

Testing and Teaching: The Dream and the Reality

by J Charles Alderson

© 1998 J Charles Alderson. *Novelty*, Volume 5, Number 4, pages 23 - 37.

This article is an abridged version of the author's plenary address to IATEFL Hungary, October 10, 1998.

Teachers often complain about the washback effect of tests on teaching: tests are said to force teachers to do things they do not want to do. But it has often been claimed that good tests will encourage teachers to do good things: communicative tests will encourage communicative teaching. However, empirical research shows (Alderson & Wall, 1993; Alderson & Hamp-Lyons, 1996; Shohamy, Donitsa-Schmidt, & Ferman, 1996; Wall & Alderson, 1993; Watanabe, 1996) that the real world is more complex. In this paper I will give examples from a number of studies that show teachers doing strange and disturbing things in classes, and I will speculate on why this is. I shall argue that if tests are to have beneficial effects on teaching, teachers will need to understand the rationale behind the test and they will need to reflect carefully on appropriate test preparation practices. Teacher training courses tend to avoid discussions of how to prepare for exams, as if this were unprofessional and undesirable, and instead concentrate on communicative teaching methodology. I shall claim that the skeleton in the cupboard needs to be unveiled and discussed, and we also need to ask ourselves whether the emperor really does have clothes on.

Teachers and tests

I vividly remember two incidents from my school days: one when learning history, one when doing an important examination in German. In the history class, I was one of four pupils identified to take an important examination one year early, and then to skip one year and go straight to the senior school, to prepare for the main school-leaving and university entrance exam. Our teacher, the Deputy Headmaster no less, gave us his four large notebooks full of facts and figures and essays about a series of significant events in British history. Our task, whilst he got on with normal teaching for the students who were not going to take the exam that year, was for one person to read aloud the notes, and the others to copy them down into their exercise books. We did this every class period of history, but there were so many notes we had to meet at weekends and in the evenings in order to have copied them out in time for revision before the exam. The teacher then told us that it was highly likely that out of the standard 20 topics in British history, the following eight topics would appear on the (centrally designed and controlled) examination. He strongly advised us to revise those eight topics (we had copied notes for all 20) and learn them by heart. I did, and six out of the eight questions appeared on the exam, and I got high marks. I have no memory of what I learned and my knowledge of British history is minimal. Why did the History teacher expect us to learn history that way? What did he think we were learning, and how were we benefitting?

When taking my A-Level German examination, in the art room at school (there were only four of us) my German teacher, who was administering the exam, walked by, and read my answer paper over my shoulder. He saw that in the translation I was doing, I had missed out a word for which I did not know the German. He walked on. Shortly thereafter he walked past me again, and put a slip of paper on my desk with the German word I did not know written on it. Why did he help me cheat? My German was very good, and I stood an excellent chance of getting high marks—which I did—without his assistance, his cheating.

Why did he do what he did? Do tests make teachers do things they would not otherwise do, and if so, why do they do them?

Teachers have known about the effect of tests on their teaching—test washback—for a long time. They frequently complain that tests force them to do things they do not want to do. I have often heard teachers say that they do not normally concentrate on grammar or spelling or pronunciation, but rather, normally, they teach communicatively, encouraging their students to speak in class, concentrating on getting their message across, on reading for meaning, on listening to authentic recordings for gist, the main idea, or some important detail.

However, they say that when it comes to exam time, they have to stop doing this and concentrate on preparing students for the exam. The exam is often grammar-focussed, testing students' ability to choose the correct structure from a choice of four in a single-sentence context. There might be a short text to translate from or into the target language, and although teachers say they do not know how to teach translation, and above all how to mark it, they feel they have to spend time on helping their students pass a translation exam, if one exists. There may be no test of listening, and so teachers stop teaching listening and concentrate on written texts instead, since the exam will focus on reading and grammar.

Teachers thus claim they face a conflict between what they do normally, what they feel is good practice in language teaching according to current theory, and what the examination tests, and therefore on how they have to prepare students for the exam.

The reform project

Recognising this pressure from the examination—and particularly high-stakes examinations like school-leaving examinations, university-entrance exams, and even examinations like the State Foreign Languages Examinations Board in Hungary—some people argue that the most sensible thing to do is to change the examination, and to ensure that it reflects more closely how teachers want to teach, how they are currently teaching, and how theory says they ought to be teaching. Thus it is suggested that if the tests become good tests, instead of being bad tests, they will encourage teachers to do good things: communicative tests will encourage communicative teaching.

