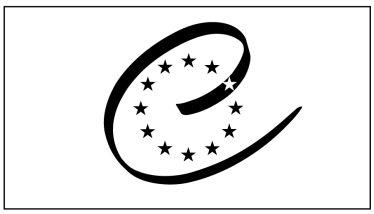
Contents



Language Policy Division
Division des Politiques linguistiques

Evaluation and assessment within the domain of Language(s) of Education

Waldemar Martyniuk (ed.), Mike Fleming, José Noijons

Languages of Education

Language Policy Division, Strasbourg
www.coe.int/lang

Evaluation and assessment within the domain of Language(s) of Education

Waldemar Martyniuk (ed.), Jagiellonian University, Poland
Mike Fleming, University of Durham, United Kingdom
José Noijons, CITO, The Netherlands

Intergovernmental Conference

Languages of schooling within a European framework for Languages of Education: learning, teaching, assessment

Prague 8-10 November 2007

Organised by the
Language Policy Division, Council of Europe, Strasbourg
in co-operation with the
Ministry of Education, Youth and Sports, Czech Republic

The opinions expressed in this work are those of the authors and do not necessarily reflect the official policy of the Council of Europe.

All correspondence concerning this publication or the reproduction or translation of all or part of the document should be addressed to the Director of School, Out of School and Higher Education of the Council of Europe (F-67075 Strasbourg Cedex).

The reproduction of extracts is authorised, except for commercial purposes, on condition that the source is quoted.

Language Policy Division
DG IV – Directorate of School, Out-of-School and Higher Education
© Council of Europe

Contents

The Challenge of Assessment within Language(s) of Education
Mike Fleming 9

The Relevance of International Assessment for the development of a Framework for the Languages of Education
José Noijons.....17

Assessing competences at the end of compulsory schooling – the Polish case
Waldemar Martyniuk.....23

Evaluation and assessment within the domain of Language(s) of Education

Introduction

Evaluation and assessment of competences related to language(s) of education is an area of special interest within the Council of Europe's project examining the feasibility of developing a common European reference document for the LE domain. The discussions carried out so far within the Working Group 'Evaluation and Assessment'[1] can be summarised as follows:

1. The fundamental question to be answered when making decisions regarding assessment is 'WHY do we want to have an assessment scheme?' The most problematic area is usually not the assessment itself but the use made of the results, the claims made on the basis for the assessment, and its impact.
2. Suitable assessment solutions and approaches can be offered only once attainment targets (competence standards) have been clearly identified.
3. Assessment designers should 'negotiate' with curriculum developers when attainment targets (competence standards) are being turned into assessment tasks and items.
4. Only those attainment targets (competence standards) that may be turned into observable behaviour can become the subject of assessment procedures (testing tasks/items).
5. Issues such as validity and reliability can be first addressed at the level of items and tasks.
6. Item level testing may not be capable of/suitable for handling all of the complexity of LE competences.
7. Sampling may be the only feasible option if educational systems (not only individual learners) need to be evaluated.
8. Introducing standards and assessment schemes should be a bottom-up process (reflecting the needs of learners and teachers) rather than a top-down procedure (reflecting only the needs of decision makers).
9. Standards and assessment schemes should be promoted as assistance tools rather than control measures.
10. The envisaged LE framework document should serve as an awareness-raising and reflective tool, broadening the notion of evaluation and assessment beyond testing and levels.

Below we present a set of studies dealing with some of these issues. The opening paper by Mike Fleming is an overview article listing and addressing challenges related to evaluation and assessment within the domain of LE that may serve as a basis for a future chapter in the envisaged LE framework document. The paper concludes with a proposal called 'Languages of Education Portfolio' – an integrated approach to evaluation and assessment within the domain of LE.

The two case studies that follow illustrate how the issues raised in the overview article are being turned into practice on an international and a national level. The point made by José Noijons when presenting the two international assessment frameworks – PISA and PIRLS – is that even if people, specifically teachers, may not be in favour of internationally produced standards, they are already with us and national policy makers wish to adhere to these

[1] Group members: Christoph Arnold, Mike Fleming, Waldemar Martyniuk (co-ordinator), José Noijons

standards. His suggestion for the work on an LE framework document is to acknowledge the already established conceptual frameworks of international surveys and the standards that go with them.

The presentation of a national-level system of compulsory external examinations where competences in a language of schooling – Polish – is assessed, is an attempt to examine how the challenges to standardise and assess competence in LS and LAC are being met in practice. Various forms of external assessment schemes are in use in most of the member states of the Council of Europe. 29 out of the 44 national respondents (including 8 from the German *Länder* and 5 from the United Kingdom) to a Preliminary survey on curricula for teaching national/official/school languages in compulsory education, carried out by the Language Policy Division in April 2005, claimed to have external examinations in place. A new survey, targeting specifically the use, the scope and the significance of these examinations was sent out by the Language Policy Division in mid July 2007. The results of the new survey will certainly serve as very valuable input for the work on an LE framework document.

Waldemar Martyniuk

The Challenge of Assessment within Language(s) of Education

Mike Fleming, University of Durham, United Kingdom

Introduction

The purpose of this paper is to provide an overview of some of the key issues and challenges related to evaluation and assessment within the domain of language(s) of education. It incorporates elements of the previous paper distributed at the Strasbourg conference in 2006[2] but is also addresses new more practical issues related to (a) the possibility of situating assessment for language as school subject (LS) within the wider construct of language(s) of education (LE); (b) the need to reconcile portfolio approaches with more formal testing. Assessment is important but it is not without controversy and it can easily lead to polarised views and unhelpful tensions. This is particularly the case in the context of language as school subject (LS) because of the diversity and complexity of its aims. Constructive debate around differences of opinion is always helpful, but too often disagreements about assessment become entrenched and unproductive. This happens for a number of reasons, including:

- a failure to recognise that assessment needs to fulfil a wide range of legitimate purposes;
- an assumption that a single assessment tool will be able to serve all needs;
- a lack of awareness that it is the *use* made of assessment, not necessarily the assessment process itself, that will largely determine its impact;
- a tendency to search for universal solutions to assessment issues and neglect the significance of context.

Teachers of language as school subject are sometimes hostile to the idea of large-scale or formal testing on the grounds that it diminishes the subject and ignores the significance of context. This view needs to be considered.

Section One will consider a range of different purposes of assessment based on different potential audiences. An ideal assessment strategy would meet the needs of all those interest groups. Section Two will examine different approaches to assessment in relation to language as school subject (LS), including the value of portfolio assessment. Section Three will consider the implications of situating the assessment of LS within a broad strategy of assessment within languages of education (LE). Section Four will develop further the concept of an integrated approach to assessment which seeks to reconcile its different purposes within an assessment strategy. It will consider the view that a portfolio approach is not necessarily in conflict with more formal testing, as is often assumed.

