HANNA KOMOROWSKA
University of Social Sciences and Humanities, Warsaw
hannakomo@data.pl

FORMATIVE EVALUATION – ITS PAST, PRESENT AND FUTURE

Abstract

The article examines psychological and pedagogical contexts of the birth and the development of formative evaluation in second and foreign language teaching and learning. Research on its educational value, meta-analyses and case studies conducted in a number of school systems under the auspices of the Organisation for Economic Co-operation and Development (OECD) are presented with examples of ways in which research informed the educational policy of various countries and impacted on school evaluation strategies. Attention is given to challenges and controversies faced by schools in the process of implementing formative evaluation and/or integrating it with summative approaches. Implications are also sought for pre- and in-service teacher education.

1. Introduction

Until the end of the 1950s teaching was considered the sole factor of a student's success as dictated by behavioural assumptions according to which learning equals the development of habits, the process is based on mechanical repetition and thus takes place in a similar way in every individual. With overlearning as the only aim, the Audiolingual Method guaranteeing automatization gained the status of the ultimate cause of students' achievement. In consequence any form of assessment other than final examinations was considered redundant. With imitation as the main learning technique, error was not expected to appear, therefore correction had no place in the teaching process nor had any form of feedback (Van Patten, Williams 2015).

A considerable change of perspective took place with the promotion of cognitive approaches in second and foreign language teaching. As varying degrees of success

observed within the same didactic and organizational framework were given due consideration, differences started to be attributed to learning rather than to teaching processes. The rapid growth of cognitive methodology in the 1960s and the birth of the Communicative Approach in language education in the early 1970s led to increased interest in ways of creating environments conducive to effective language acquisition (de Bot 2015).

Support came from educational sciences. The first sign of interest in the stages of the learning process and their implications for curricula and teaching methods appeared with the work of the team led by Benjamin Bloom and the publication of their seminal *Taxonomy of educational objectives. The classification of educational goals,* the first volume of which presenting the cognitive domain was published in 1956 and the second volume – devoted to the affective realm – eight years later (Bloom et al. 1956, 1964).

This attention to pedagogical aims, a topic of utmost importance for the curriculum theory, brought immediate results in the field of syllabus and materials design and called for methods of their evaluation (Kunnan 2015). Until then curricula and syllabi had been designed according to postulates formulated by Robert Tyler and Hilda Taba who believed the learning outcomes to be the sole measure of quality of the documents produced (Tyler 1949; Taba 1966). This approach resulted in an unprecedented interest in summative evaluation of educational achievement. Curriculum construction, therefore, did not involve any form of built-in evaluation; the value of the project was assessed at the end of the course at which a given curriculum was implemented. Michael Scriven was the first researcher to introduce innovation in this field. Instead of starting evaluation procedures related to the educational value of curricula and/or coursebooks after the actual course ended, he postulated ongoing evaluation, labelling the procedure with a newly coined term – *formative evaluation* (Scriven 1967).

The present text will examine the history of research on formative evaluation in language teaching, forms this type of assessment can take in language education, and difficulties encountered in its implementation, as well as implications of the data obtained for pre- and in-service teacher education.

2. The birth and early development of formative evaluation in second and foreign language teaching – lessons from the past

With the advent of the Communicative Approach to second and foreign language teaching it soon became obvious that, with the changing perspective on student learning, language teachers needed to engage in individualized instruction and with the changing approach to error now considered a learning step, instructors needed new ways to enhance teacher-student interaction. In general education this process started a few years earlier. As early as the 1960s Bloom postulated a frequent use of feedback and correctives in the classroom (Bloom 1968), therefore in communicative language teaching not only were various ways of preventing the occurrence of

mistakes at language lessons promoted, but also a whole spectrum of error correction techniques.

These developments rendered the final, end-of-course or end-of-term evaluation of educational achievement insufficient. Scriven's concept of formative evaluation in curriculum development was then transferred to the sphere of learning. Formative and summative evaluation of educational attainment were first introduced as equally valuable assessment formats in Bloom's *Handbook of formative and summative evaluation of student learning,* a publication that would remain a testers' bible for many years to come (Bloom et al. 1971). Since that moment most educators have come to believe that formative evaluation is criterion- rather than norm-referenced, therefore in many areas, language education included, it has been geared toward specific achievement rather than general linguistic proficiency.

