ISSUES IN LEARNER ASSESSMENT BETWEEN TRADITION AND CHANGE

Oana Maria Franţescu
Lecturer, PhD, "Al. Ioan Cuza" University of Iaşi

Abstract: Conventional teacher training – at least according to the basic syllabus of a TEFL Methodology course at BA level – places its main focus on teaching and only marginally on assessment. This is a serious shortcoming in the initial stages of teacher training as new teachers are expected, by default, to also funciton as assessors, firstly of the quality and quantity of their own class teaching, but also of other types of content and in other types of assessment. The paper aims to explore the limits of the current system of assessment in terms of aims, procedure, materials, examiner training, outcomes and quality assurance, as well as to advance a few suggestions relevant for teacher training.

Keywords: Validity, content, reliability, syllabus, skills

The changes operated in the Romanian system of education (initiated by the Ministry of Education) are reputed for their overall lack of coherence and volatility. Some of the most drastic changes were made in the area of assessment and less in teaching per se. The main reason was to change the student dynamics and boost tertiary education enrollment in an effort to increase employability and a quick return of the state's investment in what the public perceives as „free education" for all citizens. But apart from the national examinations (at mid-secondary education and then the Baccalaureate) and those organized by the schools with the aim to recruit the best students for the 5th form, another form of assessment still goes unnoticed and untouched by any wave of reform since 1990, the „oral questionning" or „*listening*" – a literal translation of the Romanian word *ascultare* which refers to a form of in-class assessment done by the teacher spontaneously during a regular lesson, usually before teaching something new. The „oral questionning" is not typical of the English lessons but a universal practice used by teachers of all subjects with all age groups, and it is generally considered part of a normal lesson, often carefully included in the Lesson Plan. It is recommended by many entities as having a number of roles, one of them being to inform further teaching.

The teaching of foreign languages follows two traditions. The more recent one dates from the first years of post-Communist transition when the Communicative Approach finally saw its way into the more advanced schools in the country's bigger cities. In the early 1990s, the old books were still in use in most schools for lack of new adequate material, and with them the teachers' old ways too, albeit not for long. With the rise of alternative textbooks and other teching materials, and with the emergence of a strong exmainations market, the methods and techniques used to teach English as a foreign language became more refined and diverse. For example, Listening was introduced and there appeared a demand for audio materials for all levels. Another significant change was the need to develop genuine speaking skills in secondary school learners, and even authentic communication activities. It took less than ten years to see new books, materials specially designed to help teachers develop speaking skills, and new curricula and syllabuses as the English programmes for secondary schools diversified to include intensive hischool courses, bilingual programmes, content-based course packages and even pre-school classes of English. Encouraged by the growing support for teachers both in state-funded institutions (universities) via newly introduced teacher training

209

programmes and in private centres, the TEFL dynamics swiftly crossed from the Communicative Approach to the Post-Method stage so as to accomodate new techniques and teaching formats.

Meanwhile, the other tradition managed to survive in parallel as widespread as the centralized state system of education. It preserves all the characteristics of the pre-transition period and is perpetuated in the recommendations of the county school inspectors despite the efforts made to adopt the Common European Framework of Reference for Languages into the Romanian teaching practice. It originates in the 1978 Act of Education (especially Art. 10, 32, 37, 43, 102, 111-114, 120, 121, 140, 141, 142, 146, 148, 155, 173, 175, 176, 185, 186, 193 of Law 28/1978) and survives despite the numerous efforts to amend the subsequent reforms of the system of education. It has visible traces in the consecrated practices perpetuated within the system of quality assurance in the school hierarchy.

Currently, all the classes of English taught in the state system are based on a national curriculum issued by the Ministry of Education. The content of the curriculum varies for each learning cycle, type of course, number of hours per week and type of institution. Thus, there are curricula issued for lower and higher secondary schools, for English taught as the first or second foreign language (abbreviated L1or L2), for bilingual or intensive courses, for theoretical highschools, professional, vocational, for common core or optional courses etc. Once a teacher has got a job in a state school (or, for that matter, a private school that folows the guidelines of the state curriculum), it is their duty to first familiarise themselves with the curriculum they need to follow for each of the classes they will teach in that school year. Depending on the organization they work in, the teacher's norm may include lower secondary classes, primary school classes, higher secondary (highschool) of some kind, common core and optional courses and as a result they may have to know more than one curriculum and plan their lessons for each of these situations. Unlike in other countries, in Romania it is common practice for teachers to have to prepare to teach more than one age group across a variety of contents and formats in the same school year, and the classes may often be of the mixt ability type and even include learners with special needs.

