The Assessment-for-learning (AfL) movement in primary and secondary education: Lessons for higher education

16 May 2018

University of Kent

Assessment and Feedback in Higher Education: A Research Symposium

Therese N. Hopfenbeck

Oxford University Centre for Educational Assessment
Some key-points

• The Assessment for Learning Movement
• What researchers agree on
• What researchers disagree on
• Critical points
• Lessons learned?
• A way forward?
Assessment in Education: Principles, Policy & Practice

Publish open access in this journal

Enter keywords, authors, DOI etc

Submit an article  New content alerts  RSS  Citation search

Current issue  Browse list of issues

This journal

Aims and scope

Instructions for authors

Journal news

Now included in the Thomson Reuters Emerging Sources Citation Index

Book review guidance notes

Call for papers

Special issue: Assessment and self-regulated learning

Special issue: Systemic influences on standard setting in national examinations
assessment - in education - principles - policy - and practice/ 20 years after the Black Box and Black and William 2018

(L-R; Therese Hopfenbeck, Ernesto Panadero, Patricia Broadfoot, Gordon Stobart, Jo-Anne Baird, Mary James, Paul Black, Dylan Wiliam)
Why AfL?

There is a body of literature claiming that AfL, or formative assessment, is one of the most powerful tools to increase students’ learning (Bangert-Drowns et al., 1991; Black & Wiliam, 1998a, 1998b; Brookhart, 2004, 2007; Crooks, 1988; Dempster, 1991, 1992; Elshout-Mohr, 1994; Fuchs & Fuchs, 1986; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Köller, 2005; Natriello, 1987; Shute, 2008; Wiliam, 2007).

The main ideas from the literature focus upon how students can use different forms of feedback and assessment to improve their learning, from peers, teachers or others, as long as it is of high quality and timely. Self-assessment is also included in the AfL processes, a procedure that demands an active student role, but also helps students to gain more insight into their learning and make them more responsible for their own learning outcomes (Stobart, 2008).

There seems to be consensus in research regarding the positive effects of AfL, as perceived by research participants (see, for example, Ofsted, 2008; DfES, 2007; Condie et al., 2005; Hayward & Spencer, 2010; Kirton et al., 2007; Webb & Jones, 2009; Kellard et al., 2008, among others).
The Assessment for Learning movement


Its take-up has been described as a ‘research epidemic’ which in a relatively few years has ‘feverishly spread into every discipline and professional field’ (Steiner-Khamsi, 2004, p. 2).
Large-scale implementation of Assessment for Learning

Therese N. Hopfenbeck & Gordon Stobart
pages 1-2

DOI: 10.1080/0969594X.2014.1001566
Published online: 30 Jan 2015
Citing articles: 0
Article Views: 125
<table>
<thead>
<tr>
<th>Where</th>
<th>Who</th>
<th>What</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>Jonsson, Holmgren and Lundahl</td>
<td>large-scale regional interventions</td>
</tr>
<tr>
<td>Canada</td>
<td>DeLuca, Klinger, Piper and Woods</td>
<td>large-scale regional interventions</td>
</tr>
<tr>
<td>US</td>
<td>Wylie and Lyon</td>
<td>large-scale regional interventions</td>
</tr>
<tr>
<td>Scotland</td>
<td>Hayward</td>
<td>National policy</td>
</tr>
<tr>
<td>Norway</td>
<td>Hopfenbeck, Tolo, Florez</td>
<td>National policy/intervention</td>
</tr>
<tr>
<td>Chile</td>
<td>Florez Tong</td>
<td>National policy</td>
</tr>
<tr>
<td>Singapore</td>
<td>Li Ratnam and Tan</td>
<td>National policy</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>De Lisle</td>
<td>National policy</td>
</tr>
</tbody>
</table>
What researchers agree on
Assessment for learning

- Clarifying and sharing learning intentions and criteria for success
- Engineering effective classroom discussions and other learning tasks that elicit evidence of student achievement
- Providing feedback that moves learners forward
- Activating students as instructional resources for one another
- Activating students as the owners of their own learning