1. Purposes of assessment

In its simplest formulation, assessment provides information on whether teaching/learning has been successful. However the information it provides has a number of potential different audiences whose precise requirements may vary. Classroom teachers need regular information on how pupils' knowledge, skills and understanding are developing, both to inform how they should adjust their teaching and to determine what kind of feedback is needed to improve pupils' learning. On the other hand, school principals and policy makers need additional, broader information on the quality of education in a school or country. The sort of comparative data required for this purpose needs a high level of reliability and uniformity. In the case of language as school subject this requirement is

[2] Intergovernmental Conference on "*Languages of Schooling: towards a Framework for Europe"*, Strasbourg, Council of Europe, 16-18 October 2006. See the Report on: www.coe.int/lang

challenging because it is difficult to create tests which are manageable but at the same time faithful to the aims of the subject. Employers and society at large also need reliable information which can help certify achievement and provide a basis for selection. Parents too require information which can help them understand their children's achievements and limitations. Learners themselves need to know how they are progressing and how to improve their performance but they may need to be protected from the potentially demotivating effects of negative assessment.

The concept of 'accountability' when used in relation to assessment usually refers to the imposition of systems of assessment external to the learning process as a form of 'policing' of standards to ensure that the education system is functioning effectively. But the term may be employed more broadly and more positively than this, referring to the different obligations that are relevant to all. Teachers have a responsibility to the learner but also to the needs of the wider society. Policy makers clearly have a duty to the public and need to ensure that the education system is delivering results but they also have responsibilities to the individual learners and need to consider consequences of policies in those terms. The concept of accountability interpreted in this way will take people outside of vested interests in order to see the larger context. Accountability needs to be linked with a process of sharing perceptions and fostering understanding. It is important therefore not to exaggerate differences between different potential 'stakeholders'; what all parties have in common is a fundamental concern that assessment should help raise achievement and improve learning. A starting point for resolving tensions related to matters of assessment is to develop understanding of other points of view. A key challenge is to develop a system of assessment that acknowledges the different functions of assessment and it helps to see these as complementary rather than being in opposition to each other.

2. Approaches to assessment

The different purposes of assessment lead to different approaches to assessment. Traditionally assessment of language as school subject took a very simple form: pupils were given a narrow written task which was then awarded a grade or mark. This allowed them (and the teacher) to make a judgement of how they ranked in relation the rest of the group (normative assessment). However the absence of clear criteria meant that the information rarely gave an indication of how they could make progress in their learning. Also the test itself often embodied a very narrow conception of what competence in language entailed; it often centred on knowledge of language form and structures (syntax and grammar) and a narrow range of language uses (often only a written form of essay). The implicit understanding of what reading literacy involved was also very narrow, often involving just decoding and literal understanding. In traditional approaches of that kind there was unlikely to be any attention to oral work, to a range of writing purposes, to a wide range of reading and response to reading. On the other hand, the advantage of a fairly narrow approach to assessment was that is was easier to provide reliable outcomes; the more complex the system of assessment becomes, the more difficult it is to ensure that the award of grades or marks for particular outcomes are consistent.

In approaches to assessment, two central tendencies emerge which are relevant to language as subject. One places emphasis on the assessment *of* learning where reliable, objective measures are a high priority. The focus here is on making summative judgements which in practice is likely to involve more formal examinations and tests with marks schemes to ensure that the process is sound. An alternative approach is to change the emphasis from assessment *of* learning to assessment *for* learning, implying a more formative approach where there is much more emphasis on feedback to improve performance. The approach here might be through course work and portfolio assessment in which diverse information can be gathered which reflects the true broad nature of the subject.

Portfolio assessment has a number of advantages for language as school subject; for example, it can motivate and empower the learner, it can provide samples of performance collected over time, evidence of use and awareness of process. Portfolio assessment incorporates evidence derived from more realistic tasks in meaningful contexts, rather than relying on artificial, decontextualised tasks undertaken in timed conditions. A further advantage of this approach is that it can embody different forms of self-assessment which can also be helpful ways of motivating learners and having them reflect on their progress. Self-assessment encourages pupils to take responsibility in the learning process although it is advisable for them to be trained in self-assessment techniques for this to work effectively. The difficulty with portfolio assessment if it is conceived *only* as the accumulation of evidence produced in informal settings is that is does not easily satisfy demands for reliability. Work which has been produced over an extended period of time, with formative guidance from the teacher and collaboration with classroom peers is not always convincing evidence of competence.

The broader the approach to assessment (incorporating the judgement of a range of different performances in different contexts), the more it can be said to constitute a meaningful assessment of performance in the subject. However, as suggested, tension emerges because it is sometimes difficult to compare, with any degree of accuracy, the results drawn from broad approaches to assessment. The quest for 'objective' and reliable methods of assessment driven by narrow ideas of accountability brings with it a number of dangers. So called 'teaching to the test' may not be a problem if the tests are sophisticated and wide-ranging but there may be practical difficulties in administering those that are too complex. If the tests are too narrow and simplistic then this may have an adverse effect on the teaching. Here then is one source of polarised opinions, one stressing the importance of objectivity, reliability and summative judgement, the other more tolerant of subjectivity in order to ensure that the assessment approach is faithful to the complexity of the aims.

A key concept is embodied in the notion of 'transparency', the view that those being assessed are aware of the criteria which are being used to make judgements about them and how those judgements are made. Knowledge of criteria can help performance and improve motivation but once again, in the context of language as subject, the issues are more complex than they first seem. A common assumption is that pupils learn best when they know what they are trying to achieve and why. While this view is largely true, there are exceptions. Because the development of language can in some ways be described as a 'natural' process learners do not always need to be fully focused on specific aspects of their performance in order to improve. In fact too much focal awareness on performance can make them too self-conscious: speakers can appear too groomed and artificial; the writer who has been told to strive for effect by using more adjectives may develop a highly artificial and awkward style. These insights do not negate the importance of transparency as a principle but highlight the fact that in pedagogical practice the principle needs to be interpreted and implemented with care.

The use of competences to describe outcomes which can be assessed has developed considerably in recent years. It is worth noting that the term 'competence' is used by writers in different ways which can be a source of confusion. Sometimes it is used in a very general way as a synonym for 'ability' or capability as in 'language competence'. Other writers use the term to describe broad language modes or domains such as reading, writing, speaking and listening. More commonly however 'competences' refers to the specific actions which a learner must perform and which in turn can be assessed to demonstrate achievement in a subject. Advocates of using competence statements for assessment purposes and syllabus design see their value largely in bringing clarity and transparency to the specification of learning outcomes. Critics of a competence approach take the view that performance statements are too narrow and specific, and do not reflect the range and subtlety of what is involved in language development. There are parallels

here with differences of opinion over the relative merits of portfolio and formal testing. The balance of advantages and disadvantages needs to be considered.

One of the challenges posed by assessing language as school subject (LS) is that the content is so varied and complex (see the paper by Florentina Sâmihaiăn[3]) which is in turn a reflection of the complexity of the aims. Each mode of 'writing', 'reading', 'speaking 'and 'listening' can be broken down into further areas. A subject that is so multidimensional raises the question as to whether an assessment task in one area is representative of achievement in the subject as a whole. For example, it is fairly safe to assume that performance in speaking is not necessarily indicative of reading competence. On the other hand it is less clear whether it is necessary to assume that reading ability varies in relation to texts of different types (fiction, non-fiction, media). The assessment of reading can easily remain at a surface level only addressing recall or literal comprehension rather than deeper understanding. Multiple choice questions on a text are easy to mark and may yield high reliability (in the technical sense) but are less equipped to assess the learner's deep and individual response to a text. Even with very young children the reading process is more than simply decoding text and any system of assessment needs to reflect that fact. It is the complexity of the subject that accounts for the hostility teachers of languages as school subject sometimes express towards large-scale and formal testing because only a fairly narrow range of competences can be assessed in a single test.