With the introduction of Common European Framework of Reference for languages, the curricula included information about the level of language taugh in the main cycles (generally, level A in lower secondary schools and B in highschools with up to two hours of English per week with a slight orientation toward higher levels), as well as additional information regarding what are called the four competences – a literal translation from the Romanian word used in the documents, but basically the four skills - , and a number of recommendations about activities to use in class. If prior to the latest version (2017) recommendations for assessment included the use of so-called „traditional techniques" (such as oral questionning, formative and final tests) and some other alternatives such as the portfolio, the project, homework/assignments, observation or self-assessment, the latest version makes no such reference but instead recommends that assessment focuses on measuring performance in each of the four skills (p.16). Towards the end of the document (the section dedicated to the English language), there is an example of activity used for evaluation and a model of scoring.

A comparison between the previous versions of the curriculum and the latest versioninvites a number of conclusions. One would be a clear need for further teacher training than is currently done in universities and during in-service training programmes. Then, the little that was said about assessment in the previous version reflected a sad continuation of the Communist and post-Communist thinking as to what is good assessment and, more importantly, what is quality education. But in the latest version assessment features even less and what is mentioned is less

effective. The proposed scoring scheme is simply unfair because it fails to account for the first four points out of ten (p. 22). And this is the point where tradition meets change.

The Act of Education of 1978 establishes a general (national and centralized) grading scale „from 10 to 1" where 5 is the pass/fail point (Art. 186). This has not changed to this day, but the proposed model of evaluation in the 2017 curriculum leaves almost half the scale unaccounted for. It is generally accepted in schools that any grade under 5 is a fail. Moreover, grade 1 is now significant of cheating, regardless of the method, and usually comes with some supplementary penalty such as disqualification from exams or expulsion from the school. This widespread practice leaves room for confusion as to the meaning and interpretation of this grade as 1 in a test can equally mean that the candidate cheated or that he did not cheat but his level of performance corresponds to one point out of ten. Also, opinions are divided as to the opportunity of the remaining „fail" grades of 2, 3, and 4. Some teachers have no difficulty giving such marks since they do reflect a candidate's performance accurately even if all these grades have the same meaning i.e., fail. Others feel these grades are superfluous as their meaning overlaps and so only 4 is sufficient to signal that a candidate's performance is substandard; once you fail does it make any difference with what grade? The latter trend seems to justify the assessment model given in the curriculum.

The evaluation aims are stated in the Order of the Ministry of Education (no.5079/2016, Art.112-113). The curriculum only suggests that evaluation should focus on the four skills (presented under the generic name of competences) in specific areas (subskills) corresponding to a particular stage of acquisition (for each year of study within the secondary cycle), but the Order of the ME clearly states that evaluation aims to measure the level of acquisition in view of improvement (Art. 112) and that it should measure competences (Art. 113). Another difference between the two documents is that the Order enumerates some assessment tools to be used for all subjects including foreign languages: oral questionning, tests, experiments, essays, projects, interviews, portfolios and practical skills, which the current curriculum for English does not. So even the most recent regulations continue the tradition of „oral questionning", and Articles 117-118 of the Order clarify that evaluation is formal and its results are expressed in grades or marks from 1 to 10 (for secondary education) that are recorded officially.

For the sake of fairness, there are other official texts that elaborate on evaluation in what concerns foreign languages, for example the methodologies of national exams. But there is no coherent national system of evaluation and, as a consequence, apart from university courses and isolated (often private) teacher training programmes, there are no other centralized sources of training or information for examiners/assessors, or a system of quality assurance for examinations. This void is filled by the recommendations given by inspectors and senior teachers with guidance duties. Thus, it has become a given fact that „oral questionning" is a valid form of assessment.

