



Content list available at <http://ijltr.urmia.ac.ir>

*Iranian Journal
of
Language Teaching Research*



Urmia University

Assessment for Learning in Language Education

Anthony Green ^{a, *}

^a *University of Bedfordshire, UK*

ABSTRACT

This paper describes the growing interest in assessment for learning (AfL) approaches in language education. It explains the term, traces the origins of AfL in developments in general education and considers the evidence for its claimed impact on learning outcomes. The paper sets out some of the challenges involved in researching, implementing and evaluating AfL initiatives in the context of language teaching and learning and considers how this may impact on our field in the future.

Keywords: assessment for learning; learning oriented assessment; feedback; classroom assessment; teacher assessment

© Urmia University Press

ARTICLE HISTORY

Received: 2 Sept. 2018

Revised version received: 20 Sept. 2018

Accepted: 21 Sept. 2018

Available online: 1 Oct. 2018

* Corresponding author: Center for Research in English Language Learning & Assessment, Bedfordshire University, UK

Email address: tony.green@beds.ac.uk

© Urmia University Press

Introduction

Building on the distinction first made in the late 1960s between summative purposes for assessment (retrospective, auditing learning and directed towards revealing the learner's state of knowledge) and formative purposes (prospective, guiding learning and directed towards changing the learner's state of knowledge) (Scriven, 1967; Bloom, Hastings & Madaus 1971), assessment for learning, often abbreviated as AfL, has emerged as a central concept in language education. AfL, which seeks to realise the educational benefits of formative assessment, originated in mainstream school and university education particularly in the UK and Australasia in the 1980s (see for example Elton, 1982; Black, 1986), but it has only relatively recently become established in relation to language assessment. In contrast to the traditional concerns of language testing, AfL assessments are often designed by teachers and may be part of regular classroom activities. Interaction between the teacher/assessor and the learner and the interpretation of student performance are fundamental (Stobart & Eggen, 2012) while standardized administration, quantitative data and formal statistical techniques, the traditional psychometric mainstays of language testing practice, are absent or peripheral.

In a pamphlet promoting the idea to teachers and policy makers titled *Assessment for Learning: Ten Principles* (Assessment Reform Group, 2002), a group of educational researchers in the UK collectively named the Assessment Reform Group, defined AfL in the following terms: "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Assessment Reform Group, 2002, pp. 2-3). AfL can be viewed as an umbrella term for a movement that seeks to harness assessment in the service of learning, using assessment data formatively to guide teaching and learning processes. Reflecting differences of origin, theoretical orientation or emphasis, terms such as dynamic assessment and learning-oriented assessment are sometimes used with similar implications.

Wiliam (2011) has used the metaphor of sports coaching to describe the relationship between teachers and learners envisaged in AfL. Just as a coach observes athletes to learn how they can improve their performance, the teacher collects evidence of how well a learner uses language, intervening in targeted ways to help them to progress. Where traditional teaching approaches follow a predetermined plan, AfL requires interaction and flexibility: adjustments are made according to how well learners are performing. Comparisons between what learners are able to do now in carrying out an assigned task and the kind of performance that would reflect fully successful learning are used to inform what happens next in the classroom. These comparisons are said to make it possible to work in a targeted way towards bridging the gap between current and intended levels of performance through appropriately chosen activities.

This contrasts with the tradition of testing, whether in the classroom or in large-scale tests (Teasdale & Leung, 2000). In a test, supporting learners with feedback during the assessment process undermines the validity of the procedure and would generally be regarded as unethical. Tests generate public statements about the success of learning and may contribute to life-changing decisions. Learners have an incentive to try to conceal or disguise any gaps in their understanding to obtain the highest possible scores. In AfL, openness about gaps in understanding is helpful to the learner as it prompts action to fill them.

Although AfL provides support for teaching, it is ultimately concerned with the development of effective learning, something for which teachers and learners are considered to be jointly responsible. The engagement of learners in assessment is therefore one of the key principles of AfL (Assessment Reform Group, 2002). When learners leave the classroom, they should have the capacity to continue their learning independently of the teacher. To this end, learners are

encouraged to assess themselves (self-assessment) and others (peer assessment), to monitor their own performance and to seek out feedback as a basis for continuing development.