A similar question arises in relation to writing. Do pupils need to be assessed on a range of different tasks reflecting the fact that writing exists for different purposes and for different audiences? The assessment of writing appears to be more straightforward than that of reading and speaking because at least there is always a product which can be referred to after the event. But there is a major challenge in determining what criteria should dominate in making a judgement (for example the accuracy of the writing as opposed to the impact of the content) and whether the criteria should change in relation to different types of writing. It is also wrong to assume that the only response to pupils' writing comes when it is completed. Formative assessment in the form of a dialogue about the work in process is an important way of improving standards. Assessment is always a selection and therefore can unwittingly value some aspects of the subject more than others.

Speaking and listening is extremely difficult to assess because, even more than other aspects of LS, performance varies with the theme, context and level of motivation. Poor performance in oral work is often to do with the nature of the task which has been devised, the atmosphere of the classroom and the dynamics of the group rather than the competence of the pupils; to provide a valid assessment of speaking and listening, evidence needs to be drawn from a variety of situations. Some might argue that the assessment of speaking and listening is so complex and context specific that it should not be assessed formally. However because assessment so often determines the curriculum and the way it is taught there are arguments to suggest that speaking and listening should be assessed despite the difficulties; the ability to articulate a point of view orally and to argue a case are essential skills for meaningful participation in a democracy.

An ideal assessment system would reflect the full complexity of language as school subject (LS), and would motivate learners by giving useful feedback, while also providing other stake-holders (e.g. policy-makers and employers) with the information they need. An integrated approach to assessment would ensure that the different purposes and approaches are balanced so that no one priority has adverse and undue influence on the system as a whole.

---

[3] Sâmihaiăn F. *Content considerations for a framework of reference for Language(s) of School Education,* www.coe.int/lang

3. Language(s) of Education (LE)

So far consideration has been given to the challenges presented by assessment of 'language as school subject' (LS). Is there any advantage in considering the assessment of LS within a broader framework of language(s) of education (LE) which incorporates 'language across the curriculum' (LAC) and foreign language learning (FL)? 'Languages(s) of education conceived in that way is not a subject but an umbrella construct; it is fairly clear therefore that the notion of examining or testing language(s) of education is entirely inappropriate. However it is conceivable to profile a pupil's competence in language(s) of education by assembling their competences and achievements in a range of domains. The present Council of Europe's European Language Portfolio (ELP) is a useful model to demonstrate the type of instrument which could be considered. This approach has a number of potential advantages. It takes seriously the concept of plurilingualism in the school context: a pupil may be a low achiever in language of schooling taught as school subject but be fluent in two or more other languages. Thus the deficit model which is often applied to assessment of LS is replaced by a more positive recognition of achievement. It would still be important to be able to identify an individual's competence in aspects of LS as a component within a broader profile, but being able to situate that description of competence in a larger context could have a positive impact on motivation and self-esteem.

Could portfolio assessment extend to embrace language competence in other subjects, to include the dimension of language across the curriculum? This is a practical challenge but perhaps not insurmountable. Some uses of language (e.g. giving presentations, writing formal reports, reading for information) are clearly required and demonstrated within different subjects and it is not inconceivable that those subjects should make a contribution to a pupil's overall language profile. There is an argument to suggest that if developing language competences across the curriculum is to be taken seriously it must have some impact on how language use is assessed. Underlying questions have to do with subject boundaries and the degree to which competence in language use can be easily separated from the context in which it is used; learners of history could also in some sense be said to be learning the language of history.