The technique was perceived as coherent with the vision of education in 1978. The Act of Education gives ample evidence as to the intended aims of schooling. In several articles (see above), the Act states that education is only done in schools (no other alternatives such as home-schooling) and includes acquisition of knowledge, skills, values and attitudes, patriotism, professional qualification, and discipline. The last in this list is a duty on the part of the educator and the beneficiary of education. And „oral questionning" has been used for more than just evaluation in view of imporvement of learning, it is perceived as a way of enforcing discipline in the classroom and prevent or stop disruptive behaviour. Implicitly, the mark that represents the result of this type of evaluation (as is known to this day according to the Order of 2016) is also an instrument of class management. By extrapolation, tests can be (and often are) used for discipline

too. The Order of 2016 also gives a formula for establishing the number of marks to be given per semester for all evaluation except the end-of-term tests, which represent a separate summative assessment. Thus, the number of hours of English per week indicates the minimum number of marks per semester. If a student is likely to fail, he will be assessed a third time in the final two weeks of school. The Order also shows that the formal summative (end-of-term) tests must be kept for a year for further consultation. However, nothing is said about all the other forms of assessment, which implies that, on one hand, it is impossible to formally challenge a mark given for „oral questionning", and that marks given for any other type of assessment enumerated can only be contested if the proof of assessment (interview, portfolio, project, test etc.) has not been destroyed.

Considering all of the above, „oral questionning" as an official classroom assessment of English is faulty in terms of rapport, reliability, validity, scorability and administrability. Good rapport is one of the indicators of good teaching and is based on predictability, positivity, fairness, support, openness, and tolerance. Rapport is a constant component of any interaction between learners and their teachers, which includes evaluation. Thus, any evaluation must be predictable: learners need to be informed (ideally, consulted especially if we include self-assessment among the instruments of evaluation) beforehand that they will be formally evaluated (for marks, of course) and about the circumstantial details (when, in what format etc.). Calling a learner to stand up or come to the front of the classroom to spontaneously answer a number of questions for one of the (at least) two marks in a semester is simply wrong, even if legal today, because of the surprise element: an overwhelming 60% of Romanian students report negative feelings about being subject to spontaneous grilling (Ionescu, 1995:225). It follows that the correct way of doing this is in fact to organize an interview and give the learners the opportunity to prepare for it. As it happens, the interview is also an authentic and communicative activity, it is used as part of the assessment process for various forms of applications in companies, schools and universities. Then, „oral questionning" is never fair as it does not give equal opportunity to students to be assessed exactly by the same standards on the same content (Cyril J. Weir's book *Language Testing and Validation* dedicates Chapter 5 to the characteristics of test takers to show how widely different is the starting point of any assessment from their point of view). It would be fair if it were conducted confidentially instead of in full view of the class: thus, the same or a similar question could be asked of any learner in turn and any subsequent candidate would not have the benefit of the precedent, but a fresh start to give an answer.

Secondly, „oral questionning" has serious reliability issues as it cannot yield comparable results if repeated to other learners, not to mention the alternative situation where this technique is in fact used as a class management instrument. And this feature goes hand in hand with validity. In the English class, „oral questionning" exploits limited parts of the curriculum. Some people in authority positions (inspectors, senior teachers, heads of department, head teachers etc.) recommend some regularity in the pace at which teachers record marks in the official register – for example three marks per week at one class – so as to ensure the appearance of continuous evaluation. In fact, the Order of the Ministry of Education of 2016 mentions that evaluation must be continuous but does not impose any rhythm. Still, having detailed the instruments of formal evaluation (including oral questionning), the choice of the word *continuous* (*evaluation*) has allowed for a literal and extreme interpretation of the text to the effect of an obligation to produce official marks and a point of contention in the way marking rhythm reflects good teaching. It has to be done in class but it cannot take all of the teaching time (otherwise it is not considered oral questionning but a different category of assessment), so it should only take a few minutes. But even if we could stretch the idea of questionning to include some reading („Please read the following text and translate these lines

212

from it") or even writing (for example, inviting the student to write an exercise on the board for a mark), in fact very little can be measured in this way. At best, one (specific) competence can be measured at one time. So any grade given for a short session of questionning is then recorded as if it represents the learner's general level of performance for half the semester. In other words, an instrument of assessment ends up measuring two things at once: a specific competence (a subskill) and all the skills and content taught up to that point. Again, this assessment technique raises a serious issue of fairness in the sense of an insufficient and inadequate sample being assessed as representative of the skills and content that were not assessed. Maybe students with high grades do not necessarily perceive the unfairness of such an evaluation, but for poorer students this is a serious demotivating factor.