Black and William (2009)
Assessment Reform Group

“The process of seeking and interpreting evidence for use by learners and their teachers, to identify 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)
"Assessment for learning (AfL) is a conscious attempt to make assessment a productive part of the learning process. It does this by making classroom assessment an essential part of effective teaching and learning" (Stobart, 2008: 144).
From Black and Wiliam 2018:

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Where the learner is going</th>
<th>Where the learner is right now</th>
<th>How to get there</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Clarifying learning intentions and criteria for success</td>
<td>2 Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding</td>
<td>3 Providing feedback that moves learners forward</td>
</tr>
<tr>
<td>Peer</td>
<td>Understanding and sharing learning intentions and criteria for success</td>
<td>4 Activating students as instructional resources for one another</td>
<td></td>
</tr>
<tr>
<td>Learner</td>
<td>Understanding learning intentions and criteria for success</td>
<td>5 Activating students as the owners of their own learning</td>
<td></td>
</tr>
</tbody>
</table>
Research-based principles of assessment for learning to guide classroom practice

Assessment for Learning is 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.
Ten principles for assessment

- Assessment
  - is part of effective learning
  - focuses on how students learn
  - is central to classroom practice
  - is a key professional skill
  - is sensitive and constructive
  - fosters motivation
  - promotes understanding of goals and criteria
  - helps learners know how to improve
  - develops the capacity for self-assessment
  - recognises all educational achievement

Assessment Reform Group (2002)
What researchers disagree on
Some background

1. Relationships between substantive learning theory & assessment
2. Theoretical and philosophical dilemmas
3. International tests – Assessment for Learning
Lack of empirical research
Review of AfL

394 journals in the field of education and assessment were selected for inclusion: peer-reviewed journal articles (907) book reviews (29) conference proceedings (481).

For the 907 articles in the field of education, we read the abstracts and categorised them according to whether they were small or large-scale studies, case studies, and theoretical or empirical.
Search Criteria:
Electronic database search: SOLO (University of Oxford)
EndNote search function: Web of Science, ERIC, Pubmed, Zetoc
Related articles search (Science Direct)
Citing articles search (Google Scholar)

References deleted:
Duplicates
Books and book sections

Articles after screening:
N = 2131

Included articles must be from subject-related journals
Education journal articles:
N = 960

Common excluded subjects:
Medical articles:
N = 376
Engineering articles:
N = 63

Included articles must be articles from peer-reviewed journals
Peer-reviewed journal articles:
N = 907

Excluded reference types: conference proceedings, book
Conference proceedings:
N = 481
Book reviews:
N = 29

Baird, Hopfenbeck, Newton, Stobart & Steen-Utheim, 2014
Findings I

• Most articles fell into the category of (1) case studies (and often one or two schools, with fewer than ten participants – such as the study by Hollingworth, 2012 and Sato 2005), (2) theoretical articles (Black and Wiliam 2009, Sadler 2010) or (3) action research (Harrison 2013, McDowell 2008).
Findings II

Very few abstracts reported how studies had been conducted and which methods had been applied. Most abstracts did not give any information on how many participated in the study or what methods of analysis had been carried out. Most empirical research studies focused upon issues such as students’ perceptions of AfL, either their experience of feedback from teachers (Hounsell et al. 2008) or their beliefs about AfL. Very few reported anything about learning outcomes.
The critique

“Whilst claims for large effect sizes are regularly made in the literature, the evidence for these has increasingly been critiqued, for example by Bennett (2011) and Kingston and Nash (2011)” from Baird et al (2014).

“Our systematic review showed that the vast majority of studies on AfL are small-scale action research designs and are published in a wide range of journals. A concern for the review is that current definitions of formative assessment/AfL cover a wide range of teaching and learning practices while research designs often lack an action theory (what is causing change), often accompanied by a lack of systematic data collection (for example baseline data before a research initiative).