Although the 1970s opened the path for innovations such as evaluating learners' progress and not only the outcome, a real breakthrough came in the 1980s as a result of further changes in curriculum theory and practice.

As identification and clarification of curricular aims became more specific, more precision was also needed in assessing degrees of achievement of particular curricular goals. What is more, dynamics of the very process of achieving those goals gained researchers' attention, becoming one of the most important criteria of course evaluation (Taylor, Richards 1985; Ornstein, Hunkins 1988). To successfully assess the value of curricula and materials in a new way, not only summative, but also formative assessment of the learners' progress proved indispensable. A new model of curriculum evaluation was promoted, stressing the need to take learners', parents' and teachers' perspectives into consideration (Sevigny 1981; Bogdan, Biklen 1982). This so-called *triangulation* built on the largely ignored and undeservedly forgotten evaluation models proposed more than a decade earlier – one presented by Stake (1967) and another by Stufflebeam (1969), the latter popularly referred to as the Context-Input-Process-Product (CIPP) Model.

Multiple perspectives contributed to the call for a closer look at what was taking place in the classroom. The increasing role of recording the learning process and reacting to what was being observed added significance to the then innovative concept of formative evaluation. Since that time the term *evaluation* has been used much more often with reference to curricula and materials, while the term *assessment* has become associated with classroom learning.

Interest in evaluation in general and in formative evaluation in particular grew throughout the the 1970s, 1980s and 1990s. Important journals emerged in this new field of research such as *Assessment in Education*, *Mésure et Evaluation en Éducation* or *Unterrichtswissenschaft.* Many reputable general educational journals started publishing articles and studies on the subject. Important books in the field were published from the 1970s (Cardinet 1977; Allal et al. 1979) throughout the 1980s (Ingenkamp 1985; Cardinet 1986; de Ketele 1986) up to the end of the 20th century (Weiss 1991; Gregoire 1996; Bonniol, Vial 1997; Hadji 1997; Perrenoud 1998; Herrmann, Höfer 1999). Summative evaluation in language education moved through the psychometric-structuralist period linked to late audiolingual and early cognitive

approaches and the psycholinguistic-sociolinguistic one linked to the integrative trends of the Communicative Approach. With the promotion of the Communicative Approach summative and formative types of evaluation were found to fulfil complementary functions in education.

A key theoretical development helping to dispel the idea of the two evaluation types being mutually exclusive was the introduction of the concept of teacher regulation and the classification of regulation activities. *Proactive regulation* was defined as an activity during which goals are differentiated depending on the needs and predispositions of the learners, *interactive regulation* as occurring throughout the learning process, and *retroactive regulation* as an act consisting in feedback and correction (Allal 1993). At this stage, however, formative evaluation was still primarily viewed as a different way of securing useful and supportive assessment than as a way of raising the quality of learning to improve both its process and its product.

Results of numerous research projects and implementation studies conducted in the first three decades of formative evaluation in the French and English educational practice were discussed in detailed reviews of literature (Fuchs, Fuchs 1986; Laveault 1992; Kluger, DeNisi 1996; Black, Wiliam 1998; Louis 1999).

3. Formative evaluation today. How does research inform the educational policy?

In the 21st century formative evaluation has gained more and more attention and status, starting with the field of curriculum studies (Figari, Achouche 2001; Pinar 2004; Tedesco et al. 2014; Alismail, McGuire 2015). Growing in importance in the contemporary pedagogical theory, the new approach was soon deemed valuable in areas far beyond syllabus and materials design and became translated into the day-to-day assessment practice of state and public schools (Heritage 2013). Rather than a mere assessment technique, formative evaluation has come to be viewed as a way of collecting important data about learners and their learning – a powerful tool to enhance classroom interaction and improve educational achievement by means of promoting learning strategies and developing *learning to learn* skills. Wiliam declared it to function as "the bridge between teaching and learning" (Wiliam 2010).

Multiple research projects in the field have been presented and evaluated over the past two decades. The first impressive meta-analysis was prepared in the last few years of the 20th century by Black and Wiliam (1998) who analyzed 250 studies from the field of formative assessment.