An important aspect of AfL research has been the use of large-scale reviews and meta-analyses of educational practice, bringing together a wide range of research findings relating to formative assessment together with a range of other variables with potential effects on learning outcomes. This work has included an influential review commissioned by the Assessment Reform Group authored by Black and Wiliam (1998) as well as such studies as Fuchs and Fuchs (1986), Kluger and DeNisi (1996), and Hattie and Timperley (2007).

Questions have been raised about the underlying research relating to the quality of the studies themselves and to the adequacy of the reporting, and the size of the effects on learning outcomes has been disputed (Bennett, 2011). Notwithstanding these criticisms, this strand of research has led to advice for educators on “what works” in the classroom. The positive classroom practices tending to lead to better learning outcomes identified by the Assessment Reform Group included the following (from Black & Wiliam, 2009):

- i) Sharing success criteria with learners: helping learners to understand the qualities of successful task performance by, for example, modelling a good performance or involving learners in developing rating criteria. AfL emphasizes the authenticity of learning tasks as representations of language use in realistic contexts that will be socially meaningful for language learners: linking success in real world language use with success in the classroom.
- ii) Classroom questioning techniques directed not only at revealing what learners know, but also at uncovering sources of misunderstanding and involving all learners in thinking through the issues at hand. This supports the provision of feedback on learner performance that highlights what can be done to improve it so that it better matches the success criteria.
- iii) Comment-only marking: avoiding giving grades to learners' work because research suggested that learners tend to ignore all other feedback once a grade has been given. In other words, the use of assessment for summative purposes such as grading may restrict or reverse its beneficial impact on the learning process.
- iv) Self- and peer-assessment: engaging learners in evaluating the quality of their own work and that of others. This helps learners to better understand success criteria and to plan and monitor their own progress. This has been described as *assessment as learning* (see for example Earl, 2012) to reflect the central role of the learner in the process. The value placed on authenticity is partly because of the role it can play in motivating learners. If learners understand the connections between what they do in the classroom and language use in the world beyond, they are more likely to take responsibility for their learning and work to improve their outcomes.
- v) The formative use of summative tests: using the results of summative tests to review what learners know and as a prompt to address weaknesses and build on successes. Although AfL can be contrasted with traditional testing, almost all education systems involve the latter. The most effective learning is likely to occur where teaching and testing can operate in concert. The importance of reconciling diverse purposes for assessment is reflected in the notion of learning oriented language assessment (Green & Hamp-Lyons, 2015), which embraces both.

Hattie and Timperley (2007) and Wiliam and Thompson (2007) have linked these strategies to Ramaprasad's (1983) three processes of learning and teaching: i) establishing where the learners are in their learning; ii) establishing where they are going and iii) establishing what needs to be done to get them there. AfL strategies such as classroom questioning and discussion support the first of these processes, sharing and clarifying success criteria support the second and feedback is fundamental to deciding what should happen next to progress learning.

Practice and theory

Its practical orientation has led to criticism that AfL lacks a coherent theoretical basis and that links to theory have been under-elaborated (Baird, Andrich, Hopfenbeck & Stobart, 2017). Although generally left implicit, there are clear points of contact between AfL and theories of learning and, as AfL has developed, its proponents have begun to formalise these (Black & William, 2009). Learner engagement in assessment invokes areas such as learner autonomy (Benson & Voller, 2014), metacognition (Oxford, 2016), motivation (Dornyei & Ushioda, 2013), self-regulation (Bailey & Heritage, 2018) and self-theories (Dweck, 2000). It is most often connected with constructivist learning theory, but has also often been associated with sociocultural approaches (Vygotsky, 1978) (Baird et al. 2017). Baird et al. (2017) note that in practice its implementation is affected by local cultural traditions, contrasting the use of data from periodic testing in Hong Kong or the US, reflecting behaviourist assumptions about learning, with greater emphasis on the role of the learner and informal classroom assessment in other anglophone countries, reflecting a more socio-constructivist or sociocultural orientation.