This is the current position in Hungary: the existing school-leaving examination (the *érettségi*) is a very traditional, largely sentence-based grammar- and translation-focussed examination, that makes no attempt to test listening or writing abilities, and whose testing of oral skills is subject to no quality control, so that students' marks vary enormously depending upon the school and the teacher where they take the test. Teachers complain that the examination does not reflect their teaching or their philosophy, and there has been considerable pressure for the examination to change, to reflect both the reality and the ideal in language teaching.

And indeed plans are underway to change the examination: the detailed requirements for the new examination have been developed and are publicly available documents, after consultation with teachers. Draft specifications for test writers have been developed, items and tasks have been written and are shortly to be trialled for the first time in secondary schools, and much more public documentation will be available shortly about the examination and its likely format and content. It is already known that writing and listening will be tested, that reading and speaking will be tested in much more communicative ways, that translation will not be used as a means of testing reading ability, and that the importance of grammar will be greatly reduced.

These are all developments that are intended to reflect what is considered to be good practice in language teaching. The practice of testing in Hungary is destined to come into line with what is preached about what should be taught in class and how it should be taught. The new examination will therefore not be revolutionary in terms of conventional wisdom about language teaching, and its introduction should be welcomed and it should be assimilated into the system without problems.

But will it? The *Baseline Study* (Fekete, Major, & Nilcolov, forthcoming), conducted under the auspices of the British Council, reports on students' levels of achievement on international examinations, teachers' attitudes to the current examination and to possible change, and about how teachers appear to be teaching at present in the average language classroom.

The results are sobering. Whilst the test results showed that many students achieved roughly Threshold level in Council of Europe terms, and thus are capable of more than teachers say they are, teachers' attitudes are ambivalent. Whilst they are critical of the existing examination, they also appreciate the fact that most students pass it. It is extremely rare for students to get a 1 or even a 2 on the *érettségi*. Teachers recognise that the examination is generally considered of low quality, and yet they feel that it is good because students pass. Of course, this poses a dilemma: if students pass a very easy exam, it can say little about their abilities, and if results are unreliable, as they are indeed generally considered to be by teachers, then they are meaningless. And indeed, it is precisely because the exam was considered to be too low-level and too unreliable that the law was changed in 1989 to allow the Rigó utca examination to be taken by school-children as well as adults, to replace the exam. Those who pass the Rigó utca examination are exempted from language classes, they receive an automatic 5 in the school-leaving examination, and bonus points for university entrance. No wonder that many students wish to pass the Rigó utca examination, and that most students, employers and universities consider the *érettségi* to be irrelevant to certifying language achievement.

But a reform of the examination is likely at least to question this status quo. If the exam is to be improved and to be respected and valued by employers and universities, it will have to become more reliable, and more relevant to the world of work and study, as well as to teaching. In other words, it may well become more demanding, on students and on teachers. An exam that everybody passes will command no respect, will have no currency or value, and thus is unlikely to have much effect on teaching, teachers or pupils.

Now, the point of examination reform is both to bring testing more into line with teaching, and possibly also to influence teaching, and to do this, the exam will need to have status and force: it will have to be an exam that is taken seriously by students and teachers alike. And the exam that is taken seriously is an exam that not everybody "passes". And there is the rub: teachers like the examination at present precisely because it is easy, and their students pass. They even interpret a high pass-

rate as reflecting well on their teaching. The reverse is even more true: a low pass-rate, or a relatively high failure rate is seen as showing that their teaching is poor, in some sense. Thus teachers have a problem: they may not like the content and form of the current exam, they may feel it ought to change to reflect current fashions or current practice, but if it does, there is a danger that it will become harder to pass, and thus present an obstacle to students. But if it remains easy to pass, it will simply not be taken seriously.