The 21st century brought new data as in 2008 Hattie examined over 800 publications on the subject (Hattie 2008). Although not all research data were conclusive (Rodriguez 2004), OECD's Center for Educational Research and Innovation (CERI), basing on numerous positive reports, commissioned reviews of the most important English, French and German publications on formative evaluation. In English-speaking countries research concentrated mainly on the evaluation of educational achievement in classes working within the formative framework, i.e. researchers used quantitative methods to assess final learning outcomes. Although studies under

examination ranged from pre-schools to colleges, learning gains were observed in all age groups. In French- and German-speaking countries research tended to be more qualitatively oriented and looked into ways in which formative assessment was being implemented in classrooms in the day-to-day work of teachers and learners. Attention moved to specific techniques used by teachers applying formative evaluation in their work as well as to types of activities that offered the most benefits to students (Allal, Lopez 2015; Black, Wiliam 2015; Köller 2015).

Teacher-made action research projects also brought very promising conclusions, especially in showing how the introduction of formative assessment encouraged teachers who had previously concentrated on performance objectives and 'teaching to the test' to considerably change their classroom strategies by targeting more specific learning goals and attempting to successfully create learning opportunities which would help to achieve set aims (Black 2010). Refocusing of this kind was in line with the guidelines of the European educational policy stressing metacognition and control strategies, postulating self-awareness, flexibility, individual learner's growth and peer-co-operation together with learning to learn competencies instead of the traditional emphasis on competition and ranking (*CEFR* 2001).

The most valuable data obtained in research projects across the world were those which nuanced and contextualized the impact of formative assessment on student learning. It was noticed, for example, that teachers were more successful following the new work format when they had the know-how to skilfully moderate classroom discourse; otherwise the results were less impressive. "Wait time", i.e. time the teacher waits for the learner's answer, was also noticed to be crucial according to basic requirements of the "pedagogy of time" as otherwise the teacher has no chance to elicit any production whatsoever, cannot become aware of the learner's potential or assess students' performance.

Formative evaluation proved most useful if feedback was offered immediately. What is more, not every kind of feedback was actually productive; the so-called ego-involving feedback addressed at the personality of the learner tended to demotivate, while performance-oriented feedback, especially carrying constructive comments on ways to improve performance, raised motivation and enhanced the learning process. The quality of feedback, types of explanations, their clarity and purposefulness were also important as feedback devoid of useful information resembles the traditional activity of simply evaluating the learner's production and labelling it with a grade (Black 2010). Eliciting feedback from learners on the teacher's success in helping the students achieve learning goals set for a particular lesson also proved very productive, e.g. by means of a traffic light technique when at the end of each lesson each student was expected to put up one of three cards – a red one signalling lack of skill that was being developed at that particular lesson, a yellow one signalling that the skill was developed only partially and a green one signalling the sense of achievement. Enhancing students' self-assessment skills proved effective when teachers decided to follow three steps – ask learners to determine the grade they expect, compare the grade received from the teacher with their own expectations and check their own work one more time against the grading criteria presented earlier.

Research also demonstrated that to make formative evaluation work, teachers need sufficient knowledge of learning theories, sufficient competence to employ scaffolding on the basis of the data obtained through formative feedback (Shepard 2005) and – last but not least – sufficient skills to engage their students in the learning process (Sterna 2016).

Effectiveness of formative assessment largely depends on the choice of learning tasks. The whole endeavour is likely to be successful if tasks selected for students are authentic and contextualized, hence the revival of Wiggins' (1989) concept of *authentic assessment* as only these kinds of language activities are likely to guarantee positive washback effect and motivational gains.