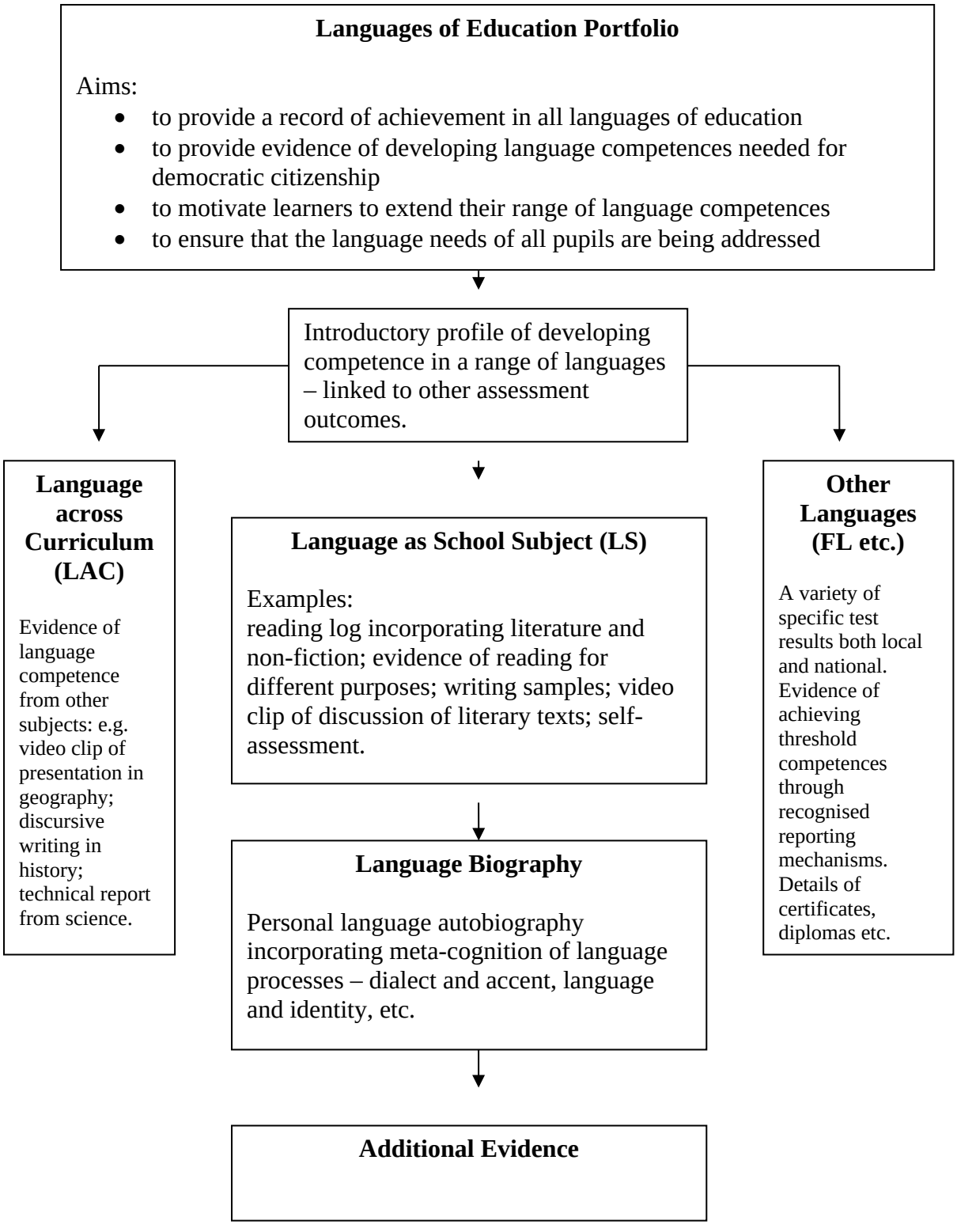
4. Integrated assessment

As described earlier a key challenge and source of tension in relation to assessment is to satisfy the different purposes assessment is expected to fulfil and to provide appropriate information for the different interest groups. As suggested, teachers of language as school subject are sometimes hostile to large-scale, formal testing because of the perceived narrowing of the subject. There are a number of points to consider in relation to this view. The testing 'industry' has become increasingly sophisticated in recent years and designers of assessment tasks and items are more adept at addressing issues of validity and reliability. That does not mean that formal tests can address all language competences but once that fact is recognised then the results of such tests may make a useful *contribution* to the overall profile of a pupils' competence in language use. Large-scale testing both at national and international level is a fact of modern life. Policy makers need information on levels of proficiency achieved by groups in schools, education authorities or countries. It is possible to discharge this 'evaluative' function of assessment by a process of sampling rather than by using summative data derived from an entire population, but it is unlikely that the quest for information of this kind will diminish. When assessment data is used to compare the progress of different cohorts of pupils, the use of a 'value-added' approach which takes account of the different base-lines from which the pupils are progressing is becoming more common. All of this information can be informative for teachers as long as the limitations of the data are recognised.

The concept of an integrated approach to assessment is intended to counter the tendency towards polarised views. As suggested above, portfolio assessment has many advantages

but without information based on some form of 'controlled' performance it will struggle to make convincing claims to reliability. The incorporation of specific results derived from appropriate tests designed to assess specific (not all) language competences into a portfolio can help strengthen both the validity and reliability of the assessment approach. The key may be to think broadly in terms of an assessment strategy which makes use of a variety of assessment tools, rather than assuming that one assessment tool will fulfil all purposes.

A sample outline of a Languages of Education Portfolio based on the present Council of Europe model is given below:


Languages of Education Portfolio
Aims:
to provide a record of achievement in all languages of education
to provide evidence of developing language competences needed for democratic citizenship
to motivate learners to extend their range of language competences
to ensure that the language needs of all pupils are being addressed
Introductory profile of developing competence in a range of languages – linked to other assessment outcomes.
Language across Curriculum (LAC)
Evidence of language competence from other subjects: e.g. video clip of presentation in geography; discursive writing in history; technical report from science.
Language as School Subject (LS)
Examples:
reading log incorporating literature and non-fiction; evidence of reading for different purposes; writing samples; video clip of discussion of literary texts; self-assessment.
Other Languages (FL etc.)
A variety of specific test results both local and national. Evidence of achieving threshold competences through recognised reporting mechanisms. Details of certificates, diplomas etc.
Language Biography
Personal language autobiography incorporating meta-cognition of language processes – dialect and accent, language and identity, etc.
Additional Evidence

The Relevance of International Assessment for the development of a Framework for the Languages of Education

José Noijons, CITO, The Netherlands

Introduction

In an ideal educational system educationists with different backgrounds sit together when new learning programmes, curricula, examination syllabi and such like are being developed. In practice it may happen that testing experts are called in rather late, when descriptors, targets, aims, levels etc have already been formulated. In this paper we hope to illustrate how useful it can be to look at the achievements in International Assessment when developing a Council of Europe's Framework for the Languages of Education (referred to here as an *LE Framework*). We will discuss what it is and does, what its salient characteristics are, but we will also dwell on its limitations. It is much appreciated that the Council of Europe has invited testing experts to take part in the development of an LE Framework. We hope that the testing expertise can be of use here, for example through showing examples of good and relevant practices in language testing.

1. What is International Assessment?

International Assessment (IA) is assessment that allows for educational achievement to be compared across countries. It is concerned with measuring trends or progress. IA may identify percentages of students in countries who meet standards for a given subject. It helps stakeholders in education – the teaching profession and policy makers – to measure the success of their country's educational policies. IA helps to make sure that children will have the knowledge and skills necessary to be productive members of an improving economy and effective citizens of a developing democracy[4]. IA is cyclical: depending on the programme measurements are repeated every three or five years. On the basis of the outcomes it can be shown if countries have progressed compared to their performances during earlier cycles, in absolute terms and also in relation to other countries.

2. Who is tested?

In the two IA programmes we will refer to here, the Programme for International Student Assessment (PISA) and the Progress in International Reading Literacy Study (PIRLS), *samples* of students are tested. These samples can be either grade- or age-based. In the case of PIRLS, they are fourth-grade students, roughly equivalent to nine and ten-year old students. In the case of PISA, they are 15-year-old students. These grades and age-groups have been chosen for a particular reason. In many countries grade-four students are about to make the change from *learning to read* to *reading to learn*[5]. Often there will be a change in the teaching circumstances. In grades one to four one teacher teaches all or most subjects, from grade five onwards in many systems there will be subject specific teachers. In the case of PISA, 15-year-old students are about to end compulsory education and they are now supposed to have mastered those skills that will enable them to function independently in society.

There are more differences between the two IA programmes: PISA has a three-year cycle and PIRLS has a five-year cycle. PIRLS only assesses reading literacy, PISA assesses reading literacy, mathematical literacy and science literacy. Although both programmes share the assessment of reading literacy, their construct of what reading constitutes differs. We shall discuss these differences in more detail later.

[4] See the NAEP programme: National Assessment of Educational Progress, www.ed.gov/programs/naep/index.html

[5] See the PIRLS Assessment Framework: www.timss.bc.edu/pirls2006/framework.html

3. The use of international assessment

As we indicated, IA helps the teaching profession and policy makers to evaluate the success of a country's educational programmes. It is *not* used to assess individual students. Samples of students throughout the years are assessed and even though data are collected from individual students, not all students are administered the same test. Thus performances of individual students are not usually compared to each other. The size of the sample usually does not even allow individual *schools* to be compared to each other.

4. The collection of background variables

Both IA programmes do not only collect data through the assessment of cognitive skills (in this particular case: reading literacy), they also collect data on background variables. These background variables may be *student* variables such as gender, age (in PIRLS), academic success, motivation, family circumstances (languages are spoken at home, social-economic status), exposure to reading texts, attitudes to reading etc; *school* variables: location, size, equipment, number of staff, number of students etc. Background variables may also relate to educational policy or to teaching methods, to mention some more. Data are collected through questionnaires that may be administered to students, parents, teachers, school principals, educational policy makers or others.