One possible solution to this is to rethink what we mean by pass and fail. At present we have a time-honoured system of grades on a scale from 1—5, where frankly the scale is effectively three points at most, and where students and teachers get upset if they are given a 4. Yet we know that students do not all have the same level of achievement, and it is ridiculous to give them all the same mark, when they vary enormously. The only sensible thing to do is to develop a finer-grained scale that will more truly reflect the differences in the population, and at the same time will certify what students have indeed learned. This will, however, represent a major change in the national system of awarding grades and rewarding achievement, and it is unlikely to be easy to implement.

Whether students are motivated by the exam is another question to which I do not have an answer. What students think of it is also unknown, but how students would react to higher failure rates is predictable: they would be unhappy. Hence the need to reconsider what is meant by passing and failing and to consider seriously a system where students' achievements, however modest, are recorded and reported, and they are not simply branded as "failures".

A major goal of examination reform, I have suggested, is for the examination to better reflect best practice in language teaching, and hopefully how teachers are actually teaching. The goal of reflecting best practice is surely laudable. What is less clear is to what extent teachers are actually following best practice.

What do normal Hungarian classes actually look like? There is surprisingly little evidence on this topic in the public domain, and so the *Baseline Study* (Fekete, Major, & Nikolov, forthcoming) made a special point of observing classrooms throughout Hungary. Nikolov Marianne (1998) reported on the findings, which make interesting but somewhat depressing reading.

A hundred and seven teachers were observed in 118 classes in 55 secondary schools (two-thirds of which were vocational schools, three-quarters in big towns). Years 10 and 11 were observed in roughly equal numbers, but fewer in Year 12 (49:42:27). On the average, teachers used Hungarian for one third of the lesson, but there was a wide range (from 0% to 95%). Students also used Hungarian for roughly one third of their utterances, showing considerable reluctance to use English.

Vocational schools revealed a strong emphasis on grammar: this was less marked in grammar schools but still substantial (in second place, after writing). The commonest oral task by far was question-answer, with picture description in second place (106:29). The commonest reading task was reading aloud, followed by matching pictures to text (64:21). The commonest writing task was copying, followed by gap-filling, translating and arranging words in sentences. Very few listening tasks were observed. There was a notable decrease in focus on listening and reading in Year 12, and a corresponding increase in focus on grammar, translation and integrated skills.

The observers commented that the majority of classes were boring and un motivating. Teachers claimed in interviews that their students were not motivated, but the observers felt that teachers did not accept responsibility for motivating them. Few teachers felt motivated or successful, most classes were teacher-fronted, with

little pair or group work, and an excessive amount of mother tongue use. Facilities (like OHPs, cassette recorders, TVs, and even computers) were available in the vast majority of schools, but were rarely used. Language laboratories existed but no lesson was observed taking place in one.

Now if most teachers are teaching like this, it is evident that this is far from best practice. The question then becomes: should attempts be made to change how teachers teach? Presumably, teacher education courses are intended to do precisely that, but we all know that it takes a long time to convince people of the need for change, and even longer for new ways of teaching to be implemented. Once we know how teachers teach, we need to know why it is that teachers teach the way they do. If we can understand their reasoning it may be possible to convince them of other ways of doing things. Certainly if we do not understand their reasoning, there is little chance of convincing them, or of their making more than superficial changes. But as of today, it is not known why the teachers observed do actually teach the way they were observed to teach.

Introducing a new examination is often advocated as a way of forcing change: a better exam will result in better teaching, using tests as levers of change. And no doubt in part the Examination Reform Project has been conceived as a way of encouraging teachers to do different things in their classes.

However, empirical research shows that the real world is more complex. Simply introducing a new test does not necessarily bring about desired behaviour. A number of studies show teachers doing strange and disturbing things in classes. I will present the findings of one such study.

TOEFL

Perhaps the most influential EFL examination in the world is the TOEFL, and it is said to influence massively the way teachers teach. Yet amazingly, I have failed to find a single empirical study of how teachers teach when preparing students for TOEFL, apart from the one which I carried out with Liz Hamp-Lyons (1996) in the USA.

We decided to see whether teachers were affected by teaching towards TOEFL by observing both their "normal" teaching and their TOEFL preparation classes. We wanted to see whether teachers taught towards TOEFL in similar ways.

What happened?