Reports demonstrate that students showed learning gains when they were well informed about final learning aims as well as about specific goals of a given lesson (Sadler 1998; Black, Wiliam 2015). In language education it usually means that functional objectives formulated in CAN DO statements, referring to what learners will be able to get done using a new language, are more clear and appealing to learners than those formulated in terminological, metalinguistic categories. Learners were observed to pay more attention during the lesson and engage in more interaction with the teacher since in the absence of grades they need more feedback. They also seemed to benefit more when they had a chance to talk their progress over with their classmates and when self- or peer assessment was based on clear guidelines offered by the teacher. Peer-assessment was particularly effective when combined with self-assessment, i.e. when the learners had to decide whether their own work would be of a higher, equal or lower quality than the one they were assessing. They also benefited from the wait time offered by the teacher, especially when an oral response was to be given. In out-of-class work projects learners were seen to work particularly well when they were explicitly encouraged to collaborate and helped to plan their work precisely, reflecting on their roles, duties and deadlines. Low-achievers were observed to benefit from the introduction of formative assessment because more frequent and more practically oriented feedback counterbalanced their weaker learning to learn and/or metacognitive skills, while high-achievers also benefited – because the new format made it possible for them to gain more opportunities and develop learner autonomy. In most cases formative assessment positively affected learners' motivation: they asked more questions, demonstrated fewer behavioural problems, felt safer in the classroom, took more risks, shared more ideas and engaged more frequently in classroom interaction (Black et al. 2003).

Case studies show that schools tended not to restrict formative evaluation to experimental classroom or subject areas treated in a specific way, but rather to introduce "whole-school policies". More individualized feedback on the student's progress (Nevo 2002) also increased schools' contact with parents. To make formative evaluation work for all the actors of the educational institution who would then enjoy all the benefits described above, a certain pre-condition had to be fulfilled on the school level. What proved indispensable was acknowledging the fact that students have individual abilities, communicative needs and ways of working. Teachers also found it necessary to convince students of the fact that results depend on their effort

rather than talent or its lack; in other words – of the value of causal attribution and internal locus of control.

The 21st century has seen a rapid growth of interest in formative evaluation on the part of policy makers. More and more frequently this form of assessment has been officially approved, while recommendations have also been made to balance it with the forms of summative evaluation by now well-rooted in schools of all types and levels. Apart from important curriculum reforms in multiple countries (Komorowska 2018) in the last two decades activities promoting formative evaluation have taken place in the United States, Great Britain, Australia, Finland, Denmark, Italy and Canada, though it should be noted that educational systems differ in the choice of aspects considered crucial – Finland concentrated on self-assessment, Denmark on feedback, Australia on implementation of feedback in core academic areas, Scotland on formative feedback in assessing young learners (ages 5–14), Italy on integrating formative feedback and individualization and New Zealand on professional development of teachers in the area of evaluation (DfES 2002; Herman et al. 2010; Looney 2011; Herman, Ripley 2012; Earl 2013; OECD 2015).

As the Organisation for Economic Co-operation and Development advises, educators should not forget that huge differences in process and product occur within schools as well as between them. Thus evaluation based on interschool comparisons tends to be discouraged, while due consideration to classroom practice is recommended. In consequence governmental guidelines in many countries encourage schools to self-assess, a trend noticed, for example, in Denmark, Finland and Scotland (OECD 2015).

As can be seen from the above, formative evaluation seems to be not only well anchored in the general perception of learners, their parents and teachers, but also strongly promoted by the educational administration in Europe and beyond. The introduction and functioning of formative assessment implemented on a large scale is, however, not devoid of difficulties and misunderstandings (Laveault, Allal 2016).

4. Challenges, controversies and the future of formative evaluation in language teaching and learning

Challenges and controversies in the field of formative evaluation are numerous. Due to space limitations in the present text only the most important of these issues will be discussed.

Early misunderstandings related to formative evaluation's nature, purpose and function had their roots in the **terminology** used and meanings ascribed to words. Initially the term *formative evaluation* was used to denote more frequent and more individualised forms of assessing learning outcomes. In this sense assessment, though labelled formative, took the form of mini-summative instances of evaluation addressed to particular students and was still meant to be *assessment of learning*. Later the term assumed a broader meaning and began to denote enhancing student learning, developing learning to learn skills and intensifying teacher-student

interaction in the classroom through feedback. In this sense it became *assessment for learning*. In Great Britain the Department of Education and Skills (DfES) decided to promote the definition of assessment for learning adopted by the Assessment Reform Group (ARG): "Assessment for learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where they are in their learning, where they need to go and how best to get there" (DfES 2002). This definition became very popular among teachers of various subject areas. Soon other terms became popular as well, such as *assessment as learning* (Dann 2002) and *dynamic assessment* (Scallon 2000; Poehner, Lantolf 2005). Most of these terms encompass making students aware of goals and deadlines, but also helping them to monitor and regulate their learning processes. These two ways of using the terms refer in fact to two very different sets of educational phenomena. Unless teachers become aware of this distinction, misunderstandings are likely to continue.

