An instrument to audit teachers' use of assessment for learning

Zita Lysaght[1] and Michael O'Leary

St. Patrick's College, Dublin

Abstract

This paper traces the design, development and trialling of an assessment for learning audit instrument (AfLAi) in use currently in the Republic of Ireland to gauge teachers' baseline understanding of AfL practices and the extent to which AfL is embedded in their teaching. As described in the paper, the AfLAi consists of 58 items distributed across four scales based on the following key AfL strategies: *sharing learning intentions and success criteria*, *questioning and classroom discussion*, *feedback*, and *peer- and –self assessment*. Preliminary data from the study provide a window into current formative assessment practices in Irish primary schools and teachers' professional needs in AfL.

Keywords

Assessment for learning, Formative assessment, Audit instrument, Teacher professional development, Assessment audit instruments

Introduction

In an editorial reflecting on authors' contributions to a special issue of *Educational Measurement: Issues and Practice,* Brookhart (2009, 2) reminds us that 'formative assessment is as much about learning as it is about assessment'. Her remarks reflect a growing acknowledgement within the education community that the warranty for formative assessment, established in the main by Black and Wiliam (1998), inadvertently - though not altogether unexpectedly, given the powerful backwash effects of high stakes assessment - spawned a disproportionate interest in the design, development and marketing of tests, assessments and data-mining systems. As a consequence, educators are now challenged with

---

[1] Zita.lysaght@spd.dcu.ie

the task of restoring 'the balance between "assessment" and "formation" in formative assessment' (3-4) and, *ipso facto,* refocusing attention on the dynamics of classroom life. Required are coal-face investigations of the processes involved in the democratisation of learning and teaching, including the contingency and unpredictability which of necessity ensue when teachers and students prioritise learning over everything else.

This paper argues (somewhat ironically, given the disproportionate interest noted previously) that an integral element of this work is the design of assessment instruments to support teachers' personal review of existing classroom practices and provide them with site-based information about the extent to which they use assessment information on a minute-by-minute, day-by-day basis (Leahy, Lyon, Thompson and Wiliam 2005). Specifically, the paper describes the design, development and trialling of one such instrument - the AfL audit instrument (AfLAi) – the information from which is currently being used by practitioners in the Republic of Ireland to guide their use of formative assessment in their schools. The paper is framed in the context of a brief reflection on what the implementation of AfL with fidelity actually demands of teachers, with reference to some key systemic issues that typically contribute to, rather than alleviate, classroom-based, teacher-specific challenges.

Zoning in on the technical core

In an effort to avoid any confusion arising from the interchangeable use of the terms formative assessment and assessment for learning (AfL) in this paper, the following understanding, generated from the Third International Conference on Assessment for Learning that took place in New Zealand in 2009, is assumed:

> AfL is part of everyday practice by students, teachers, and peers that seeks, reflects upon, and responds to information from dialogue, demonstration, and observation in ways that enhance ongoing learning. (Klenowski 2009, 264)

This *second generation* definition of AfL, as it is now termed, foregrounds the classroom as the *technical core* (Elmore 2000), and students and teachers as the key agents of educational change. In so doing, it refocuses attention on the privacy of classroom life and

on the 'black box' referred to originally by Black and Wiliam (1998). Perhaps, crucially, it reaffirms that:

> Detailed decisions about what should be taught at any given time, how it should be taught, what students should be expected to learn at any given time…and perhaps most importantly, how their learning should be evaluated - resides in individual classrooms, not in the organizations that surround them. (Elmore 2000, 5-6)

Hence, by adopting this definition, we are immediately challenged to try to identify and address whatever obstacles are likely to prevent and/or dissuade teachers – and, by default, students - from embracing AfL because, as Thompson and Goe (2006, 1) note, 'each teacher has to independently get it (i.e., AfL) and do it right'. This begs the question, what exactly does it mean to do AfL right?