The next point, scorability, continues the above discussion about the model of assessment given in the current curriculum for English. The Order of the Ministry of Education of 2016 does not give details as to the way each instrument of assessment should be scored. The methodologies elaborated for national exams come with a scoring scheme and a formula for calculating the final grade, but for formal class evaluation (including oral questionning) there are no suggestions. The Order does mention that all evaluation should come with feedback for the students and their parents as well as suggestions for help and improvement, and that it is the duty of the Teaching Committees in each school (the equivalent of departments for each subject) to design assessment instruments – but nothing further. All formal evaluation should be based on a set of clearly defined criteria; the fact that we ask students to say/read/write something only represents that half of the assessment where we harvest the material we assess (Ur, 2009:245). But oral questionning fails to justify a mark from 1 to 10 in detail. It may be clear that a student does not answer a number of questions, for instance, and that his performance will be considered insufficient for a 5, but any grade, even the fail ones (1 to 4), must be justified as the total performance in terms of specific competences. If a student can't read properly, will he fail at English? Does his reading reflect his ability to communicate? In the mother tongue this may be the case, but the level of communicative competence differs between the same language as mother tongue and as foreign language (Braniste, 2016:24). If he can read very well does he get a 10 or should the teacher press on with the questions until the student makes a mistake and loses his 10? And does a question carry more weight than others? These examples illustrate the lack of a set of criteria for marking. Once the criteria are defined based on the competences for that level, oral questionning will no longer be sufficient to assess a student's performance.

Finally, there is no prescribed pattern for oral questionning, it is left to the teachers to decide how to administer it to the best of their ability. A number of factors must be considered such as time, resources, format, context etc., but in general oral questionning is considered both a necessary instrument of evaluation, and a comfortable way of keeping one's superiors happy about the marking speed. At the same time, it has a considerable impact on teaching in the sense that it disrupts the involvement in the day's topic and it significantly reduces the lesson time. A typical lesson that includes oral questionning has about two thirds of the time left for the new lesson. It means that the opening stages (Warm-up or Lead-in optionally followed by Presentation) will come much later in the lesson, and the practice activities are squeezed in the second half of the lesson. The quality of these activities varies greatly if we factor in the size of the class and the type of activity. Teachers complain that they have too little time to cover all the content in the curriculum and are therefore forced to sacrifice communication in order to focus on grammar or to make sure they cover the textbook in time, which defeats the very purpose of having an English course at all. Having to plan a year's syllabus at the beginning of the school year, teachers find that what they

include in the syllabus is allocated only half of the time. Also, learner motivation to participate in the lesson is affected by the oral questionning session whether because of the marks or because of the lack of predictability in the teacher's behaviour – in other words, rapport. In her book, *Assessing English Language Learners. Bridges to Educational Equity*, Margo Gottlieb distinguishes between many categories of learners (English language learners, long-term English language learners, emergent bilinguals, and culturally diverse language learners would be the most illustrative for Romania) and refines the types of learner motivation to account for a variety of teaching decisions and teacher responsibilities in approaching both the students and the very content of the syllabus. She places a high emphasis on rapport as a matter of not just didactic, but cultural awareness to the dynamics of the learners' interest, which confirms a real need for more refined planning.

Oral questionning does not have a prescribed format though it is commonly understood that there should be more than one question, more than one question type, and that the questions should refer to both Low and High Order Thinking Skills. The problem with administrability though is that, in order to ensure a good quality assessment, questions should be pre-tested and fine-tuned with other peer teachers, which contradicts the spontaneous character of oral questionning. Thus, lack of feedback from other teachers or any measurement as to the quality of the assessment makes it difficult to ensure the quality of this type of assessment. Last but not least, the format of the oral questionning is closely connected with the teacher's role of examiner. The sparcity of information related to class assessment in the official documents suggests that much of the organization is left to teachers. The language departments must, among other things, elaborate instruments for assessment. However, they do not have any responsibility in quality assurance and in fact there is no formal training for examiners apart from what is occasionally taught during the TEFL teacher training course at BA level. Thus, teachers must juggle the roles of interlocutor and assessor with assessment material that is not pre-tested. To make matters worse, the need to „produce" official marks, as it were, at a certain rate changes the rules of the game with each new session of oral questionning. The later in the school year the assessment occurs, the more complex it is, quantitatively and qualitatively, because new content will be included in the spontaneous questions. Consequently, the first mark of the two recorded officially for all the students in a class will in fact be given for very different reasons, which, again, is a serious drawback for quality assurance.

