Eger Journal of English Studies **IV** (*2004*) *149–154*

Testing Written English through Pairwork

Katalin Lazarevics

Introduction

No one doubts that testing is necessary. Lots of things are tested day by day and so is language performance in both oral and written forms. Normally, we want to find out about individual achievement of the candidates to check their progress, place them in appropriate groups, give them feedback, or award a certificate. It was not until the 70s of the past century that, due to political and economic reasons, crowds of people wanted to, or had to, learn a second or foreign language in order to get a job or survive in a foreign country. The expectations for intensive and efficient language training as well as for valid and objective assessment contributed to the development of methodology in general, and testing techniques in particular.

Most of us have some kind of experience about being tested. Usually, it is not part of everyday routine but a special occasion for which you have previously prepared. How do you feel when you know you are going to be tested? Are you relaxed or excited, looking forward to showing how much you know or, on the contrary: worried, nervous or in panic at the thought of being caught on something you do not? Most people feel the latter. There is some evidence that the higher the level of anxiety, the lower the performance (Vekerdy). It seems that we are trying to measure something that is being unfavourably influenced by psychological factors so, in most of the cases, examination situations are unlikely to give reliable results. Is there a way out? How can stress be reduced? Testing specialists have long been trying to make examinations less stressful, more human and more real life like (McNamara).

Test designers agree that a good test is objective, valid and realiable. We can add one more feature: testing normally happens individually. Candidates have to face the examination board alone, which can be rather discouraging, or sit for a written test to solve unknown and, sometimes, unexpected problems. Loneliness of candidates may contribute to their feeling of anxiety in examination situations, which could possibly be reduced by co-operation and sharing responsibility with another candidate. Would it affect the features of an effective test if testing happened in pairwork?

Testing writing through pairwork may sound unusual. Still, there is some evidence to support its feasibility. Having tested my own college students through pairwork, I asked some primary and secondary school teachers to cooperate and contribute to my investigation by piloting some tests of their own. Now a wider population has been involved and, as a result, more data are available to support the idea.

Theoretical Background

When we look back at the history of language testing in its written form we can see that, after the end of the hegemony of the grammar-translation method, the first attempts to create grammar tests were made around the 60s and 70s which resulted in multiple choice tests (MCQ). To their greatest pleasure, test designers managed to create a test that was really objective (Weir, 1990). Moreover, these tests were easy to correct with the help of optical readers (computers), and were suitable for assessing language competences for a lot of people in a short period of time. No wonder that MCQ tests quickly became popular all over the world.

But, is it the same skill to select one good answer out of four options as to produce it on your own? The answer is certainly: no (Weir, 1993). What do MCQ tests measure, then? Among other things, they measure the candidate's skill to solve MCQ tests which testees can be (and actually are) trained for. Apart from some possible, usually formal, deficiencies of test rubrics the main problem is that candidates have no opportunity to think: a limited (usually very short) time is given for this type of test. If you happen to know the correct answer, you are lucky. If not, there are some strategies that might help: "When in doubt, choose C or the longest", as some advice goes. Even if you do not understand anything of the stem or the distractors, you can be successful: a probability of 25% is guaranteed in four-option tests.

It did not take too long for testing specialists and item writers to realise the shortcomings of MCQs. In the early 80s they developed new and more reliable test types to measure language competencies. With the help of Cloze-tests and C-tests, yes/no type of decision making was replaced by more productive and flexible ways to find out about candidates' general understanding, vocabulary and grammar skills. These tests are still popular and, slowly but surely, are spreading even in a typically slow and resistant public education system like the one in Hungary (e.g. in intermediate and advanced level school leaving exams, from the year 2005).

No doubt that even the most candidate friendly test is able to create a lot of stress and anxiety which, in turn, can lower testees' performance.

Test designers today are striving to change the character of the tasks: once so popular multiple choice questions have gradually been replaced by more life like exercises to make candidates forget about the fact that they are being tested. Some of these tests look like puzzles, riddles or interesting problems to be solved in the target language. Another, really promising, experiment is connected to oral testing: pairwork is used by some official language examination boards where the candidates are free to choose their partners. Each performance is assessed individually, which, according to the findings of Csépes (2003) is not influenced in any way by the partner's level of competencies. This kind of exam organisation can definitely reduce the level of stress and can create a candidate friendly atmosphere.

Feasibility of Written Pairwork

However convincing the practice of paired oral testing can be, pairwork for written performance sounds astonishingly strange. You can ask questions like "How do you make pairs? How can you find out how much each candidate knows? What about marks? Will everybody get a five, then? There will be noise in the classroom! What should a worksheet look like?", and many more. With the kind help of co-operating colleagues, investigation was conducted in different school types and age groups last year to find out whether, and how, testing through pairwok made sense. It is a common fear that the points of score and, consequently, the mean will go higher, and the normal distribution curve will be deformed just because two co-operating people are supposed to produce better results than a single one.