From routine to adaptive expertise

> An expert in A*f*L is able to rapidly note essential details of the complex social and psychological situation of a lesson (especially the state of student learning), while disregarding distracting yet nonessential details. The expert teacher is then able to swiftly compare that situation with the intended goals for the lesson, the teacher's knowledge of the content being taught, the teacher's developmental knowledge of students in general and of these students in particular, and other relevant schema. Guided by the results of these comparisons, the teacher then selects the next instructional moves from a wide array of options—most well-rehearsed, some less familiar, and some invented on the spot—such that these next steps address the students' immediate learning needs in real time. (Thompson and Goe 2006, 13)

In reality, it is only when one begins to unpack Thompson and Goe's (2006) description that the enormity of the challenges facing teachers - based on our expectations of them to embrace the *spirit* of AfL - really emerges. What is made explicit here is that pedagogical and content knowledge, although always necessary (and frequently, though not always justifiably, assumed, [Bennett 2011]), is no longer sufficient to bring about the kinds of changes envisaged by researchers. Alternatively phrased, the focus has quietly shifted from routine to adaptive expertise – AfL requires teachers to routinely display both efficiency and innovation in their teaching (Schwartz, Bransford, and Sears 2005). As explained elsewhere (Lysaght, forthcoming 2012):

> Innovation skills represent the knowledge-building skills associated with expanding knowledge in pursuit of solutions to novel problems and … efficiency skills…

represent the collection of knowledge and experiences (schemas) experts leverage for solving routine problems quickly and efficiently.

The term *an expert in AfL*, as employed by Thompson and Goe (2006), then might more accurately be rephrased as *an adaptive expert in AfL*, given that the habits of mind, attitudes, ways of thinking, and organising knowledge described differ fundamentally from those attributed to a teacher displaying *routine expertise* (Bransford 2001).

In this context, the finding that 'teachers are better at drawing reasonable inferences about student levels of understanding from assessment information than they are at deciding the next instructional steps' (Heritage, Kim, Vendlinski and Herman 2009, 24) is hardly surprising given (a) what is actually involved in drawing such inferences and decisions and (b) the *apprenticeship of observation* (Lortie 1975) to which teachers are exposed.

Potential challenges emerging from teachers' apprenticeships of observation

Arguably, one of the greatest challenges to emerge from teachers' apprenticeship years is what Pellegrino (2006, 3) refers to as 'faulty mental models', that is, inappropriate, out-dated yet deeply held beliefs, assumptions and convictions about teaching and learning. Sometimes called habitudes (Flores, Cousin and Diaz 1991), or paradigms (Senge, 2006), such mental models are typically highly resistant to change (Pajares, 1992), not least because they frequently remain hidden, uncontested and unchallenged by the individual and/or his/her peers. And if, as has been argued (e.g., Shepard, 2000), the shift in emphasis from assessment *of* learning to assessment *for* learning, in effect, represents what Kuhn (1970) would have described as a paradigm shift for many teachers, then mental models based on social efficiency, behaviourist learning theories, and scientific measurement will not easily give way to those based on social constructivism and socio-cultural theories of teaching and learning. Such auguries do not bode well for the seamless integration of AfL. Rather, they suggest that teachers may have difficulties in coming to terms with some of the most fundamental strategies of AfL, such as peer- and self-assessment, classroom questioning and discussion.

Such observations challenge us to examine the status and use of AfL in the two formal arenas in which teachers typically develop such mental models: primary and post-primary schools and teacher education colleges. In the process, we are forced to confront the underlying learning theories that shape the teaching, learning and assessment principles and

practices to which teachers (pre- and in-service) are typically exposed (see Lysaght, forthcoming 2012, for exploration of these issues). Black and Wiliam's (2006) caution - which we would argue is as relevant to teachers at third level as it is to those at first and second levels - is noteworthy in this regard:

> The beliefs of teachers about learning, about their role as assessors and about the 'abilities' and prospects of their students, will affect their interpretations of their students' learning and will thereby determine the quality of their formative assessment. (23)

Highlighting the issues of adaptive expertise and teachers' mental models here, albeit briefly and with broad-brush stokes, is at once a comforting and a deeply disconcerting exercise. On the one hand, it helps to explain why AfL has not taken hold as easily or broadly as might have been expected following the publication of more than five seminal reviews (Black and Wiliam 1998; Crooks 1988; Kluger and DeNisi 1996; Natriello 1987; Nyquist 2003) that collectively synthesised in excess of 4,000 research studies over a 40 year period, all attesting to the warranty of AfL. Stated honestly: integrating AfL optimally is a high risk, immeasurably challenging task that demands the routine application in real time of advanced adaptive expertise. As such, it is not something that happens overnight or without very considerable effort, and extended support and professional development. And that is what is so disconcerting about unravelling what is involved in AfL: the enormity of the challenge that presents - principally for teachers, researchers and policy-makers - emerges.