Another difficulty is connected with identifying and clarifying educational objectives. In the era of promoting learner autonomy, it is firmly believed that **participative goal-setting** is the best curricular solution in education. Research demonstrates that setting goals by an individual enhances his/her performance more than having goals imposed by an external source or power (Locke 2000; Wegge 2001). The same is true for groups of students for whom the process of negotiating evolving objectives has been demonstrated to be more conducive to raising the quality of their learning. Yet this is true only for easy-to-attain goals or objectives of moderate difficulty. When a goal is difficult to achieve, the situation changes. It is then better for the teacher to impose the goal, informing students at the very beginning about what is expected of them. Haslam, for example, states that "if they [goals] are participatively set, there would appear to be value in approximating to them gradually, but if they are imposed then this strategy may prove counterproductive and it may be better to set the difficult goal at the outset" (Haslam et al. 2009: 441). Yet extreme ends of the scale are more often selected as teachers in state schools tend to impose goals on their students irrespective of the difficulty level, while those working for language schools often run into classroom management problems caused by excessive negotiating time.

Feedback versus grading is another problematic issue. From the early days of the implementation of formative evaluation teachers have complained that this type of assessment clashed with the **grading system** as students do not concentrate on comments and ignore information on what they have done well and what needs improvement if at the same time they are being graded (Black, Wiliam 2015). Few schools are ready to drop grades, although more and more often principals decide to introduce parallel formats of descriptive assessment. As middle and secondary schools usually expect teachers to grade their students, the way to overcome this difficulty lies in drawing up a classroom contract which sets clear assessment guidelines with the teacher's feedback on the student's performance coming first and grading being done later. Sharing grading criteria with students also sparks controversy. Some teachers claim that – although they see value in sharing assessment criteria before a class test or an examination – doing so at the beginning of a sequence of lessons might stifle students'

creativity and thus prove counterproductive (OECD 2015). The question of timing is crucial here and so is the teaching objective; accuracy issues lend themselves easily to early criteria sharing, while discussions related to tasks already completed might prove more beneficial for those connected with fluency, interactive, mediating and intercultural skills.

Even though promotion of language learning on a mass scale has been successful, quality of education seems to be an issue across school systems, and as such is a problem addressed by the most important European institutions: the Council of Europe, the European Union, the Organisation for Economic Co-operation and Development as well as the European Centre for Modern Languages. As quality is most often assessed on the basis of **examinations and testing**, summative rather than formative evaluation finds itself at the centre of attention of educational administration. Learners and parents focus on test scores and expect teachers to offer instruction that guarantees desired results in high-stakes examinations. Due to institutional ranking of schools, comparison and competition gain more and more importance in educational institutions overshadowing individual learning paths and learners' growth. Formative evaluation – even if actually implemented in the language classrooms – concentrates on helping students to become more test-wise by developing test-taking strategies and in consequence to achieve results which will look good on their certification documents. In this situation formative assessment becomes a humble servant of summative evaluation.

In its original function formative evaluation is likely to be restricted to primary education levels where typically high-stakes examinations are not yet administered and to those language schools that do not offer exam preparation courses. In other words, it runs the risk of being treated seriously only in those few classrooms of the so-called progressive schools where improvement is understood as quality learning rather than just a high score on a final test. To remedy the situation, language teaching needs to be allowed more classroom time and smaller class size. Realistically speaking, not many state schools can be expected to secure sufficient time and budget. To overcome this hurdle, another solution is sometimes recommended, i.e. making formative use of summative tests. The solution is reasonable, especially because each of the two types of evaluation has its advantages, though they differ in function and purpose (Shohamy 2001; Kucharczyk 2015). Several possibilities are offered here. Before a test, students (a) can be asked to prepare a list of topics, skills or vocabulary groups they expect to appear on the test and to decide which items have been fully acquired, which call for more practice and which have not been mastered at all, or (b) can be asked to prepare their own tasks or test items of the type they expect to appear on the test. After the test – if they have the opportunity to get their copies back – they can be asked to peer-assess their work (Black, Wiliam 2015).