We grouped Teacher 1 and Teacher 2's TOEFL classes together and compared them with their pooled non-TOEFL classes. We found that:

- Test-taking is more common in TOEFL classes;
- Teachers talk more and students have less time available to talk in TOEFL classes;
- There is less turn-taking and turns are somewhat longer in TOEFL classes;
- Much less time is spent on pairwork;
- The TOEFL is referred to much more in TOEFL classes;
- Metalanguage is used much more in TOEFL classes;
- TOEFL classes are somewhat more routinised;
- There is much more laughter in non-TOEFL classes.

In other words, there seemed to be an effect of TOEFL teaching. However, when we analysed Teacher 1's classes separately from Teacher 2's classes we found substantial differences. Teaching TOEFL had an effect on Teacher 1's use of testing in class time—he did much more of it than in non-TOEFL—but did not affect Teacher 2, who never used class time for testing whether teaching TOEFL or another kind of class.

In non-TOEFL classes Teacher 1's procedure was basically the same as in TOEFL classes, except that the work being done and then gone over was not test practice but exercises in the textbook. His nomination of students followed a strict and predictable pattern.

Teacher 2's TOEFL and non-TOEFL classes showed more variation in pattern: he called on students unpredictably, and often selected items to be completed out of their sequence in the textbook.

Teacher 2 talked less than Teacher 1 in TOEFL classes, so students had more time available to talk. The classes were more interactive (having more turns) regardless of whether they were TOEFL or non-TOEFL, and although he used no pair-work in his TOEFL classes, he used much more pair-work than Teacher 1 in his non-TOEFL classes.

His classes were more innovative, and his TOEFL class had as much laughter in it as Teacher 1's non-TOEFL classes.

In fact, we feel that the differences between the two teachers are at least as great as the differences between TOEFL and non-TOEFL classes. To attribute the features of any one class to the presence or absence of the TOEFL test as a goal proves, on these data, too simplistic: individual teacher styles clearly contribute enormously to the nature of the classes we observed.

(Alderson & Hamp-Lyons, 1996, p. 290)

Table 1 sets out what we felt to be the main contrasts between the two teachers.

Table 1: Impressions of differences in teaching style

Teacher 1	Teacher 2
Room	
On podium	Same level
Behind desk	In front of desk
Seats in rows	Seats in horseshoe
Method	
Calls on SS in sequence	Calls on SS unpredictably
Never varies item order	Varies order of items
Often calls attendance	Never calls attendance
Uses blackboard frequently	Uses blackboard little
Often starts late	Sometimes starts early
Sometimes is unprepared	Appears unprepared
Little happens	Class is hectic
Always completes on time	Overruns on task time
Never changes voice	Changes voice to mimic
Explanations articulate	Explanations ad hoc
Uses textbook examples	Makes up own examples
Reads aloud from book	Rarely reads from book
Extensive metalanguage	Uses some metalanguage
Answers are rarely wrong	Deliberately gets answers wrong
Never uses attention getters	Uses attention getters
Often gives test	Never gives test in class
Tends to slow down	Talks at normal pace
Goes through all items first without stopping	Stops after each item for discussion
Very familiar with TOEFL	Not familiar with TOEFL

Why does this happen?

We asked teachers, not just the two observed, to complete the following sentence: "What I like most about teaching TOEFL is." Many teachers replied: "Not having to do lesson planning," and "not having to mark homework." Imagine: the two most important things that teachers might be said to do - plan what they are going to do, and then give students feedback on how well they have performed - and yet teachers feel no need to do this in TOEFL classes. Why?

One possibility is that they have simply never thought about how to teach TOEFL: thick textbooks are available complete with answer keys (and significantly without guidelines to teachers on how to teach the class): the tests are there, all the student has to do is complete the exercises and check the answers in the back of the book. What does the teacher have to prepare or even give feedback on?

However, we felt that teachers could have found ways to make closer connections between test preparation and their preferred teaching methods and activities, and between the content of their 'regular' courses and successful preparation for TOEFL, if they only spent the time and imagination on these classes that they spend on their regular classes.

In addition, class sizes were different: 12-16 students in regular classes and up to 30 in TOEFL classes. Clearly, what a teacher can achieve with 12 students is different from what she can achieve with 30. This decision about class size suggests that the administration either believed that TOEFL teaching was possible in larger classes, or that it was somehow less important, or that it made no difference to TOEFL-type teaching how many students there were in a class.