In this paper, we offer no quick fixes to the challenges raised. Rather, we introduce an assessment instrument that is proving useful to researchers and teachers and, in so doing, suggest it as a valid and practical way to begin the process of changing practices which, for all the reasons cited, are notoriously difficult to change.

The development of the Assessment for learning Audit Instrument (AfLAi)

The AfL audit instrument was designed as a mechanism to support schools in conducting site-based reviews of their existing knowledge, skills and practices in formative assessment. Building on audit instruments developed previously by the Association of Assessment Inspectors and Advisors (AAIA) in the UK, and Lysaght (2009) in the Irish context, and informed by the research of Thompson and Wiliam (2007), a draft AfL audit instrument was

developed. As conceived, the instrument was intended for use by teachers in schools to (a) identify individual and collective levels of understanding and use of AfL in teaching and learning and (b) develop plans for site-based teacher professional development.

To-date, the AfLAi has gone through three stages of development. Initially, the researchers consulted with five classroom teachers with knowledge of both assessment instruments (the AAIA and that developed by Lysaght 2009) to identify the relative strengths and weaknesses of each. As a consequence, most of the original items were either rewritten or discarded and a new instrument – the AfLAi – was developed. In the second stage of the project, the AfLAi was administered to 50 teachers who were attending an in-service professional development course in a college of education in Dublin. These teachers were asked to complete the instrument and comment on it. Analysis of the data generated led to further revisions to the AfLAi and the organisation of items across four scales, comprising 58 statements about classroom practices:

- Sharing learning Intentions and Success Criteria (LISC; 16 statements/items);
- Questioning and Classroom Discussion (QCD; 16 statements/items);
- Feedback (FB; 12 statements/items);
- Peer- and Self-Assessment (PSA; 14 statements/items).

From this point on, the acronyms LISC, QCD, FB and PSA are used to signify the four scales of the AfLAi. Teachers responding to the statements within each were asked to report the extent to which the statements reflected their current classroom practice using the following rating scale:

- Embedded   (Happens 90% of the time)
- Established   (Happens 75% of the time)
- Emerging   (Happens 50% of the time)
- Sporadic   (Happens 25% of the time)
- Never   (Never happens)
- Don't Understand   (I don't understand what the statement means).

Study design

The data presented in this section of the paper derive from the third stage of the project which involved the administration of the AfLAi between January 2011 and June 2012 to a

convenience sample of 476 teachers working in a range of classrooms across 36 primary schools in the Republic of Ireland. Participants in the study included schools in which individuals known to the researchers worked and schools whose principals responded to an invitation to participate issued in an article about the project published in a teacher's journal (Lysaght 2010). The aim was to secure enough data to evaluate the psychometric properties of the four scales, to provide baseline data on the AfL practices of a reasonably large (though non-representative) sample of Irish primary teachers and to ensure that the instrument could be used as a means of auditing teachers' AfL practices.

Most of the schools involved in the study were situated in Dublin and its surrounding counties (with the exception of three schools from Belfast, Cork, and Mayo). A range of school types was included in the sample - single and mixed gender, disadvantaged and non-disadvantaged, urban and rural, and Gaelscoileanna. While the sample of respondents cannot be regarded as representative, in general its characteristics reflect the population of teachers in Ireland. Most respondents were female (89%) and most worked as mainstream classroom teachers or teaching principals (70%). Approximately 30% of the respondents indicated that they worked as special educators providing supplementary teaching for students with special educational needs, either on a within-class or withdrawal basis, a figure slightly greater than the population value of approximately 25% (Department for Education and Skills 2011; National Council for Special Education 2011). It should be noted in this context that subsequent analyses revealed no significant difference between the AfL practices of mainstream and special educators. The sample was distributed almost equally between those who were teaching for five years or less (37%), those who were teaching for between six and twenty years (33%), and those who had more than 20 years teaching experience (30%). Finally, approximately equal numbers of respondents taught at each class level from Junior Infants to Sixth Class.

Psychometic properties of the four AfLAi scales

In order to examine some of the psychometric properties of the four scales of the AfLAi, separate principal components factor analyses were run on the data for each scale using SPSS software. For quantitative analyses purposes, each of the scale points was given a numeric value from 6 to 1. So for example, an *embedded* practice was given a score of 6, an *established* practice a score of 5 and so on. Results from the Kaiser-Meyer-Olkin test (none

statistically significant), and Bartlett's Test of Sphericity (all statistically significant), supported the application of factor analysis to all four scales.