Peer-evaluation, which is often seen as a remedy or at least a way to reduce the role of high- stakes examinations, also creates problems. Controversy over its educational value springs from the fact that some students reveal a tendency to criticize and diminish their peers, most probably acting out some former interpersonal or intrapersonal conflicts. For that reason very precise assessment rubrics

need to be prepared and a certain amount of coaching offered before systematic peer-assessment activities are implemented. Effective preparation and practice activity consists in comparing the assignment analyzed with one's own already submitted work or with the idea of one's own potential assignment of the same kind. Problems also arise when students, functioning as even relatively objective assessors, try to grade the work of a classmate – a solution strongly advised against by most of the teachers.

As can be seen from the above, formative evaluation – in spite of its empirically confirmed values – still involves issues which make this type of assessment difficult to implement. No wonder teachers place solving the set of problems related to evaluation among their most important professional needs (Czetwertyńska 2015).

5. Conclusions. Implications for teacher education

Formative evaluation has so far been successfully implemented in initial and primary education where most of the educational systems have already replaced grading with descriptive assessment. In secondary education, although more and more innovative schools tend to introduce formative evaluation in parallel with the summative one, it is far more often considered a way to stimulate and support student learning. In light of the increasing role of high- stakes examinations and the practical requirements of certification in mass education, formative evaluation does not seem likely to be much further promoted in most of the state educational systems. Smaller and/or private schools can afford more flexibility in this field. Prospective teachers being trained to use various forms of assessment need to be able to reflect on role expectations and possibilities of role enactment in the context of their future employment.

Formative evaluation – even with all the benefits discussed above, such as contribution to the development of metacognition, motivation, learning to learn skills and self-assessment as well as to the promotion of alternative non-test assessment techniques such as portfolios and logs – is not devoid of educational dangers. One risk, inherent in *assessment for learning*, is constant observation of learners whose sense of being at the centre of the teacher's attention, however gratifying it might be at the outset, often leads to loss of privacy and subsequent avoidance behaviour. Another, inherent in *assessment of learning*, lies in the frequency of assessment activities which might bring about disadvantages typical of those observed in gamification, i.e. loss of interest in long-term objectives, development of instrumental motivation and competitiveness. Self-assessment is also often considered difficult for the learners, although they tend to have fewer doubts when it comes to peer-assessment. Pre-service teacher education should be responsible for awareness raising *vis-à-vis* those issues, while in-service teacher development needs to help in the selection of prevention strategies appropriate in a given context and techniques of dealing with problems which have already arisen.

Schools, students and teachers cannot make good use of formative evaluation unless certain indispensable measures are taken on higher administrative levels in

ways permitted by the national educational policy. When administration is open to innovative solutions, teachers feel free to try out new methods and techniques even though new approaches do not immediately bring about improvement in learners' performance. When administration provides opportunities for in-service teacher development, schools can base their practice not only on friendly exchanges in the staff room and reflective peer coaching (Vidmar 2005), but also on research data provided by professionals and the know-how of teacher trainers. When budget is provided for workshops and conferences, schools across the region can share experience and work out new solutions exchanging and circulating new knowledge and skills. But as long as schools are evaluated and judged solely on the results of summative interschool tests, classroom-based formative assessment will be treated as no more than an extra burden for language teachers and an impractical, unrealistic scheme designed by researchers far removed from the educational practice.

References

Alismail H.A., McGuire P. 2015. 21st century standards and curriculum: Current research and practice. – *Journal of Education and Practice* 6.6: 150–154.

Allal L. 1993. Régulations métacognitives. – Allal L., Bain D., Perrenoud P. (eds.). *L'évaluation formative et didactique du français*. Neuchâtel: 81–98.

Allal L., Cardinet J., Perrenoud P. (eds.). 1979. *L'évaluation formative dans un enseignement differencié*. Bern.

Allal L., Lopez L.M. 2015. Formative assessment of learning. A review of publications in French. – Black P., William D. (eds.). *Office of Economic Co-operation and Development: Formative assessment. Improving learning in secondary classrooms*. Paris: 241–264.

Black P. 2010. *Assessment of and for learning. Improving the quality and achieving a positive interaction*. Brussels.