Experience

In the experiment, language tests were written through pairwork in one German and two English groups of primary school students, and in several students' groups at college. Altogether 116 students (58 pairs) and four teachers were involved. This form of testing came as a surprise to the participants: they learnt about it on the spot. They were asked to form pairs and work with their partners. Surprisingly, test scores tended to give normal distribution, that is, points of score, percentages and marks did not seem affected by the mere fact of pairwork. The teachers themselves found that the results were not much different to the usual. Moreover, the German group took the same test twice: first individually and a second time through pairwork. The difference in marks was as low as 0.05 percent.

Participants

Testing written performance through pairwork is not possible unless candidates share the same idea about it, i.e. everybody agrees. Young adults (students) are usually happy to cooperate because of obvious reasons and, in return, they are ready to promise not to use cheat sheets. Primary school pupils were surprised at their teacher's proposal about paired testing but, after the first shock, they were quite positive about it. Anyway, testing in pairwork should never happen because of the authority, or under the pressure of, the teacher. It is also important that each student give their consent prior to test taking to accept the same score and the same mark awarded by the teacher, with no regard to their individual contribution.

Pairing Pupils

Making pairs can be organized in two ways: (1) candidates are free to choose their partners, or (2) it is the teacher who decides about it. Provided that pairwork had not been announced in advance and students did not have the opportunity to agree who is learning what for the test we can assume that they come to the classroom decently prepared. In this case friendship, realiability and trust are the major considerations for their choice. Motivation for choosing a partner was investigated in all groups concerned, and it was interesting to find that nobody wished to benefit from a much better student's performance. Moreover, all college students held the belief that everybody else was better prepared than themselves, consequently they were ready to accept anyone for a partner in the group.

Those college students who came late could work with a person left alone if there was one, or join an existing couple. A third choice was to work individually, which was not popular and can be considered evidence for feeling safer with a partner. The case is a little bit different in the lower forms of primary school where gender differences are a serious concern. Young children are not willing to choose a partner from the opposite sex which is natural for their age, but makes their choice limited. In certain cases it is the teacher who must take the responsibility for making couples.

Marks and Results

Is it always crucial to find out about each candidate's individual performance? Is it really so very important? In some cases, it is. But, if the couples have agreed to accept the same mark why should the teacher not rely on their judgement? Each pupil has an idea about how they are going to perform and they choose a partner accordingly. Accepting the same mark is part of their self assessment and peer assessment because they expect their

partner to equally contribute. So from the teacher's point of view, there should not be any concern about marks.

All teachers who participated in the experiment were looking forward to the paired test results. If you think that a miracle happened to the pupils and students when they worked in pairs, and they all got good marks, you are wrong. Let us see the evidence. Firstly, both primary school teachers and myself found that test results were realistic. That is, the mean of the given marks was more or less the same as usual. So there is no "threat" to the good old bell shape curve which, due to many other factors, can be easily distorted. Secondly, one of the teachers administered the same test twice: first through individual work, and then through pairwork. She got very similar results: 3.80 for individual work and 3.75 for pairwork, which sounds convincing.

Some more results are provided as shown in the table below:

GROUP	NUMBER	MARKS					MEAN
		1	2	3	4	5	
Form 3	6 pairs	0	0	1	0	5	4.300
Form 8	13 pairs	2	0	0	1	10	4.300
Form 9	8 pairs	0	0	4	2	2	3.750
Students	31 pairs	5	7	6	7	6	3.0645
All	58 pairs	7	7	11	10	23	3.603

As we can see, test results are not significantly higher than they normally are in a mainstream group.

Conclusion

We have some limited evidence to show that pairwork does not influence the objectivity, validity and reliability of tests that otherwise are objective, valid and reliable. Provided that they are valid for individuals, they should be equally valid for pairs of students in the same group. Drawing a conclusion, however, is impossible without further investigation. In addition, paired testing is not likely to be adaptable for all assessment types.

When looking at different test types we can see that placement tests, diagnostic tests and proficiency tests are designed for the specific purpose of gathering information about individual performance (Harmer, 2001). Moreover, very often there is a competition among candidates for a job, promotion or higher salary which would act against co-operation. Still, it

1.000,-

would be interesting to find out if, and how, pairwork can be used in these test types.

Achievement tests however, which do not carry any special risk or stakes, are suitable for being conducted through pairwork. Both of the team members can benefit from collaboration and each will do their best for their partner and themselves. What is the benefit to be gained by this idea? Not more than making assessment less stressful, more relaxed, more of a pleasant experience for those who are acting as candidates. I firmly believe that paired testing can foster important competencies like self and peer assessment, solidarity and cooperation which are vital for our children. Even in a highly traditional educational environment, I can only encourage my colleagues to try it out at least once in a lifetime to see how it works and to check if my ideas are correct.

References

Csépes, Ildikó (2003). *Szóbeli nyelvtudásmérés páros vizsga formátumban.*

Harmer, Jeremy (2001). *The Practice of English Language Teaching.* Longman.

McNamara, Tim (2000). *Language Testing.* OUP.

Vekerdy, Tamás (2001). *Gyerekek, óvodák, iskolák.* Saxum Kiadó.

Weir, Cyril (1990). *Communicative Language Testing.* Prentice Hall.

Weir, Cyril (1993). *Understanding & Developing Language Tests.* Prentice Hall.