Insert Figure 1 about here

The factor analysis of each scale and inspections of scree plots revealed the presence of one large factor with Eigenvalues ranging in size from 4.4 to 7.1. Indeed, it is clear from Figure 1 that for all four scales, the point at which the curve begins to straighten (scree test criterion) occurs at the point of the second factor where too large a proportion of unique variance makes this, and subsequent factors, unacceptable. As the data in Table 1 show, the analyses also revealed that the proportion of variance explained by the first factor in each scale was large in each case.

Insert Table 1 about here

All items in each scale had factor loadings of 0.49 or above on the first component/factor, with average factor loadings of 0.67, 0.62, 0.60 and 0.62 for the four scales LISC, QCD, FB and PSA, respectively. It should be noted that, with a small number of exceptions, all items within each scale loaded more strongly on the first factor than on any subsequent factor. For example, the loadings on the LISC scale could not be used to make a case that learning intentions and success criteria were separate meaningful factors. In addition, the Cronbach's alpha reliabilities reported in Table 1 can be considered very satisfactory. There was no instance of any case where removing an item from any scale would improve the overall reliability measure for that scale. Finally, factor and Rasch analyses reported in another paper dealing with the construction of a 20-item reduced form of the AfLAi provide evidence of a strong psychometric link between the four scales (O'Leary, Lysaght and Ludlow in press). Taking all these data into account, it was concluded that the underlying structure of the relationships between the items/statements within each of the four scales was coherent and that the scales were interpretable in terms of the theoretical framework used to construct the instrument originally

Findings from the administration of the AfLAi in 36 schools

Findings from this stage of the study are now presented in separate tables for each scale. While the statements are numbered according to how they appeared in the original instrument, they are rank ordered in the tables starting with the AfL practices reported by teachers as being most *embedded/established* in their classrooms. It should be noted that the scale used implies that the closer the mean rating is to 6, the more *embedded* the practice. Smaller mean ratings signify that the practice is either *sporadic* or *never happens* (closer to 1). It should also be noted that since all teachers did not respond to every statement across the four scales, the number of responses (N) recorded in the tables is less than 476.

Insert Table 2 about here

Table 2 contains the findings for the learning intentions/success criteria strategy (LISC). According to the data, teachers use student-friendly language to share learning a lot of the time – the mean in this case being 5.26 - somewhere between *established* and *embedded* on the AfLAi. However, given concerns about the reliability of self-reported data (e.g., Ryan, Gannon-Slater and Culbertson 2012), caution must be exercised when drawing conclusions here as respondents may not always judge and/or report accurately what they do in class. Indeed, as noted in recent Irish research, teachers may be reticent to report any practices which they perceive might reflect negatively on their schools (Lysaght 2009). What has also emerged from researchers working with participant schools is that sharing learning intentions and success criteria frequently pose considerable challenges for teachers, with the extent of use closer to the response of teachers to statement 8 (foot of Table 2) than statement 5 (top of Table 2). That said, there is a consistency between the data pertaining to two very similar statements in the scale (statements 1 and 12) which teachers reported as *emerging* in their practice (means of 4.68 and 4.58, respectively). There is little evidence of widespread use of AfL techniques such as WALT (we are learning to) and WILF (what I'm looking for). The data here suggest *sporadic* use at best (mean = 3.29) and considerable variability in how teachers responded to the statement (standard deviation = 1.52). This may have resulted from the fact that, while the terminology, as well as the approaches, were familiar to some teachers, they were not well understood by many others. Overall, while it can be argued that practices of reminding pupils about the relevance of what they are learning (statement 3) and differentiating success criteria according to pupil needs (statement 9) are more *established* in the classrooms of these respondents, the sense is that most practices associated with this scale

are at best *emerging* – a finding that is consistent with conclusions drawn in the Department of Education and Skills (DES) report on incidental inspections of schools (DES 2010) and the Educational Research Centre's (ERC) report on classroom practices reported as part of the 2009 national assessments of mathematics and English reading (Eivers et al. 2010).