It is not enough to recommend that oral questionning be dropped altogether as an instrument of formal assessment even though this alone would be a major step away from a dubious tradition in education. And the root of the problem is not actually the spontaneous questioning (which is a valid evaluative and teaching technique), but the fact that it is now formal assessment – in other words, it has official consequences for the long-term evaluation of a student's performance. If used informally, oral questionning is quite a productive technique, highly administrable and motivating for learners.

So how can teachers continue to give two marks per semester and do so continuously? The solution lies in planning and cooperation. The first important duty teachers have is really to teach. Only what has been taught can be assessed, nothing outside the syllabus and nothing outside the competences of using English – not intelligence, not creativity, not emotions or attitudes. It follows that the first possible formal assessment can be the at least the second progress test after a number of weeks since the start of school. All assessment done before that date must be regarded as informal. Similarly, the next formal assessment will probably take an equal number of weeks to prepare if the learners' progress is to be reliably compared to the previous one. All formal assessment should be predictable. It should be announced as early as possible (even at the start of school) and prepared in cooperation with the students, their families, and the other teachers so that

assessment is done in comfortable conditions for the learners. If they will have a test in English, maybe teachers of other subjects will show some understanding around that time and avoid pressing the students' schedule with other tests. It is also a good idea to negotiate the time with the learners in order to find a comfortable date.

The format and scoring scheme should also be presented to the students, and assessment should include familiar operations for all the specific competences. The format should be carefully planned so as to give equal representation to all the four skills; which means on the one hand that sufficient time must be allotted for productive skills (speaking and writing) and, on the other, that an integrative test may be more efficient than a discrete item test even though the latter appears to be less time-consuming and more to the point. In their book, *How to Teach for Exams*, Sally Burgess and Katie Head give numerous examples of test patterns that include all four skills. The assessment can be split into parts that take longer and can be organized in different (but close) days, and parts that can be done together. If you need an average of at least 5 minutes to assess one student's speaking accurately, then assessing an entire class requires careful planning. Ideally, teachers can do assessment in cooperation, but there is no certainty that adequate help will be available at all times in a school year for all teachers. This is why maybe the departments (or Teacher Committees as they are called in the Order of the Ministry of Education of 2016) can compensate with some standardized practical solutions for assessment such as common assessment formats (since all assessment, as all teaching, is based on the same curriculum), a bank of materials and peer-assessment so as to help teachers handle tests single-handedly.

In terms of planning, several things can be done both on short term and long term. Accurate and reliable formal assessment is time-consuming and needs to be planned separately, which has consequences in re-thinking the entire syllabus for a year. Some of the lessons will count as teaching in the general calendar, whereas others will count either as a mixture of revision and assessment, or just assessment in their own right, with their specific stages and timing. Then, lesson plans will gain in timing practice properly by eliminating the formal oral questioning. Informal questioning can be spread throughout the entire lesson and, as a consequence, will become shorter and less intimidating for students. Instead of opening the meeting with formal assessment of previous knowledge, the teacher can just dive into the topic of the day and reserve the final minutes for remedial work if necessary, or feedback. Alternatively, (temporary) assessment can also be done quietly while the students are working. Pending on the size of the class, it can be organized in one session or in two, for half a class at a time, or groups. The results can be noted separately and added to the formal evaluation subsequently. Another strategy is to include some of the homework in the formal assesment (again, in a predictable manner). These are not new techniques, admittedly, and homework features among the designated instruments for assessment in the 2016 Order of the ME, but the above observations about oral questioning are also true of homework taken separately as an independent tool for assessment.