Our overall conclusion is that the effects of formative assessment upon learning have been over-sold by some authors. This is unfortunate because the limited empirical research suggests a modest, but educationally significant, impact on teaching and learning” (ibid p.6)
Formative assessment: a critical review

Randy Elliot Bennett*

Research and Development, Educational Testing Service, Princeton, NJ, USA

This paper covers six interrelated issues in formative assessment (aka, ‘assessment for learning’). The issues concern the definition of formative assessment, the claims commonly made for its effectiveness, the limited attention given to domain considerations in its conceptualisation, the under-representation of measurement principles in that conceptualisation, the teacher-support demands formative assessment entails, and the impact of the larger educational system. The paper concludes that the term, ‘formative assessment’, does not yet represent a well-defined set of artefacts or practices. Although research suggests that the general practices associated with formative assessment can facilitate learning, existing definitions admit such a wide variety of implementations that effects should be expected to vary widely from one implementation and student population to the next. In addition, the magnitude of commonly made quantitative claims for effectiveness is suspect, deriving from untraceable, flawed, dated, or unpublished sources. To realise maximum benefit from formative assessment, new development should focus on conceptualising well-specified approaches built around process and methodology rooted within specific content domains. Those conceptualisations should incorporate fundamental measurement principles that encourage teachers and students to recognise the inferential nature of assessment. The conceptualisations should also allow for the substantial time and professional support needed if the vast majority of teachers are to become proficient users of formative assessment. Finally, for greatest benefit, formative approaches should be conceptualised as part of a comprehensive system in which all components work together to facilitate learning.
Formative Assessment: A critical Review

- The definition of formative assessment/assessment for learning
- The claims commonly made for its effectiveness
- The limited attention given to domain concerns in its conceptualisation
- The under-representation of measurement principles in that conceptualisation
- The impact of the larger education system

Bennet (2011)
In ‘Inside the Black Box’, Black and Wiliam (1998a) estimated that feedback would increase students’ learning within the range of effect sizes from 0.4 to 0.7, while the review by Shute (2008) found effect sizes between 0.4 and 0.8. In the review of 74 meta-analyses, conducted by Hattie and Timperley (2007), an average effect size of 0.95 standard deviations was reported from an analysis of 4,157 studies. Other writers reported learning gains with effective sizes between 0.4 and 0.7 (Popham, 2008b; Stiggins, 1999).
“A total of 24 teachers (2 science and 2 mathematics teachers, in each of six schools in two LEAs) were supported over a six-month period in exploring and planning their approach to formative assessment, and then, beginning in September 1999, the teachers put these plans into action with selected classes. In order to compute effect sizes, a measure of prior attainment and at least one comparison group was established for each class (typically either an equivalent class taught in the previous year by the same teacher, or a parallel class taught by another teacher). The mean effect size in favour of the intervention was 0.32.”

Thus the improvement of formative assessment cannot be a simple matter. There is no ‘quick fix’ that can be added to existing practice with promise of rapid reward.
Critique of AfL

Given that Black and Wiliam’s (1998b) review was not a meta-analysis, but is generally reported as one, Bennett described this as a ‘mischaracterisation that has essentially become the educational equivalent of urban legend’ (p12). He also critically reviewed widely quoted reviews by Nyquist (2003), Meisels et al. (2003) and Rodriguez (2004).

Kingston and Nash's review was in turn critiqued for its methodology by Derek Briggs and colleagues (2012). They had concerns about the narrow initial search for studies and about inconsistencies in what was selected. Some of the variability in effect sizes could be explained by the different calculation techniques, it was argued. They concluded that as a consequence, even the 0.20 estimate cannot be claimed with confidence – the effect sizes from a different sample might have been significantly different from than this.
Bennett’s (2011, p17) claim was that the formative assessment literature gave too little attention to the interpretation of assessment evidence. The process of finding out what a learner knows and understands involves a formative hypothesis that requires further validation through the collection of further evidence. Such inferences are uncertain and are subject to systematic, irrelevant influences that may be associated with gender, race, ethnicity, disability, English language proficiency, or other student characteristics. Put simply, a teacher’s formative assessment may be unintentionally biased. (Baird et al p.47)
A way forward?
Many schools give the impression of having implemented AfL when in reality the change in pedagogy that it requires has not taken place. This happen, for example, when teachers feel constrained by external tests over which they have no control. As a result they are unlikely to give pupils a greater role in directing their learning, as is required in AfL, in order to develop the capacity to continue learning throughout life (2006:10)
Explore