Insert Table 3 about here

Table 3 presents the data on the questioning and classroom discussion (QCD) scale. According to the teachers surveyed, questioning to elicit prior knowledge of students' learning (statement 3), assessment techniques to facilitate classroom discussion (statement 2), and the use of open-ended questions (statement 1), happen quite frequently in their classrooms (means = 5.45, 5.02 and 4.94, respectively). However, techniques that expand the repertoire of approaches to questioning are less common. The mean scores for the last three statements concerning pupils sharing the questioning role (statement 7), taking answers 'round the class' (statement 12), and the use of techniques encouraging questioning of the teacher by the pupils (8), suggest that these events are either just *emerging* or occur somewhat *sporadically* in the classrooms of the respondents. The mean ratings for most other statements, particularly those pertaining to classroom discussion (e.g., the issue of pacing highlighted in statement 10) lie in the band between *emerging* and *established*. The greatest variability in practices was that pertaining to statements 6 and 12.

Few would argue with the assertion that questioning and classroom discussion is a normal part of classroom activity in Irish schools so findings from this study that many techniques associated with them are not more *established* is particularly interesting. The data here point to the tension between the more traditional, teacher-led approaches to assessment (e.g., statement 3) and pupil-led approaches (e.g., statement 7), and also flag the potential difficulty of getting teachers to implement AfL in a way that changes the teacher-pupil relationship and democratises learning – a concern raised in the introductory section of this paper.

Insert Table 4 about here

Data in Table 4 relate to the use of feedback (FB) to guide teaching and learning. The data suggest that the teachers in the study believe that practices related to making a link between feedback and the learning intentions/success criteria (statement 1), the diagnostic use

of teacher-made tests (statement 5), and feedback that specifies the nature of progress made (statement 4), are close to being *established* in their classrooms (mean ratings of 4.82, 4.82 and 4.70, respectively). However, the data also indicate that practices such as students giving information to their parents about their learning and/or teachers providing *closing the-gap feedback* (statement 9), and the involvement of pupils in providing feedback to parents/guardians (statement 7), are not nearly as common (means of 2.96 and 3.64, respectively). These data point once more to the challenges of introducing student-led approaches in classrooms and resonate with the discussion on the impact of teachers' mental models on teaching, learning and assessment. It is especially noteworthy that statement 12, which encapsulates the essence of AfL - pupils are provided with information on their learning on a *minute-by-minute, day-by-day* basis rather than end of week/month/term - is considered by the teachers to be an *emerging* feature of their classrooms. Of note also is the fact that the use of standardised tests to identify strengths and needs in teaching and learning is somewhere between an *emerging* and *established* practice in many of these classrooms (statement 6; mean = 4.56). This statement is also characterised by a good deal of variability in how the teachers responded to it (standard deviation = 1.32) – which one might interpret as resulting from a gradual response to the relatively recent initiatives by bodies such as the Professional Development Service for Teachers (PDST, n.d.) to improve teachers' ability to analyse standardised test scores.

Insert Table 5 about here

With the exception of practices associated with pupils being invited to reflect on their prior learning and on their learning progress and goals (statements 3 and 4), the data in Table 5 highlight that many techniques associated with student peer- and self-assessment (PSA) are reported as being *sporadic* in the classrooms of the teachers in this study (means approaching 3). It is significant that statement 8 - time is set aside during lessons for peer- and self-assessment – has an average rating of just 3.04. Indeed, it is also clear from these data that, even in a context where teachers are self-reporting, peer- and self-assessment practices are not commonplace; the means, even for the more *embedded* practices, are generally lower than the equivalent values for the strategies included in the previous tables. This message is highlighted further in Table 6.

Insert Table 6 about here

This table shows the overall average rating for each of the four scales in rank order beginning with the *most embedded*. In the case of the first three scales, the average ratings suggest that teachers view the three AFL strategies - questioning and classroom discussion, sharing learning intentions, and success criteria and feedback – as *emerging* in their classrooms (an average close to 4 on the scale). However, peer- and self-assessment is reported as being more *sporadic* (an average closer to 3). A repeated measures ANOVA confirmed that mean ratings for the AfL strategies were statistically significantly different (F(3, 987) = 433.06, $P < 0.0005$) and the effect size difference was large (eta squared = 0.57). Post hoc tests, using the Bonferroni correction, revealed that the mean for the PSA scale was statistically significantly different to the means for the other three scales (P< 0.0005 in all cases).