Once data have been collected both on the cognitive tests and on the questionnaires, countries may like to relate student performance on the tests to the background variables. Especially malleable factors, background variables that can be changed in the light of the outcomes on the cognitive tests, would seem to be of special relevance in the context of the Council of Europe's brief and an LE framework in particular. If representative samples of identifiable groups of students under-perform on cognitive tests there may be reasons to look at the background variables of these groups and find out if there are significant correlations. If a theory can be developed that would predict an improvement in performance through a change in background factors, a country might consider to change relevant elements in its educational policy. This latter type of research can be very interesting and revealing for countries. However, from IA programmes in progress we know that research into these areas can be quite sensitive.

5. Constructs of Language Proficiency

Both the PIRLS and the PISA programmes have developed Assessment Frameworks. These contain definitions of the *construct* of what is tested (reading literacy), among others. Of course the two programmes differ in their constructs for reasons given above.

For the PIRLS assessment, reading ability for fourth-grade students is defined as:

> *…the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to earn, to participate in communities of readers in school and every day life, and for enjoyment.*

For the PISA assessment, the construct of reading literacy for 15-year-old students is defined as follows:

> *Reading literacy is understanding, using, and reflecting on written tests, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society*.

What is important here is that both IA programmes *define* the construct to be assessed, they define what they consider reading literacy to be. This is relevant in the context of the development of an LE Framework. Those who are familiar with the Council of Europe's framework for foreign language learning, teaching and assessment, the Common European Framework of Reference (CEFR), will know that that framework does *not* offer a construct definition of foreign language competence. Rather, it defines levels of competence

through descriptors, through the so-called can-do statements. Critics considered this to be one of the serious weaknesses of the CEFR.

6. To what language(s) does reading literacy in the PIRLS and the PISA programmes refer to?

In fact in neither of the IA programmes' constructs explicit reference is made to the language in which reading literacy is assessed. It is left to the countries participating to choose which language(s) this/these would have to be. It would seem that in both programmes language as a school subject (LS) may be implied, if in that language the skill of reading is taught. However, strictly speaking the language itself is not part of the construct; it is about reading *per se.* In most countries this means that the tests are administered in the language(s) of instruction. In the case of PISA, attempts have been made to assess reading in the context of language across the curriculum, as part of an umbrella construct, not so much with reference to foreign languages but with reference to math and science. The reason for assessing reading in this way is that cross-curricular skills are thought to be *life-skills* that modern education systems often strive to train their students in.

7. Assessment Frameworks

Both programmes further define the construct they are testing. We give examples from PISA here, but PIRLS has comparable definitions. In the process of reading, PISA distinguishes a number of components that can be manipulated in the testing: the reading situation, the structure of the text and the characteristics of the questions (items) that are asked about the text. Each of these factors is further specified.[6]

Text Format

In the PISA tests, a distinction is made between two text types:

- Continuous texts, typically composed of sentences that may be organised in turn into and paragraphs, sections, chapters and /or books;
- Non-continuous texts, such as charts and graphs, tables and matrices, diagrams, maps, forms, information sheets, calls and advertisements, vouchers and certificates.

Reading processes

In the PISA tests five reading processes are distinguished:

- Retrieving information;
- Forming a broad general understanding;
- Developing an interpretation;
- Reflecting on and evaluating the content of a text;
- Reflecting on and evaluating the form of a text.

Item types

In the PISA 2000 and 2003 studies, around 43 per cent of the reading literacy tasks were open constructed-response items which required judgment on the part of the marker. The remaining tasks consist of closed constructed-response items that require little judgment on the part of the marker, as well as simple multiple-choice items, for which students choose one of several alternative answers, and complex multiple-choice items, for which students choose more than one response.

[6] See the PISA Reading Framework: www.pisa.oecd.org

Situations

In the PISA tests the four situation variables as identified in the CEFR as *domains of language use* have been taken over:

- Reading for private use (personal);
- Reading for public use;
- Reading for work (occupational);
- Reading for education.

8. Levels in Reading Literacy

PISA also distinguishes levels – or standards – of reading proficiency. Items have been scaled (on the basis of the response data) from less difficult to more difficult. The members of the PISA reading expert group and test developers identified a number of variables that can influence the difficulty of any reading task. The progression of difficulty in the set of test items was captured in a composite reading literacy scale with three reading processes (*retrieving information, interpreting texts* and *reflecting and evaluating*). For each of the three subscales five levels were identified. Expert panels judged that the tasks within each level of reading literacy shared many task features and requirements, and differed systematically from tasks at higher and lower levels. In this way a reading literacy level map could be drawn up.

This map somewhat resembles the levels as distinguished in the CEFR. However, the PISA levels are based on *response data* from students answering the questions. Judges in PISA grouped tasks within levels and characterized the levels on the basis of the content of the items. In the case of the CEFR the *descriptors* have been scaled and levels have been formulated in terms of these descriptors. Validation was done through comparison with student performance. Items in CEFR-related tests have been derived from these descriptors. In that way the CEFR approach differs considerably from procedures in IA.

9. Bias in IA

As part of IA, research has been carried out into the phenomenon that the same item seems to measure different things in different countries. Thus it happens that this same item is found to be easy for weak students in one country and more difficult for more able students in another country. This phenomenon is called Differential Item Functioning (DIF). For the sake of validity such items tend to be removed from a testing system. Yet they may teach us something about the differences in education policies and testing practices between countries, much in the same way as data from background questionnaires may tell us that. Both approaches may help us to develop theories why there are differences between performances between countries. This in turn can help countries to change elements in their educational systems if they so wish.

10. The limitations of IA with respect to the development of the LE Framework

We must be aware that most IA is limited in its aims. These are limitations as to the skills that are tested: receptive skills, rather than productive skills, and to the age groups tested. There are some *national* assessment systems that do more, but their limitation is that only one educational system is surveyed. For an *international* LE Framework the format of national assessments and their data may be too specific.

11. The importance of IA with respect to the development of an LE Framework

The most relevant feature of IA for an LE Framework would seem to be that it is based on an assessment framework that incorporates a definition of the construct its tests are based on. The next important phenomenon in this context is that this framework and construct is accepted by a great many countries (commitment is essential), otherwise countries would not take part in the survey. Even if these surveys contain low-stakes tests from a student's point of view, for the countries' educationalists these surveys are high-stakes. They would not accept data from programmes whose fundamentals they did not agree with. In the case of PISA 2009, some 65 countries are participating. It therefore would seem to be possible to develop an LE Framework that is acceptable to a large number of member states of the Council of Europe, certainly if it is based on the principles of the Council of Europe that all member states have signed up for.

It goes without saying that the international language testing community is happy to be involved in the development of an LE Framework at an early stage. If that framework is to be an influential policy instrument it will be useful to develop instruments that can test the extent to which countries achieve the aims (derived from the Framework) they have set themselves. But we must remember that we cannot produce tests on the basis of aims, goals, standards that are not translatable into assessment targets.

12. Recommendations for the development of an LE Framework

Using the experiences of some programmes of International Assessment, it would be advisable to pay attention to the following while developing an LE Framework:

- Define the construct of competence in Languages of Education.
- Define what concepts in the Council of Europe's Charter this construct is based upon.
- Seek support from member states in accepting the construct of competence in Languages of Education.
- Define what skills learners need to acquire to be competent in Languages of Education.
- Identify key stages in the acquisition of competence in Languages of Education.
- Specify skills and sub-skills at key stages in the acquisition of competence in Languages of Education.
- Define standards that typically should have been reached at key stages in the acquisition of competence in Languages of Education.