Black P., Harrison C., Lee C., Marshall B., Wiliam D. 2003. *Assessment for learning. Putting it into practice*. Berkshire.

Black P., Wiliam D. 1998. Assessment and classroom learning. – *Assessment in Education* 5: 7–71.

Black P., Wiliam D. 2015. Changing teaching through formative assessment. Research and practice. – Office of Economic Co-operation and Development: *Formative assessment. Improving learning in secondary classrooms*. Paris: 223–240.

Bloom B.S. 1968. Learning for mastery. – *Evaluation Comment* 1.2: 1–8.

Bloom B.S., Englehart M.D., Furst H.J., Hill W.H., Krathwohl D.R. 1956. *Taxonomy of educational objectives. The classification of educational goals. Handbook I. Cognitive domain*. New York.

Bloom B.S., Englehart M.D., Furst H.J., Hill W.H., Krathwohl D.R. 1964. *Taxonomy of educational objectives. The classification of educational goals. Handbook II. Affective domain*. New York.

Bloom B.S., Hasting J.T., Madaus G.F. 1971. *Handbook on formative and summative evaluation of student learning*. New York.

Bogdan R.C., Biklen S.K. 1982. *Qualitative research for education. Introduction to theory and methods*. Boston.

Bonniol J.J., Vial M. 1997. *Les modèles de l'évaluation*. Brussels.

Cardinet J. 1977. *Objectifs éducatifs et evaluation individualisée*. Neuchâtel.

Cardinet J. 1986. *Évaluation scolaire et pratique*. Brussels.

CEFR 2001 = *Common European Framework of Reference for Languages. Teaching – Learning – Assessment*. 2001. Cambridge.

Czetwertyńska G. 2015. Ocenianie jako nauczanie i uczenie się. – *Języki Obce w Szkole* 3: 4–9.

Dann R. 2002. *Promoting assessment as learning. Improving the learning process*. London.

De Bot K. 2015. *A history of applied linguistics*. London.

De Ketele J.M. (ed.). 1986. *L'évaluation: approche descriptive ou prescriptive?* Brussels.

DfES 2002 = Department for Education and Skills. 2002. *Supporting Pupils Learning English as an Additional Language*. Revised edition. Annesley (Notts).

Earl L. 2013. *Assessment as learning. Using classroom assessment to maximize student learning*. Thousand Oaks.

Figari G., Achouche M. 2001. *L'activité évaluative réinterrogée*. Brussels.

Fuchs L.S., Fuchs D. 1986. Effects of systematic formative evaluation: A meta-analysis. – *Exceptional Children* 53: 199–208.

Gregoire J. 1996. *Évaluer les apprentissages. Les apports dela psychologie cognitive*. Brussels.

Hadji C. 1997. *Évaluation démystifiée*. Paris.

Haslam S.A., Wegge J., Postmes T. 2009. Are we on a learning curve or a treadmill? The benefits of participative group goal setting become apparent as tasks become increasingly challenging over time. – *The European Journal of Social Psychology* 39: 430–446.

Hattie J. 2008. *Visible learning. A synthesis of over 800 meta-analysis relating to achievement*. London.

Heritage M. 2013. *Formative assessment in practice. A process of inquiry and action*. Cambridge.

Herman J., Osmundson E., Silver D. 2010. *Capturing quality in formative assessment practice. Measurement challenges*. Los Angeles.

Herman J., Ripley M. 2012. *Assessment and teaching of the 21st century*. New York.

Herrmann J., Höfer C. 1999. *Evaluation in der Schule – Unterrichtsevaluation*. Gütersloh.

Ingenkamp K. 1985. *Lehrbuch der Pädagogischen Diagnostik*. Weinheim.

Kluger A.N., DeNisi A. 1996. The effects of feedback interventions on performance. A historial review, a meta-analysis and a preliminary feedback intervention theory. – *Psychological Bulletin* 119: 254–284.

Köller O. 2015. Formative assessment in classrooms. A review of the empirical German literature. – Black P., William D. (eds.). *Office of Economic Co-operation and Development: Formative assessment. Improving learning in secondary classrooms*. Paris: 265–279.

Komorowska H. 2018. Feedback in language learning and teaching. – *Glottodidactica* 45.2: 185–199.