The two teachers we observed certainly adjusted their teaching methods when teaching TOEFL classes, but we questioned the rationale for such change. It seemed to us that the use of pair and group-work was likely to be just as useful and effective in TOEFL classes as in non-TOEFL classes; TOEFL classes could be smaller—the same size as regular classes—and they could be streamed by ability. TOEFL classes could be made fun, by a variety of means. There is no obvious reason why TOEFL classes have to be taught the way they are.

There is much teacher-dominated lengthy explanation of grammatical points, with the use of a great deal of metalanguage and little opportunity for students to practice or to experiment with what has been said. Such behaviour suggests a rather naïve learning theory on the part of the teacher: “Give students an explanation, they will understand it without problems, and this will automatically change the way they use language and answer test items” (Alderson & Hamp-Lyons, 1996, p. 294).

We conclude that there needs to be much more questioning of what the best way of teaching TOEFL might be.

There ought, in a mature profession, to be much more critical questioning of the assumption that test-taking practice will automatically lead to improved test performance, or that explanation will lead to internalisation. (Alderson & Hamp-Lyons, 1996, p. 294)

In short, teachers showed little evidence of having thought about what best practice might be in TOEFL preparation, and how this might square with their general teaching philosophy and practice.

Central Europe

Although I have not yet conducted similar systematic research in Europe, I had a rather similar experience recently in a Central European country, where I had the privilege of watching a teacher teach two consecutive classes, one Year 11, and the other Year 12. The two classes were very different, and the only reason I could see was that the Year 12 class was about to take the school-leaving examination.

In the Year 11 class, the teacher pinned verbal descriptions of cities around the walls, and handed out photos of some of these cities. The students' task was to walk around the room and match pictures to descriptions, and to identify the two descriptions for which no picture was provided. The aim of the activity was reading for meaning and no attention was paid to unknown words or to structures. The teacher did not dominate the class, and students were left to solve the problem using English and whatever other resources they could muster.

The next lesson was a Year 12 class taught by the same teacher. The students were said to be at the same level of ability as the Year 11 class. But the lesson was very different. They were due to take the school-leaving exam at the end of the year and the teacher was preparing them for an oral test on the topic of Environment. She distributed three pictures of a polluted place and an arid place, and of recyclable

products. She wrote “Problem and Advice” on the board, and asked them to list problems first. They had clearly done this topic before as they volunteered phrases like “global warming,” “greenhouse effect” and “CFCs,” which she wrote up. She asked them to volunteer advice and wrote up: “Don’t emit smoke,” “buy recyclable products,” “don’t wash your car in the river,” and the like. Often, students offered these ideas in L1 and the teacher wrote them up in English.

The students diligently wrote down every word: I got the impression they expected to and indeed would learn these by heart later. The teacher seemed to be coaching them by giving them things to say and words to say them with. She acknowledged that in the past she had been responsible for setting the topics for the oral, so it is likely that students and she expected that she would select the topic of The Environment.

This class was very different from what and how she was teaching the Year 11 class: teacher-fronted rather than pair and group-work, focus on learning and practising phrases rather than processing meaning. It is unclear why the teacher taught differently, what the students would have learned, or how it would benefit them in the exam.

In a second but similar context, I observed a class which consisted of students of computing, electrical and mechanical engineering, who had four lessons of English a week. They would be taking the school-leaving examination at the end of the school year. They were using *Headway Pre Intermediate*.

First the teacher reviewed the homework, during which students were asked to ask each other questions about their holiday: “What are the beaches like? What are the kangaroos like, what is the weather like?” Having told them in English to ask each other questions in pairs, the teacher then repeated the instruction in L1, and translated the questions for the students.

Nevertheless, the students showed some confusion as to what they had to do. The pair-work—more like a drill—was boring, the students seemed bored, but yet were amazingly obedient. To my surprise, the students actually asked for the English equivalent of words like “three-star hotel”, “fish soup” etc. Then they performed in public:

Q: “What was the food like?” A: “Chips and beefsteak.”

After this review of homework, the teacher came to the main point of the lesson. She said: “You will hear a text with adjectives. Some come three times. Note the form. Do you understand?” She then played the tape, which was very fuzzy and hard to understand. The students didn’t do anything much, although when asked for an adjective, they volunteered “exciting.”

T: “Which three forms of ‘exciting’ did you hear?”