Activities such as the above may result in an LE Framework that will be challenging for language testers to produce valid tests for. These tests can be used by member states to gauge to what extent they have achieved the goals of the Council of Europe in the area of Languages of Education.

Assessing competences at the end of compulsory schooling – the Polish case

Waldemar Martyniuk, Jagiellonian University, Kraków, Poland

Introduction

The aim of this paper is to present a system of compulsory external examinations where competences in a language of schooling – Polish – are assessed. It is an illustration of how the challenges of assessing competence in LS and LAC are being met in practice. The paper starts with a brief introduction to the Polish system of education, continues with an overview on the national system of external examinations and concludes with a brief presentation of one of the examinations – the Lower-Secondary School Examination (*Egzamin gimnazjalny*).

1. Education in Poland

In accordance with the School Education Act of 7 September 1991 the Polish education system comprises pre-school institutions, primary schools, lower-secondary and upper-secondary schools. Institutions of higher education form a separate higher education system or sector.

A child aged 3 to 5 may receive pre-school education, which is not compulsory. In the school year 2004/05 a one-year obligatory pre-school preparation for 6-year-olds was introduced, to be provided either by kindergartens or primary schools.

Children aged 7-13 attend a 6-grade primary school (*Szkoła podstawowa*). Education in primary schools is divided into two stages: the first stage (grades 1 to 3) offering subject-integrated learning and the second stage (grades 4 to 6) at which subject teaching is provided. At the end of the 6-year primary school pupils take their first external interdisciplinary competence test. This test was introduced in April 2002.

All primary-school leavers continue their education in a 3-year *Gimnazjum*, a compulsory lower-secondary school. At the end of this cycle pupils take their second compulsory external examination organised by a regional examination commission. Upon the completion of the *Gimnazjum*, they may choose between two options:

- *Liceum profilowane* (3-year specialised upper-secondary school with the external *Matura* examination at the end);
- *Zasadnicza szkoła zawodowa* (2-year vocational upper-secondary school).

Graduates from vocational schools may also take the *Matura* examination upon successful completion of a 2-year supplementary general secondary school (*Uzupełniające liceum ogólnokształcące*) or a 3-year supplementary technical secondary school (*Technikum uzupełniające*). The new external *Matura* examination introduced in the spring of 2005 has replaced entrance examinations to universities.

The Act of 25 July 1998 amending the School Education Act of 1991, Article 15, defines compulsory education in Poland in the following way:

- education is compulsory until the age of 18;
- full-time compulsory education starts during the calendar year in which the pupil reaches 7 years of age and lasts until the end of *Gimnazjum* (lower-secondary school), yet no longer than when he/she reaches 18 years of age.

Upon completion of a *Gimnazjum,* compulsory education can be implemented through:

- attendance of a public or private post-*Gimnazjum* school;

- participation in classes organised in out-of-school forms in public or non-public institutions (with accreditation);
- participation in classes organised by legal bodies or individual persons with accreditation for running educational activities;
- vocational training offered by an employer.

2. Lower-secondary education

Teaching at this stage is arranged in subjects taught by specialist teachers. Since 2002 there have been national Core Curricula for all subjects taught. The Core Curriculum has to be respected by every school, although teachers are free to follow one of the authoring curricula based on the Core and approved for use in schools by the Minister, as well as use various textbooks selected from a list approved by the Minister.

Besides separate subjects, the following "educational paths" have been introduced at this stage:

- philosophical education;
- reading and media education;
- health education;
- ecological education;
- regional education – cultural heritage of the region;
- civil defence;
- European education;
- Polish culture in the context of Mediterranean civilisation.

Each School Head is responsible for the inclusion of these paths in the school curriculum. Implementation of these paths is assured by subject teachers through the integration of their content in the subject curricula.

3. Assessment

The education system reform, started in 1991, encompasses the introduction of a new coherent evaluation system including internal and external assessment.

Internal evaluation is carried out by school teachers on the basis of requirements defined in relation to the curricula implemented in the given school. This type of evaluation aims mainly at supporting the pupil's development.

External assessment is carried out by institutions external to the schools and it is based on national examination standards. These are defined and announced in the regulations issued by the relevant minister of education. In order to prepare and implement this type of assessment, a Central Examination Commission and eight Regional Examination Commissions have been established.

3.1. Internal Evaluation

Pupils are assessed separately in each subject by their teachers. The results of the formative assessment carried out during a school year are taken into account in the end-of-year summative assessment.

In grades 1-3 of primary school the evaluation is only descriptive and applies to educational attainments and behaviour (conduct) separately. Starting with grade 4, each subject teacher has the following scale of marks at his/her disposal: 6 - excellent, 5 - very good, 4 - good, 3 - satisfactory, 2 - acceptable, 1 - unsatisfactory. School marks as well as the criteria on which they are based should be open to the pupil and his/her parents.

Pupils also receive marks for their behaviour (conduct) according to the following scale: excellent, very good, good, acceptable, unacceptable, and inadmissible. The mark for behaviour (conduct) should not influence subject marks or promotion to a higher grade.

Pupils may take a verifying examination if the mid-year or end-of-year summative mark, given by the teacher, is too low in their or their parents' opinion. This kind of exam is also commissioned for pupils whose results cannot be assessed for other reasons, such as an individual study programme, out-of-school study or a long period of illness.

3.2. External Evaluation

The new system of external assessment, implemented gradually from 2002 on, aims at measuring the achievements and identifying shortcomings of students' education, assessing the effectiveness of teaching, and comparing objectively current certificates and diplomas irrespective of the place where they have been issued.

The external examination system in Poland encompasses:

- The Competence Test (*Test kompetencyjny*) taken at the end of the 6-grade primary school;
- The Lower-Secondary School Examination (*Egzamin gimnazjalny*) conducted at the end of lower-secondary school (9th grade);
- The *Matura* Examination for graduates of general secondary schools, specialised secondary schools, technical secondary schools, supplementary secondary schools or post-secondary schools;
- The Vocational Examination (examination confirming vocational qualifications) for graduates of vocational schools, technical schools and supplementary technical schools.

3.2.1. The Competence Test (*Test kompetencyjny*)

The Competence Test, introduced in 2002, is administered at the end of primary school (Grade 6) and is the first, compulsory external examination a pupil experiences in his/her school education. It is usually conducted at the pupil's own school. The results do not have any influence on a pupil's promotion and should not be the basis of any selection process. Any pupil who has graduated from primary school, independent of test results, must be admitted to a lower-secondary school in the appropriate area (until the age of 16).

The test content integrates all school subjects and tests the skills and abilities specified in the attainment standards related to: reading, writing, reasoning, using information, applying knowledge. Pupils have 60 minutes to complete the test and can get a maximum score of 40 points. External examiners mark the exam papers.

3.2.2. The Lower-Secondary School Examination (*Egzamin gimnazjalny*)

(See the detailed description below, section 4.2.)