Developing adequate materials for assessment is an important part of quality assurance. Ideally, these materials should be based on the requirements of the curriculum – if there is a curriculum for each type of English course, then there should be adequate materials for assessment for each of the curricula. This ensures both content and face validity, and prevents tests from being used for other purposes. Then, they should be elaborated and tested in a large group of teachers, at least those teachers who use the same textbook for a certain level. Each of the stages in the process of writing a test implies a negotiation among several authors. Sampling, for example, means harvesting a wide variety of material that is representative for each component/skill of the test (Hughes, 2003: 44-45). The more test writers do it for the same curriculum, the more relevant the

test is. The same goes for writing the specifications, writing the items, testing the items and calibrating the test as a whole. Another side effect of this process is that, for the same content a large number of test versions can be generated, which helps teachers ensure the security and fairness of the test, and assess its reliability. Also, versions of a test can be adapted for students with special needs, which is hardly ever the case with formal oral questioning.

In the relationship with the students and their families, a new approach to testing encourages learner autonomy and promotes reliable self-assessment. It is imporant for students to be able to predict their future performance not only for the next progress test, but also for assessment outside the school context. If they can predict the outcome of a test, it means they were assessed objectively rather than subjectively, and that the assessment aligns with the teaching. But for this to happen, the teachers must create a bank of assessment materials to be consulted by the students (and, of course, by other teachers). The idea of a Self-Access Centre is not new, but in Romania the tradition of formal oral questioning has excluded the need for such materials being available for out-of-class use. Technically, fairness and predictability in the teacher-student relationship mean that the student is entitled to access to a collection of previous tests with keys and explanations just like any other school materials in the library. Unfortunately, this situation is yet to be addressed in the Romanian system of education.

Another shortcoming, both in teacher training and in the system of assessment, is the lack practice in writing assessment scales. Teachers find it easy to teach and test grammar using discrete items because they are obvious. However, the curriculum and the Order of the ME make it very clear this is not the same as communicative competence. To objectively assess performance for just one mark at least four scales are necessary, one for each skill (more if we add project work etc.). For a second or third mark given as a result of assessing a different content, a new set of assessment scales must be used as any subsequent mark is given for a new level of knowledge and skills.

The change in the Romanian landscape of instructional assessment depends on the understanding of the relationship between the quality of the form of assessment and the quality of further teaching. The washback effect is understood as the influence a test can have on the didactic choices prior to the test in terms of content and format. But the quality of class assessment (expressed in stability, discrimination, test length and homogeneity) impacts not just the future classroom teaching (Fulcher, 2007: 31), but the parallel systems of learner self-assessment and third-party assessment. For the time being the system of assesment implemented by the Ministry of Education remains partly rooted in the past, but the much-needed reform has already been introduced via private centres of teaching and assessment with considerable variations from the conventional practices and outcomes, and at this point any ideological divergence doe the learner a disfavour.

BIBLIOGRAPHY

Legea Educației si Învățământului nr. 28/1978 http://www.monitoruljuridic.ro/act/legea-educatiei-si-invatamintului-nr-28-din-21-decembrie-1978-legea-educatiei-si-invatamintului-emitent-marea-adunare-nationala-publicat-n-466.html
Ordinul 5079/2016 al Ministerului Educației Naționale și Cercetării Științifice http://legislatie.just.ro/Public/DetaliiDocument/181773
Programa Școlară pentru disciplina Limba modernă 1 (clasele a V-a – a VIII-a) http://programe.ise.ro/Portals/1/Curriculum/2017-progr/19-Limba%20moderna1.pdf

Braniste L., *Rolul dezvoltării competenţei comunicative în procesul de predare-învăţare-evaluare a limbii române ca limbă străină*, in „Studii şi cercetări ştiinţifice”: „Plurilingvism şi interculturalitate. Textul în context: cuvânt şi imagine”, Bacău, Editura Alma Mater, nr. 35, p. 23-28, ISSN 1224- 841X, 2016

Burgess S., Head K., *How to Teach for Exams*, Pearson Education Ltd., 2005

Fulcher G., Davidson F., *Language Testing and Assessment*, Routledge, 2007

Gottlieb M., *Assessing English Language Learners. Bridges to Educational Equity*, Corwin, 2016

Hughes, A., *Testing for Language Teachers*, Cambridge University Press, 2003

Ionescu, M., Radu, I., *Didactica modernă*, Ed. Dacia, Cluj-Napoca, 1995

Ur, P. *A Course in Language Teaching: Practice and Theory*, Cambridge University Press, 2009.

Weir, C., *Language Testing and Validation*, Palgrave Macmillan, 2005