As Baird et al. (2017) observed, "AfL is an approach that needs to be adapted for each subject rather than a general framework that can be directly applied to any subject" (p. 337). In applying AfL principles in language education, there is no requirement for users to adopt a particular perspective on language, but much depends on language teachers' beliefs about progress in language learning and their ability to sequence instruction in ways that reflect these. Whether based solely on classroom interaction or involving planned, formal assessment, the success of AfL in language classrooms rests on the capacity for teachers and learners as designers of assessment procedures and users of assessment information to generate appropriate evidence and to draw appropriate inferences from it, leading to appropriate actions that enhance learning.

Because AfL represents a general approach to assessment and learning, it provides only limited guidance on the kinds of task that might best elicit evidence of language development, or drive it forward. Although it is important to use tasks to gain insights into language learning processes, understanding of these processes must come from elsewhere. One source might be theories of second language acquisition (SLA). Feedback from teachers has been an issue of interest to SLA researchers, although the way in which feedback is conceptualized (the focus being on the treatment of grammatical errors) is generally rather different from the way it is understood in AfL (which includes all aspects of learner performance). Contrasting positions on the benefits of error correction are exemplified by Truscott (1999), who argues that error correction is ineffective, and Long and Robinson (1998), who consider that feedback on errors is essential to the acquisition of form-meaning relationships. Depending on which school of thought they favour, teachers might make very different decisions about what kinds of information to feedback to learners or how to use assessment data to inform their next steps. Given that some forms of feedback appear to have a negative impact on learning (Hattie & Timperley, 2007), it is important to be able to distinguish between different types of feedback and to understand the processes through which these affect learning.

Unfortunately, as Fulcher (2011) has observed, SLA evidence on the order in which words and structures are acquired provides only limited guidance for teachers on how best to sequence instruction. Another source of support may come from frameworks such as the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001), which suggests a sequence of levels of functional language ability which may inform the development of locally appropriate learning objectives based on a collective pedagogic understanding of progression in language learning.

Some work has been done to reconcile this pedagogic perspective with second language acquisition research by relating the CEFR to evidence from corpora of learner language (Green, 2012) to build a more grounded picture of how learners develop their language abilities. On the other hand, sociocultural views of language question whether it is possible to trace any general pattern of progression in language learning, but envisage the teacher and learner working collaboratively to negotiate meaning in the context of each interaction. In any case, the sequencing of learning in the classroom is likely to be mediated by curricula and text books that vary in the extent to which they reflect realistic learning objectives or engage a learner's interest.

AfL and quality standards

Whereas there now exist consensus standards for quality in traditional language testing, enshrined in documents such as the International Language Testing Association Guidelines for Practice (ILTA, 2007), there is no such consensus on the qualities that should be demonstrated for AfL or on evidence that should be reported to demonstrate its value.

Some suggest that traditional standards for reliability and validity should apply to all assessments (see for example, Brown and Hudson (1998), but that expectations may be relaxed for AfL because mistakes are more easily remedied and their consequences are not as serious. However, traditional psychometric approaches to test validation rely on quantitative results that can be aggregated and analyzed using statistical techniques. As noted above, AfL does not typically generate such data, or involve claims based on variance within populations. Efforts have been made to develop novel psychometric approaches that are better able to address the contingencies of performance data, but the pressure to demonstrate reliability continues to encourage conservatism and the use of test-like procedures. Statistical techniques are not suited to determining in real time whether a teacher's interpretation of a classroom interaction is justified by the available evidence.

To date, the most influential evidence for the success of AfL initiatives has come from test scores at the end of a period of instruction, with learners involved in AfL making greater gains (Baird et al., 2017). Although this points to the success of AfL approaches where these have been embraced by teachers, it is much less helpful in guiding the day to day practice of teachers involved in implementing AfL in the classroom. There is general agreement that for AfL to succeed, a number of conditions must be met. First, teachers must be able to design tasks that will generate useful evidence of learners' abilities. They must be able to interpret the evidence appropriately and so adjust their teaching in ways that will benefit learning. Second, they must help learners to engage in the process of learning: to set themselves appropriate goals and to come to understand the criteria for success. Third, feedback must be taken up and used effectively both to engage learners and to move learning forward. On the other hand, it has proved challenging to introduce AfL on a large scale as teachers and other stakeholders have often proved reluctant or unable to change their established methods. It is clear that simply

mandating AfL at an institutional or national level is unlikely to result in effective enactment in the classroom.