3.2.3. The *Matura* Examination

The new *Matura* Examination, implemented fully in 2005, is an external examination, uniform throughout the country, but not compulsory. It was introduced to ensure the uniformity of examinations and criteria for assessment, comparability of results and objectivity of assessment (the coded answer sheets are marked by external examiners). The *Matura* Examination is the basis for entry into universities and colleges of tertiary education.

3.2.4. The Vocational Examination

The Vocational Examination (examination confirming vocational qualifications) is an external examination, not obligatory, conducted by Regional Examination Boards. Its task is to assess and certify vocational qualifications for a particular profession. It is conducted for graduates of: basic vocational schools, technical secondary schools, supplementary technical secondary schools and post secondary schools.

4. Polish Language as school subject (LS) at Lower-Secondary Level (*Gimnazjum*)

The following documents, procedures and standards create the basis on which the Polish language is taught as school subject:

- Core Curriculum, issued by the Ministry of Education;
- local curricula, written by local authoring teams, validated by experts appointed by the Ministry;
- textbooks, written by authoring teams and validated by the Ministry, edited in separate sets for Culture + Literature and for Language;
- Examination standards, issued by the Central Examination Commission (incl. separate versions for hearing-impaired candidates, and for [mother tongues of] national and ethnic minorities);
- External examination: Lower-Secondary School Examination (*Egzamin gimnazjalny*), conducted by Regional Examination Commissions.

4.1. Core Curriculum for the subject 'Polish Language' in Lower-Secondary Schools (1999)

The document starts with a Preamble that underlines the general educational value of mother tongue training:

> "Knowledge is acquired most of all through language. Mother-tongue teaching creates a basis for pupils' general development, contributes to the shaping of their personal identity, and thus functions as the main reference point for the whole of school education. Developing language abilities in writing and speaking (the repertoire of language terms, orthography, and the aesthetics of the written code included) is a shared responsibility of all teachers, whatever the subject taught."

The following chapter specifies the general Educational aims for this stage (the same set of aims is listed for upper-primary and for lower-secondary level):

1. Developing skills in speaking, listening, reading and writing in different private and public communicative situations – especially those of special importance for life in a democratic, civic state.
2. Raising awareness of language as part of cultural heritage.
3. Unveiling pupils' interests, capacities, needs, and linguistic skills by appropriate goal setting using suitable tasks, texts and activities.
4. Stimulating pupils' motivation to read and interpret literary works and texts of culture (including audio-visual ones) as a way to better understand the world and our place in it.
5. Presenting national and European cultural tradition.
6. Teaching to take part in and identify with culture, especially in its symbolic and axiological aspects.

Derived from these general aims, a set of specific Tasks for the school is defined (1-5 in common with the upper-primary level):

1. Developing motivation to read literary works and other texts of culture that contribute to growing knowledge of human life in the world – contemporary and past;
2. Refining communicative competence, i.e. ability to speak, listen, read and write, and to interpret a variety of texts of culture;
3. Creating situations in which language learning happens by conscious and thoughtful language use (without unnecessary burden of theoretical, abstract knowledge of the language system);
4. Supporting pupils' creativity in developing skills and accumulating knowledge;
5. Integrating cultural experiences;
6. Presenting texts of culture with a view to acquiring general knowledge, participating in contemporary culture and social communication, and to practise cultural traditions;
7. Supporting pupils in their process of building integrated systems of personal values;
8. Developing pupils' learning strategies and equipping them with tools for independent intellectual work.

The Teaching content is specified in the Core Curriculum 1999 in the following, quite technical way:

1. Formal and informal speaking situations.
2. Advanced forms of structured writing, e.g. essay, reportage, review, interview.
3. Elements of rhetoric and logic in speech: thesis, argument, conclusion, etc.
4. Syntax – inflected utterances.
5. Morphology and Flexion – their function in making meaning.
6. Basic terms of Semantics.
7. Basic terms of Stylistics.
8. Assimilation mechanisms in pronunciation – spoken versus written code.
9. Terms of literary analysis: irony, synthesis, allegory, etc.
10. Literary genres.

The expected learning outcomes – the Attainments – are presented in relation to linguistic skills, learning skills and cultural competences:

1. Speaking, listening, reading, writing.

   At the end of the *Gimnazjum* the student should be able to:

   a. produce intentionally and situationally appropriate spoken and written texts;
   b. recognise and produce utterances that inform, describe, evaluate, express opinions, convince, build an argument;
   c. use various situationally appropriate speech modes (especially the standard one – spoken and written);
   d. recognise and understand communicational principles (based on situations, emotions, judgments) and implications of language use; discuss the ethical aspect of language use;
   e. process texts (summarise, develop, adjust style, etc.);

f. transform pictures and sounds to texts (spoken and written);

g. use situationally and contextually appropriate grammatical structures producing clear and coherent texts;

h. work on sentence level – develop and condense content using simple and inflected structures, Passive voice, etc.

2. Autonomous learning

At the end of the *Gimnazjum* the student should be able to:

i. retrieve information from various sources;

j. prepare documentation, take and make notes, select and store information;

k. make wise use of mass media.

3. Cultural reception of texts

At the end of the *Gimnazjum* the student should be able to:

l. recognise and comment on specific features (e.g. means of expression) of literary and other texts of culture;

m. use literary and cultural terminology in a natural and functional way;

n. relate experiences, feelings and ambitions present in works from different periods in time, especially with reference to universal cultural traditions;

o. compare literary works with the works in other disciplines of art;

p. interpret literary works in their different contextual settings;

q. produce creative interpretations of literary works in theatrical forms (e.g. staging);

r. recognise values in literary works and relate them to the own system of values;

s. evaluate cultural phenomena independently and critically: build and defend an argument, opinion, interpretation – in writing and in speech;

t. relate literary works to personal biographies of their authors and to the historical context.

The Core Curriculum ends with a Reading list with a selection of recommended books from Polish and international literature, as well as other text types recommended for use in the classroom (newspaper articles, instructions, advertisements, movies, television programmes, internet sources, etc.).

4.2. External assessment of competences – the Lower-Secondary School Examination (*Egzamin gimnazjalny*)

The compulsory Lower-Secondary School Exam is an interdisciplinary examination consisting of two parts: Arts and Sciences. The first part tests abilities and knowledge in the Humanities: the Polish Language, History, Civic Education, Art, Music and educational

paths (see above) such as Philosophy, Regional Studies, Reading and Media, Polish Culture and the Mediterranean Civilisation.

The second part tests abilities and knowledge in science subjects: Maths, Biology, Geography, Chemistry, Physics and Astronomy and educational paths such as Health Education or Ecology.

In each part of the examination the maximum score is 50 points. The result is not significant for the graduation from lower-secondary school but it is noted on the certificate and taken seriously into account by upper-secondary schools during the admission process.

From 2008 on, it is planned to implement a third part of the examination: assessing competences in a modern foreign language.