Most read articles

Article
Teachers developing assessment for learning: impact on student achievement

Wiliam * et al.
Volume 11, 2004 - Issue 1
Published online: 27 Sep 2010
Views: 60510

Article
Assessment and Classroom Learning

Black et al.
Volume 5, 1998 - Issue 1
Published online: 28 Jul 2006
Views: 52744

Article
Formative assessment: a critical review

Bennett
Volume 18, 2011 - Issue 1
Published online: 25 Jan 2011
Views: 35658
Self-regulated learning and AfL

Self-regulated learning (is) an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment (Pintrich, 2000c, p. 453).
Why self-regulation?

“We are in the midst of a revolution in educational thinking and practice. Scientific advances in a number of fields point to a similar argument – how well students do in school can be determined by how well they are able to self-regulate. Some theorists believe that self-regulation should now be considered a more important indicator of educational performance than IQ, (Blair & Diamond, 2008, Duckworth & Seligman, 2005, Schonkoff & Phillips, 2000)” in (Shanker, 2013, p.ix)

“We need to develop an awareness of what we are doing, where we are going, and how we are going there, we need to know what to do when we do not know what to do. Such self-regulation, or meta-cognitive skills are one of the ultimate goals of all learning, they are what we often mean by life-long learning and it is why we want students to become their own teachers”.

(Hattie 2012: 115).
Life long learners?

Self-regulating learners know how to read situations, set goals and take control over their actions (Butler, Schnellert, Perry, 2017).
The Secret of Self-Control

Children who are able to pass the marshmallow test enjoy greater success as adult. Read more http://www.newyorker.com/reporting/2009/05/18/090518fa_fact_lehrer#ixzz1j67UWWUE
Chamot, 1999
Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going

Ernesto Panadero\textsuperscript{1} \cdot Heidi Andrade\textsuperscript{2} \cdot Susan Brookhart\textsuperscript{3}

Received: 16 November 2017 / Accepted: 29 January 2018 / Published online: 8 February 2018
© The Australian Association for Research in Education, Inc. 2018

Abstract We have known for a long time that a relationship exists between how learning is assessed and the learning processes and strategies students employ when engaged in those assessments. Black and Wiliam pointed out in 1998 that self-regulated learning should be a primary goal of formative assessment (FA). Since then, a growing body of research on this relationship has been produced. The purpose of this paper is to present and discuss keystone publications that inform our current understandings of the relationship between FA and self-regulated learning. The result is a roadmap of the development of the field and directions for future research.

Keywords Self-regulated learning \cdot Coregulated learning \cdot Classroom assessment \cdot Formative assessment \cdot Assessment for learning \cdot Self-assessment
Fig. 1: Formative assessment and self-regulated learning publications timeline.

Early years (1988-2000)
- Crooks, 1988
- Sadler, 1989
- Black & Wiliam, 1998
- Perrenoud, 1998

Connections made between assessment and two SRL-related concepts: learning strategies and self-monitoring.

Middle years (2001-2012)
- Paris & Paris, 2001
- Allal & López, 2005
- Brookhart, 2005
- Nicol & McFarlane-Dick, 2006
- Hattie & Timperley, 2007
- Shute, 2008
- Andrade & Valtcheva, 2009

Connections made b/w SRL and self-assessment.
English publications by French-language scholars about assessment as regulation.

Present time (2013-2018)
- Allal, 2010
- Andrade, 2010
- Wiliam, 2011
- Clark, 2012
- Andrade & Brookhart, 2016
- Dinsmore & Wilson, 2016
- Panadero, Jonsson & Strijbos, 2016
- Reinholz, 2016
- Panadero, Jonsson & Botella, 2017
- Panadero & Broadbent, 2018
- Andrade & Brookhart, under review

Level of detail increases as specific mechanisms are explored, especially self-assessment. Recognition that SRL and FA scholarship tend to approach the same phenomena from different perspectives, and the benefits of unifying them. First empirical reviews of FA and SRL. AERA 2014 symposium on assessment as regulation.