Kucharczyk R. 2015. Jak definiować poprawność językową w epoce wielojęzyczności? Kryteria oceny wypowiedzi ustnej na egzaminie z języka obcego nowożytnego (na przykładzie egzaminów maturalnych). – Sujecka-Zając J. (ed.). *Ewaluacja biegłości językowej*. Warszawa: 119–132.

Kunnan J. (ed.). 2015. *Talking about language assessment*. London.

Laveault D. (ed.). 1992. *Les pratiques d'évaluation en éducation*. Montreal.

Laveault D., Allal L. 2016. *Assessment for learning. Meeting the challenge of implementation*. Hamilton.

Locke E.A. 2000. Motivation, cognition and action. An analysis of studies on task goals and knowledge. – *Applied Psychology. An International Review* 49: 408–429.

Looney J.W. 2011. *Integrating formative and summative evaluation*. Paris.

Louis R. 1999. *L'évaluation des apprentissages en classe: théorie et pratique*. Laval.

Nevo D. (ed.). 2002. *School-based evaluation. An international perspective*. Oxford.

OECD 2015 = Office of Economic Co-operation and Development: *Formative assessment. Improving learning in secondary classrooms*. 2015. Paris.

Ornstein A.C., Hunkins F.P. 1988. *Curriculum. Foundations, principles and issues.* Boston.

Perrenoud P. 1998. From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. – *Assessment in Education* 5.1: 85–102.

Pinar W.F. 2004. *Understanding curriculum*. New York.

Poehner M.E., Lantolf J.P. 2005. Dynamic assessment in the language classroom. – *Language Teaching Research* 9.3: 233–265.

Rodriguez M.C. 2004. The role of classroom assessment in student performance on TIMSS. – *Applied Measurement in Education* 17: 1–24.

Sadler R. 1998. Formative assessment. Revisiting the territory. – *Assessment in Education* 5: 77–84.

Scallon G. 2000. *L'évaluation formative.* Bruxelles.

Scriven M. 1967. The methodology of evaluation. – *AERA Monograph Series on Evaluation* 1: 39–83.

Sevigny M. 1981. Triangulated inquiry. A methodology for the analysis of classroom interaction. – Green J., Wallat C. (eds.). *Ethnography and language in educational settings*. Norwood: 65–85.

Shepard L.A. 2005. Linking formative assessment to scaffolding. – *Educational Leadership* 63.3: 66–70.

Shohamy E. 2001. *The power of tests. A critical perspective on the uses of language tests*. Singapore.

Stake R.E. 1967. The countenance of educational evaluation. – *Teachers College Record* 68.7: 523–540.

Sterna D. 2016. *Uczyć się uczyć. Ocenianie kształtujące w praktyce*. Warszawa.

Stufflebeam D.L. 1969. Evaluation as enlightenment for decision making. – Beatty W.H. (ed.). *Improving educational assessment and an inventory of measures of affective behaviour.* Washington, DC.

Taba H. 1966. *Teaching strategies and cognitive functioning in elementary school children. Cooperative research project 2404*. San Francisco.

Taylor Ph., Richards M. 1985. *An introduction to curriculum studies.* Milton Keynes.

Tedesco J.C., Opertii R., Amadio M. 2014. The curriculum debate: Why it is important today. – *Prospects* 44: 527–546.

Tyler R. 1949. *Basic principles of curriculum and instruction*. Chicago.

Van Patten B., Williams J. (eds.). 2015. *Theories in second language acquisition.* London.

Vidmar D.J. 2005. Reflective peer coaching. Crafting collaborative self-assessment in teaching. – *Research Strategies* 20.3: 135–148.

Wegge J. 2001. Motivation, information processing and performance. Effects of goal setting on basic cognitive processes. – Efklides A., Kuhl J., Sorrentino R. (eds.). *Trends and prospects in motivation research*. Dordrecht: 269–296.

Weiss J. (ed.). 1991. *Évaluation – problème de communication*. Cousset.

Wiggins G. 1989. The case for authentic assessment. ERIC Documents ED328611. [available at www.eric.ed.gov; last access: February 8, 2018].

Wiliam D. 2010. The role of formative assessment in effective learning environments. – Dumont H., Istance D., Benavides F. (eds.). *The nature of learning. Using research to inspire practice.* Paris: 135–159.