4.2.1. Examination standards (requirements) and their application in the tests

The following seven requirements have been specified to assess the Reading skill:

1. Reading and interpreting 'Texts of Culture' (ToC) on their literal, metaphoric, and symbolic levels.
2. Identifying the author's intention, facts and opinions, historical truth and fiction, persuasiveness, manipulation, valorisation.
3. Retrieving information from different types of ToC, such as literary, popular scientific and normative (legal) texts, newspaper articles, illustrations, tables, graphs, maps, figures, schemata.
4. Identifying types and functions of linguistic means for making meaning in literary texts, newspaper articles, pieces of art and music.
5. Identifying and interpreting logical relations in the domains of politics, economy, culture, and social life, civilisation development in Poland and in the world.
6. Perception and analysis of contextual aspects necessary for appropriate interpretation of ToC, such as historical, biographical, philosophical, religious, literary, artistic, musical, and regional; ability to explain relationships between different types of ToC (arts, music, and literature).
7. Perception of values embedded in ToC.

The application of the requirements in real tests conducted over the last three years is shown in Table 1:

Requirement number – Reading	Number of test items related to the requirement		
	2005	2006	2007
1	8	7	8
2	2	7	2
3	5	4	6
4	3	2	3
5	2	2	1
6	5	3	5
7	0	0	0

Table 1. Application of requirements in tests. Reading.

The analysis shows that most test items used in the years 2005-2007 were related to Requirement No. 1: *Reading and interpreting 'Texts of Culture' (ToC) on their literal, metaphoric, and symbolic levels*. The next most frequently used were standards No. 3 (*Retrieving information from different types of ToC, such as literary, popular scientific, and normative [legal] texts, newspaper articles, illustrations, tables, graphs, maps, figures, schemata*) and 6 (*Perception and analysis of contextual aspects necessary for appropriate interpretation of ToC, such as historical, biographical, philosophical, religious, literary, artistic, musical, regional; ability to explain relationships between different types of ToC [arts, music, literature]*). These three requirements generated together between 56% (2006) and 76% (2007) of the score points available in this part of the examination. At the opposite end is Requirement No. 7 (*Perception of values embedded in ToC*) not used at all in these years – in fact it generated only one single test item (in 2004) in the six years for which the system has been in use.

The assessment of the Writing skill has been standardised by a set of nine requirements:

1. Producing linguistically and stylistically correct texts, such as description, narrative, report, review, essay, note, plan, newspaper article, interview, announcement, advertisement, invitation, dedication, request, letter, memoirs.
2. Using terms and categories specific to the domain of humanistic subjects and disciplines.
3. Producing situationally appropriate texts of informational or persuasive character.
4. Knowledge and use of organisational principles to produce logically and syntactically coherent texts appropriate to the topic given.
5. Putting forward, organising and valorising arguments in favour of one's own or somebody else's opinion.
6. Analysing, comparing, organising, and synthesising information embedded in ToC.
7. Processing texts: summarising, developing, transforming (stylistically).
8. Expressing own opinion regarding relationships between native culture and other cultures; commenting on historical and contemporary links between the Polish and the Mediterranean cultures in such domains as politics, economy, and everyday life.
9. Tackling problematic issues present in ToC, suggesting solutions and drawing conclusions.

The operationalisation of these standards in real tests is shown in Table 2:

Requirement number – Writing	Number of test items related to the requirement		
	2005	2006	2007
1	8	9	8
2	1	0	1
3	7	5	7
4	5	6	5
5	3	2	1
6	1	1	2
7	0	2	1
8	0	0	0
9	0	0	0

Table 2. Application of requirements in tests. Writing.

We can see that it is Requirement No. 1 (*Producing linguistically and stylistically correct texts, such as description, narrative, report, review, essay, note, plan, newspaper article, interview, announcement, advertisement, invitation, dedication, request, letter, memoirs*) that has been used most frequently, followed by No. 3 (*Producing situationally appropriate texts of informational or persuasive character*) and 4 (*Knowledge and use of organisational principles to produce logically and syntactically coherent texts appropriate to the topic given*). It is striking that no use at all has been made of Requirements Nos. 8 and 9 (*Expressing own opinion regarding relationships between native culture and other cultures; commenting on historical and contemporary links between the Polish and the Mediterranean cultures in such domains as politics, economy, everyday life* and *Tackling problematic issues present in ToC, suggesting solutions and drawing conclusions*).

In the part of the examination that assesses competences in Sciences, the following four areas have been identified to serve for setting requirements:

1. Using terms and applying procedures
2. Retrieving and using information
3. Identifying and describing facts and relationships (logical, functional, spatial, and temporal)
4. Applying integrated knowledge and skills to problem solving

It is interesting to note that most of these requirements (1, 2, and 3) are clearly related to the use of language across curriculum (LAC) – at least to a similar extent as they are related to the subject matter. What is tested here is thus as much language competence as subject-matter competence.

The proportion of test items produced with reference to these requirement areas is being kept stable in all tests:

Requirement area number – Sciences	Number of test items related to the requirement (2005, 2006, 2007)
1	15 (30%)
2	12 (24%)
3	15 (30%)
4	8 (16%)

Table 3. Application of requirement areas in tests. Sciences.

4.2.2. The results (2006)

The following two tables – based on the results of the 2006 test – show representative tendencies for the whole set of exams (results of tests conducted in other years are quite similar).

	Arts		Sciences			
	Reading	Writing	Using terms and applying procedures	Retrieving and using information	Identifying and describing facts and relationships	Applying integrated knowledge and skills to problem solving
Average facility of test items	.72	.54	.44	.67	.46	.30
Average for each part	.63		.48			

Table 4. Average facility of test items (2006)

The average test item facility index indicates that testing competences in the Sciences is based on items that are significantly more difficult for the candidates than those assessing the reading and writing skills tested under the Arts. Test items assessing the reading skill (most of them closed multiple-choice items) prove to be less difficult to the candidates than those evaluating writing ability, and far less difficult than those testing the ability to apply integrated knowledge and skills to problem solving (in both cases mostly open-ended tasks).

General findings: Arts (Reading and Writing, 2006)	
Abilities confirmed	• Retrieving information • Identifying the author's intention • Organising text • Producing coherent, communicatively appropriate text
Deficits identified	• Reading instructions to test items • Ignoring facts clearly present in texts • Building an argument • Accuracy, orthography, punctuation • Lack of historical context knowledge • Using terms related to literary and language sciences

Table 5. General findings: Arts (Reading and Writing, 2006)

In general, the results of the test parts assessing the competences in Arts indicate that while the requirements related to the communicative use of language are being largely met by the graduates of the lower-secondary school in Poland, the standards, by which linguistic accuracy, the use of (scientific) terms and the application of historical knowledge are assessed, are much less satisfactorily fulfilled.

Conclusions

The aim with this presentation of a concrete case of compulsory external examination scheme where competences in a language of schooling are assessed was to illustrate some of the issues relevant for the development of an LE framework. The following three general points confirming the initial observations noted in the introduction to this set of studies may serve as a conclusion here:

1. Only those attainment targets (competence standards) that may be turned into observable behaviour can become real subjects for assessment procedures (testing tasks/items);
2. Item level testing may not be capable/suitable of handling all of the complexity of LE competences;
3. Most (if not all) competences that are subject to assessment in education are language related (LAC + LS + FL).