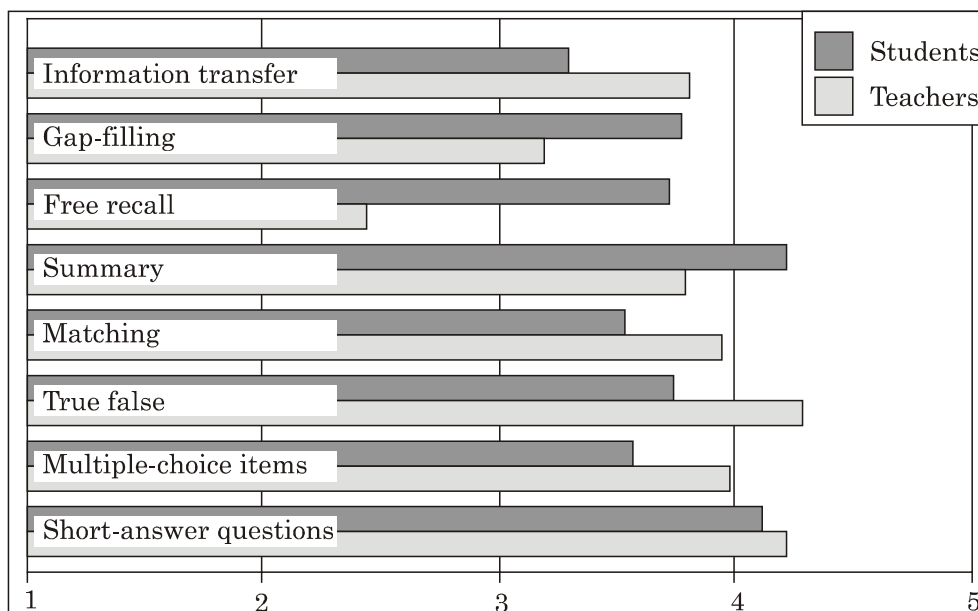


Figure 1
Suitability rating of task types by teachers and students

Apart from short-answer questions, which both teachers and students ranked very high, no other correspondences were found between their rank orders. It was particularly salient that teachers ranked the task types very low which involved extensive language production (summary and free recall) and seemed to prioritise recognition tasks such as true and false or multiple-choice items. Quite surprisingly, they ranked true and false statements highest among all the reading comprehension tasks. True and false statements can be regarded as the recognition version of short-answer questions in the sense that whereas short answer questions expect test takers to produce statements about a text, true and false items present the statements and require test takers only to recognize if they do or do not truly represent text meaning. If test takers did not guess when they judge a

statement, and skipped the items they cannot answer, this task type could be the ideal measure of reading comprehension. However, as students guess extensively – they are even encouraged to do so by their teachers - scores on a test will always involve a high risk of error in measurement (there is a 50 % chance of guessing correctly).

The sharp discrepancy between teachers' and students' opinions has raised doubts whether teachers and students interpreted 'suitable task' in the same way. It is assumed that, despite the explanations in the questionnaire introduction, for teachers suitability did not simply mean that the task can elicit well-measurable performance, but involved practical considerations as well. The assumption was supported by two of the teacher interviews in which the interviewees admitted they had considered task types' suitability in terms of larger-scale assessments, so they automatically rejected tasks such as free recall.

As for language use, neither L1 or L2 was considered universally better by the respondents. Opinions varied according to task types. Students' and teachers' opinions were basically the same: both groups agreed that tasks including language production might measure more reliably in L1, whereas they appreciated recognition tasks in L2 more. In general, tasks including language production (short-answer questions, free recall, summary, and information transfer) were considered more objective measures of reading comprehension when administered in L1. The students' opinions suggest that in production tasks like these it is not code switching, but rather L2 use that is perceived as a contaminating factor. It was interesting that while some teachers (23.5 %) showed a preference for either L1 or L2 use exclusively regardless of task type, no students indicated an absolute preference for either language. This may lead to speculations about teachers' pre-conceptions.

Summary

Students' and teachers' varying opinions echo contradictory research results in the literature. Empirical studies are needed to find out which task types have the potential to elicit language performance for objective measurement in a Hungarian pedagogical context. On the other hand, as students' preferences concerning task types and language use also vary, it is advisable to include various task types in test booklets to enhance the face validity of testing.

References

- ALDERSON, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- BACHMAN, L. F. (1985). Performance on cloze tests with fixed ratio and rational deletions. *TESOL Quarterly*, 19, 535-556.
- BACHMAN, L. F. (1991). *Fundamental considerations in language testing* (2nd ed.). Oxford: Oxford University Press.

- CHAPELLE, C. A., & ABRAHAM, R. G. (1990). Cloze method: What difference does it make? *Language Testing*, 7, 121-146.
- GORDON, C. M., & HANAUER, D. (1995). The interaction between task and meaning construction in EFL reading comprehension tests. *TESOL Quarterly*, 29, 299-322.
- HELTAI, P. (2001). Communicative language tests, authenticity, and the mother tongue. *novELTy*, 8(2), 4-21.
- RILEY, G. L., & LEE, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing*, 13, 173-190.
- SHOHAMY, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 146-169.
- WOLF, D. F. (1993). A Comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal*, 77, 473-489.