Teachers as well as other stakeholders such as policy makers and parents tend to base their understanding of assessment on their past experiences. They conceive of it primarily as a tool for grading and ranking and find it difficult to adjust to thinking of it primarily as a means of driving learning. This perspective also tends to predominate in teacher training, where assessment is associated with the end of a teaching sequence, rather than being embedded in the learning process.

In practice, it has proved very hard to extend the success of AfL beyond small-scale projects involving teachers who are already open to the concept. Teachers involved in AfL initiatives often report that the experience has helped them to improve their own pedagogy and knowledge of their subject, but such benefits are more likely to occur when teachers are sympathetic to the aims of the innovation. Attempts to oblige teachers to engage in AfL by asking them to record progress against mandated levels or standards risk alienating teachers and promoting a culture of bureaucracy (Brindley, 2001). On the other hand, grassroots initiatives by enthusiastic teachers may fail unless provided with institutional support.

Accounts such as Rea-Dickins (2001) have demonstrated that teachers may struggle to balance the demands of AfL with continuing need to award grades for purposes of external accountability. The importance of obtaining high scores on national or international tests may encourage a narrow focus on rehearsing test tasks and on covering the content that will be tested, at the same time discouraging more flexible, interactive approaches. When individual teachers attempt to experiment with AfL in their classes, this may conflict with established grading policies. AfL projects that have brought success have tended to involve long term commitment and investment in teacher development and dialogue between teachers and education authorities (Hopfenbeck, Flórez Petour & Tolo, 2015).

Bennett (2011) recommended that the introduction of AfL should be based on an explicit 'theory of action' presenting the rationale for each of the elements of the system, the basis for interpreting performance as evidence of learning, intended effects of introducing the assessment system, steps to be taken to achieve those effects and steps to be taken to reduce the risks of any unintended negative consequences. Bennett suggested that a number of issues would need to be addressed if the apparent advantages of AfL were to be realized on a large scale. These included greater clarity of descriptions of AfL in the research literature; standards for evaluating AfL procedures; the interpretation and use of evidence collected through AfL for the guidance of learning; professional development for teachers; and management of change.

Conclusions

Although interest in AfL in the language classroom has grown, research into teachers' formative assessment practices in language classrooms remains limited (Leung & Mohan, 2004; Inbar-Lourie & Donitsa-Schmidt, 2009; Tang, Leung, Chow & Wong, 2010). To guide research in language education, Hill and McNamara (2012) and Turner and Purpura (2016) have suggested approaches to categorizing salient features of feedback and assessment systems that can provide clearer descriptions of what they involve. Use of such systematic approaches to description should help to clarify how practices compare with each other and how they might contribute to learning outcomes. They would also support practitioners who wish to adopt similar approaches, or adapt systems of assessment to other settings.

Experience from general education suggests that teacher beliefs and teacher development are essential to the successful implementation of AfL. Work will be needed to establish how the implementation of AfL can be effected under a variety of social and cultural conditions and whether the beneficial effects are generalizable across diverse contexts. It has also been suggested that consideration of assessment issues needs to be better integrated into teacher training programmes so that teachers have better models for practice than those they experienced as students. This has led to calls for improving the assessment literacy of language teachers through better training programmes (Inbar-Lourie, 2008).

With technological advances, consideration is being given to how best to realise the benefits of AfL through new assessment tools. Computer assisted language learning has often involved reversion to the very mechanistic forms of textbook exercise that are easiest to deliver using the available systems, but this situation is improving. The emerging field of learning analytics involves exploiting the massive quantities of learner data that can be gathered through online learning tools to design and deliver more personalised learning experiences (Thomas & Gelan, 2018). This promises to provide new insights and adds a dimension of feedback to teachers and learners that could enhance learning. However, this will require considerable investment in research to accompany the development of new learning tools.

References

- Assessment Reform Group. (2002) *Assessment for learning: 10 principles*. Cambridge: University of Cambridge School of Education.
- Bailey, A. L., & Heritage, M. (2018). *Self-Regulation in learning: The role of language and formative assessment*. Cambridge, MA: Harvard Education Press.
- Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart?. *Assessment in Education: Principles, Policy & Practice*, 24(3), 317-350.
- Bennett, R. E. (2011) Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Benson, P., & Voller, P. (2014). *Autonomy and independence in language learning*. Abingdon, Oxon: Routledge.
- Black, H. D. (1986). Assessment for learning. In D. L. Nuttall (Ed.), *Assessing educational achievement* (pp. 7-18). London: Falmer.
- Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook of formative and summative evaluation of student learning*. New York: McGraw-Hill.

- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing*, 15(1), 45-85.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393-407.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: a review of the literature. In J. H. McMillan (Ed.), *Formative Classroom Assessment: Theory into Practice* (pp. 43-62). New York, NY: Teachers College Press.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL quarterly*, 32(4), 653-675.
- Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.
- Dörnyei, Z. (2018). Motivating Students and Teachers. *The TESOL Encyclopedia of English Language Teaching*, pp. 1-6. Medford MA: Wiley.
- Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching motivation*. Abingdon, Oxon: Routledge.
- Dweck, C. S. (2000). *Self-Theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology press.
- Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Elton, L. (1982). Assessment for Learning, in Bligh, D. (ed), *Professionalism and Flexibility in Learning, Programme of Study into the Future of Higher Education* (The Leverhulme Study), Society for Research into Higher Education, Vol 6, pp. 106 – 135. Guildford: The Society for Research into Higher Education.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation—a meta-analysis. *Exceptional Children*, 53(3), 199-208.
- Fulcher, G. (2011). *Practical language testing*. Abingdon, Oxon.: Routledge.
- Green, A. (2012). *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range* (Vol. 2). Cambridge: Cambridge University Press.
- Green, A., & Hamp-Lyons, L. (2015). *Introducing opportunities for learning-oriented assessment to large-scale speaking tests*. Final project report for Cambridge English Language Assessment research funding Round 4.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420.

- Hopfenbeck, T. N., Flórez Petour, M. T., & Tolo, A. (2015). Balancing tensions in educational policy reforms: Large-scale implementation of assessment for learning in Norway. *Assessment in Education: Principles, Policy & Practice*, 22(1), 44-60.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16, 185-204.
- International Language Testing Association. (2007). *ILTA guidelines for practice*. Retrieved on June 30, 2018 from <https://www.iltaonline.com>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing*, 21(3), 335-359.
- Long, M., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- O'Malley, J. M., & Valdez Pierce, L. (1996) *Authentic assessment for English language learners: Practical approaches for teachers*. Reading, MA: Addison-Wesley.
- Oxford, R. L. (2016). *Teaching and researching language learning strategies: Self-regulation in context*. Abingdon, Oxon: Routledge.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives on curriculum evaluation. AERA monograph series – Curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally & Co.
- Stobart, G., & Eggen, T. (2012). High-stakes testing – Value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, 19(1), 1-6.
- Tang, S. Y. F., Leung, P. P. W., Chow, A. W. K., & Wong, P. M. (2010). A case study of teacher learning in an assessment for learning project in Hong Kong. *Professional Development in Education*, 36, 621-636.
- Teasdale, A., & Leung, C. (2000) Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, 17(2), 163-184.

- Thomas, M. & Gelan, A. (2018) Editorial: Special edition on language learning and learning analytics, *Computer Assisted Language Learning*, 31(3), 181-184.
- Truscott, J. (1999). What's wrong with oral grammar correction? *The Canadian Modern Language Review*, 55(4), 437-56.
- Turner, C. E., & Purpura, J. E. (2016) Learning-oriented assessment in second & foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.) *Handbook of Second Language Assessment* (pp. 255–274), Boston, MA: De Gruyter, Inc.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wiliam, D. (2011) *Embedded formative assessment*. Bloomington, IN: Solution Tree.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The Future of Assessment: Shaping Teaching and Learning* (pp. 53–82). Mahwah, NJ: Erlbaum.

Anthony Green is Professor in Language Assessment and Director of the Centre for Research in English Language Learning and Assessment at the University of Bedfordshire, UK. His main research interests concern relationships between language assessment, teaching and learning. He has published widely on language assessment and was President of the International Language Testing Association (ILTA) in 2015-16. His book *Exploring Language Assessment and Testing* (Routledge, 2014) provides trainee teachers and others with an introduction to this field.