No answer. We listen again.

T: “Listen for an irregular adjective. Which do we know?”

Ss volunteer “good” “bad”, “fat.”

T replays tape, and Ss have no difficulty identifying “more” “most.” “Many” she points out, “is an irregular adjective: what are its forms?” (more, most).

This low-level exercise continues:

“Last year, we learned about comparisons of adjectives.”

“What are the three forms of ‘interesting’?” Ss volunteer. “Good” “bad” “little” (“less, least”!) Ss volunteer, having no difficulty.

Relentlessly, T: “Turn to page 41: give the three forms of these adjectives. You remember the rules from last year? What happens to words with two syllables ending in y?”

Ss answer.

“What happens to words with two syllables that do not end in y?” Ss answer.

“What happens to words with three syllables?”

Ss answer.

Three Ss put their heads on their arms and start sleeping.

T: “I’ll show you the transparency you saw last year to remind you’ T puts it on OHP—showing the rules, and sample forms.

T: “You know it well because it is in your exercise book. This is from last year, so you needn’t write it down. I just wanted to show it to you.”

T repeats “ugly, uglier, ugliest, big, bigger, biggest”

T then asks individual students to read out the three forms of the words in the book: “modern” “ugly” “big”, etc: roughly 30 words in total. No context is given or required, just the three forms. They have no difficulty at all.

T then asks: “What’s the opposite of ‘wet’, Peter?”

Peter: “wet, wetter, wettest”

T: “Most of you remember quite well. If you have problems, you can revise it from last year’s exercise book.”

Next, T: “Bottom of page: facts and opinions. Complete the sentences in pairs.” T then has to go round class explaining in L1 what they have to do.

T: “This is not new: we learned about it last year.”

T: “Start writing.” (It’s unclear why they have to write in pairs, and most don’t, initially. Lots of L1 chat)

T: “Now listen to Tim telling you about the cities again”: T plays the tape again with the adjectives.

T: “I will stop the recorder for you to check the sentences, fill in the comparatives and superlatives in the textbook”

Now Ss are asked to read out the sentences.

Complete boredom and lack of attention.

Homework: to write in the workbook page 32 exercise 3 the correct forms of adjectives. As students get up and struggle to leave, T shouts: “Please revise everything about comparative forms and sentences. You will get marks next time. This is only revision.”

I was deeply struck by the meaningless way the teacher repeated again and again something that the students could already do, presumably because she felt it important for the exam, yet the students were clearly able to do the exercise anyway. I was also struck by the dissonance between the content (and methods) of the English class, and the life I presumed they lead outside the English class. Why did she teach that way? Why did the students tolerate it?

Hungary

There is little systematic evidence yet on how teachers prepare students for examinations in Hungary, but there are plenty of anecdotes, especially with respect to the oral examination, which is teacher-controlled and subject to no external quality control.

In February, teachers give out the likely broad topics that students can expect to be tested on in the oral examination. But within each topic, teachers give model conversations and students write out and learn word-for-word their speech on these topics (much like the Central European teacher I have described). Weaker students are told exactly which topic to prepare, and since the class teacher is the examiner, and the Chair of the Examinations Board often does not speak a word of English, the teacher can ensure that the weak student is indeed tested only on what she has learned by rote in advance. Conversations are developed in class, sentences are constructed and written up on the board, they are rehearsed and rote-learned in class.

Students are expected to tell the content of a reading passage in the oral examination, in Hungarian, supposedly a test of reading comprehension. They have up to 15 minutes to prepare the text, and teachers often let them summarise the text in English rather than Hungarian, which of course they have attempted to learn by heart in the preparation period. Other teachers on the panel often do not speak English and are persuaded by the English teacher that it is harder to retell a text in English than in Hungarian, whereas in fact students are simply regurgitating the text without comprehension! Many teachers use modern textbooks, know about communicative teaching, have attended workshops and seminars on modern approaches to teaching, and yet as the school-leaving examination approaches, they appear to change, and their methods begin to resemble ancient, traditional, form-focussed, rule-based teaching. Two months or so before the written exam, teachers are said to bring to class the previous year's examination paper, and discuss grammar points: modal verbs, reported speech, passive voice, sequence of tenses, conditionals and the like. Stannard Allen, Thomson and Martinet or Hungarian exam preparation books are used in class and for homework. Students are expected to remember correct patterns and sequences, and they take copious notes on appropriate structures.