Higher level of detail in theoretical and empirical treatments of FA and SRL. Empirical reviews grow in number. Use of co-regulation increases.
Supporting and developing learner self-regulation

1. Clarify what good performance is
2. Facilitate self-assessment
3. Deliver high quality feedback information
4. Encourage teacher and peer dialogue
5. Encourage positive motivation and self-esteem
6. Provide opportunities to close the gap
7. Use feedback to improve teaching
How to use AfL in the classroom?

“Thus the improvement of formative assessment cannot be a simple matter. There is no ‘quick fix’ that can be added to existing practice with promise of rapid reward. On the contrary, if the substantial rewards of which the evidence holds out promise are to be secured, this will only come about if each teacher finds his or her own ways of incorporating the lessons and ideas that are set out above into her or his own patterns of classroom work.

This can only happen relatively slowly, and through sustained programmes of professional development and support. This does not weaken the message here—indeed, it should be a sign of its authenticity, for lasting and fundamental improvements in teaching and learning can only happen in this way”.

# Student participation in assessment

<table>
<thead>
<tr>
<th>Positive responses in percentage</th>
<th>Primary 6 - 12</th>
<th>Lower: 13 – 15</th>
<th>Upper 16 – 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>The teacher asks me how I assess my own work</td>
<td>61 (69)</td>
<td>40 (46)</td>
<td>35</td>
</tr>
<tr>
<td>The teacher asks me whether I have any suggestions for how to improve my work</td>
<td>74 (78)</td>
<td>47 (56)</td>
<td>49</td>
</tr>
<tr>
<td>The teachers asks students to assess each others work</td>
<td>54 (48)</td>
<td>36 (29)</td>
<td>30</td>
</tr>
<tr>
<td>Teachers asks students to take part in the decisions of what should be assessed</td>
<td>50 (58)</td>
<td>36 (38)</td>
<td>31</td>
</tr>
</tbody>
</table>

Hopfenbeck, Thronsen, Lie og Dale 2012)
Please cite this paper as:

http://dx.doi.org/10.1787/5k3txnplsn-7en

OECD Education Working Papers No. 97

Balancing Trust and Accountability? The Assessment for Learning Programme in Norway

A GOVERNING COMPLEX EDUCATION SYSTEMS CASE STUDY
An example: The Norwegian Education Act

• Understand what they are supposed to learn and what is expected of them;
• Receive feedback that informs them about the quality of their work or performance;
• Receive advice on how they can improve;
• Are involved in their own learning activities for example through assessing their own work and development.
Assessment for Learning in Africa (AFLA): Improving Pedagogy and Assessment for Numeracy in Foundation Years

Major ESRC-DfID Research Grant Confirmed (ES/N010515/1)

Oxford University Centre for Educational Assessment (OUCEA) has been successful in its bid for ASSESSMENT FOR LEARNING IN AFRICA (AFLA): Improving Pedagogy and Assessment for Numeracy in Foundation Years to the ESRC-DfID and awarded £695,210 to conduct research on formative assessment in primary school early years’ numeracy contexts in Tanzania, East Africa and two sites in South Africa. OUCEA will lead the project with Principal Investigator Dr. Therese N. Hopfenbeck. To find out about the research team, click here.

Our approach is to develop classroom materials for primary school numeracy development and use these as the basis for workshops and for the development of teacher learning communities in each.
Feedback in Higher Education

Steen-Utheim & Hopfenbeck (accepted for publication) “To do or not to do with feedback: A study of undergraduate students’ engagement and use of feedback within a portfolio assessment design”, *Assessment & Evaluation in Higher Education*.

Some take-home messages, confirming previous studies:

- **Students prefer oral feedback**
- **Students struggle with understanding written feedback**
- **SRL influences the way students’ act upon feedback**
Questions or feedback?

therese.hopfenbeck@education.ox.ac.uk

@TNHopfenbeck