Why do teachers do things like this?

I want to argue that if tests are to have beneficial effects on teaching, teachers will need to understand fully the rationale behind the test and they will need to reflect carefully on appropriate test preparation practices, and on their own teaching practices, as well as on their own test preparation practices. Yet teacher-training courses tend to avoid discussions of how to prepare students for exams, as if this were unprofessional and undesirable, and instead they concentrate on general teaching methodology.

In an ideal world, perhaps—the Dream of my title—the way teachers prepare students for examinations would look no different from how they teach every day, and how they teach every day would match both the aims and the methods that the examinations intend to inculcate, and the methodology that the textbooks and the teacher education courses advocate. However, it is absolutely clear that teachers will indeed engage in special test preparation practices, and therefore it is important to consider what best practice in test preparation should look like. What we need to develop is teacher education programmes that will encourage such reflection on test preparation and likely best practice, and address what inappropriate and bad test preparation practice might be, and why. I believe strongly that this skeleton in the cupboard needs to be unveiled and discussed.

Thoughts on changing teacher thinking

We need to ask ourselves whether what is going on in regular classes really is good practice, desirable and thoroughly professional, or unreflective, traditional, habit-based activity which allows under-paid and overworked teachers to survive in the real world of classrooms. And if it is the latter, then it is clear that we face a Herculean task if we want to get teachers to change the way they teach. Changing the examination is but one small part in the process. It is essential that teachers are given full and open information about examination reform, they should be encouraged to attend courses and workshops on what appropriate examination preparation might look like, they should be encouraged to reflect on this and to come to terms with the dissonance between what is so frequently done, both in test preparation, and in regular teaching, and the aims of language teaching generally, and the examination in particular. Above all, we should be clear that supporting teachers to come to terms with new exams, indeed with any innovation, takes time, and needs patience and understanding.

It is, however, not clear what we can or should do about unprofessional practice as illustrated in my examples, or in the anecdotes about current Hungarian practice. Teachers may well do such things for the best of reasons: to help their students overcome important obstacles, in the only way they know how. Who can blame teachers for wanting their students to succeed? Nevertheless, I strongly believe that we should discuss why teachers do such things, what such practices imply for teachers' professionalism, and how best we can bring about a situation in which the ethics of such practices are addressed, and such practices are considered inappropriate and unnecessary.

Conclusion

I have considered the importance of test washback on teaching, and suggested that there is indeed evidence to show that tests impact on what teachers teach, but the impact on how they teach is much more complex. Thus it is important that test designers pay attention to the content of their tests from the point of view of washback: if listening is not tested, it will not be taught, and similarly with other skills. If translation is tested, it will be taught, somehow. Thus good test design is a prerequisite to good test preparation practice. However, it is not quite clear how test designers can impact on the way teachers think and teach. Indeed, to my knowledge test designers have not yet seriously considered the methodological implications of their tests. Is it possible to design tests in such a way that they will influence how teachers teach, and not just what they teach? That is at present an unanswered question, and one that test designers need to ask themselves and attempt to answer.

However, it is also clear that whatever the test looks like, much more needs to be done if we are to change the way teachers prepare students for tests. Teachers need full and open information about exams, they need to understand the rationale for the test, and above all considerable effort needs to go into changing the way they think about test preparation, and into encouraging them to adopt a critical, reflective approach to their own practice, in order to develop more mature, responsible and appropriate ways of preparing students for examinations.

If we are able to do this, through constant reflection and the incorporation of such reflection into our teacher education, through constant and open communication about the new examinations, through explanation of the rationale for the examination

and through exemplification of how best to prepare for it, then perhaps one day the dream may become a reality.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13 (3), 280-297.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115-129.
- Fekete, H., Major, E, & Nikolov, M. (Eds.) (forthcoming). *Baseline study*. Budapest: The British Council.
- Nikolov, M. (1998). Classroom observation project. *Baseline study* [Online]. Available <http://ipisun.jpte.hu/~joe/baseline/observation>
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13 (3), 298-317.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10 (1), 41-69.
- Watanabe, Y. (1996). Does grammar-translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13 (3), 319-333.