Indeed, in other analyses conducted, when all 58 items from the four scales were rank ordered by mean teacher-ratings, it was found that 10 of the 13 items with ratings of 3.5 or lower came from the peer- and self-assessment scale, suggesting that this particular AfL strategy features most infrequently in the classrooms surveyed. Notably, others have argued that pupils' ability to self-regulate is an important life-long learning skill (Popham 2008; Wiliam, 2011) and, that being so, the expectation must be that techniques such as those listed in the PSA scale should be at the heart of changes in assessment practices in every classroom. Popham (2008, 95), for example, describes the successful implementation of peer- and self-assessment as one of the 'key classroom climate shifts', in the sense that the teacher surrenders the role of 'prime instructional mover' and 'students assume meaningful responsibility for their learning and for the learning of their classmates'. Once again, findings deriving from the use of the AfLAi highlight the considerable challenge that lies ahead in engaging pupils more fully in assessment processes that integrate teaching and learning in a seamless manner. Reviewing the data in this table, in concert with those from the previous four, it is noteworthy that the findings are generally consistent with recent evidence about classroom practice in Ireland (DES 2010; Eivers et al. 2010).

Conclusion

Earlier in the paper, the question was raised: what exactly does it mean to do AfL right? In response, attention was drawn to the complex adaptive expertise required by teachers to implement AfL successfully. In addition, the influence of out-dated mental models on learning and teaching, and the enduring influence of teachers' apprenticeship of observation years, were highlighted as notable barriers to progress. In this context, the AfLAi was advanced as one mechanism that might be used in addressing some of these challenges. Reflecting on the collective findings from the data generated by the instrument and presented here, two conclusions are tentatively drawn. First, the instrument, as designed and developed, is fit for purpose; findings from statistical analyses confirm that the statements and scales capture both the extent of teachers' use of AfL and the degree to which the *letter* and *spirit* of AfL are differentiated in practice. Second, the AfLAi provides an in-depth snapshot of the formative assessment practices of a large cohort of primary teachers in Ireland and, in doing so, adds to existing data from national agencies in a manner that enriches understanding of classroom practice.

Building on the work of this project, the researchers have provided each of the participant schools with disaggregated data from the study and are currently supporting a number of them in using the data as a springboard for school-based, professional development in assessment. This work is the subject of a forthcoming publication.

Reference list

Association of Assessment Inspectors and Advisors. 2004. *Secondary assessment practice: Self evaluation and development materials.* http://www.canterbury.ac.uk/education/protected/spss/docs/assessment-self-evaluation.pdf (accessed March 1, 2012).

Bennett, R.E. 2011. Formative assessment: a critical review. *Assessment in Education: Principles, Policy and Practice* 18, no.1: 5-25.

Black, P., and D. Wiliam, 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5, no. 1: 7-74.

Black, P., and D. Wiliam. 2006. Assessment for learning in the classroom. In *Assessment and Learning*, ed. J. R. Gardner , 9-26. Thousand Oaks, CA: Sage.

Brookhart, S. 2009. Editorial. *Educational Measurement: Issues and Practice* 28, no. 3: 1-4.

Crooks, T. J. 1988. The impact of classroom evaluation practices on students. *Review of Educational Research* 58, 438-481.

Department of Education and Skills. 2010. *Incidental inspection findings 2010. A report on the teaching and learning of English and Mathematics in primary schools.* http://www.education.ie (accessed November 11, 2012).

Eivers, E., S. Close, G. Shiel, A. Clerkin, L. Gileese, and L. Kiniry. 2010. *The 2009 national assessments of Mathematics and Reading.* http://www.erc.ie/ (accessed May 16, 2012).

Elmore, R. 2000. *Building a new structure for school leadership.* http://www.politicalscience.uncc.edu/godwink/PPOL8687/Wk10%20March%2022%20Accountability/Elmore%20Building%20a%20New%20Structure%20for%20Leadership.pdf (accessed November 21, 2012).

Flores, B., P. Cousin, and S. Diaz. 1991. Transforming deficit myths about learning. *Language, and Culture. Language Arts* 68, 369-379.

Heritage, M., J. Kim, T. Vendlinski, and J. Herman. 2009. From evidence to action: A seamless proves in formative assessment? *Educational Measurement: Issues and Practice* 28, no.3: 24-31.

Klenowski, V. 2009. Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice* 16, no. 3: 263–268.

Kluger, A. N., and A. DeNisi. 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119, no. 2: 254-284.

Leahy, S., C. Lyon, M. Thompson, and D. Wiliam, 2005. Classroom assessment: minute-by-minute, day-by-day. *Assessment to Promote Learning* 63, no. 3: 19-24.

Lortie, D.C. 1975. *Schoolteacher*. Chicago: The University of Chicago Press.

Lysaght, Z. 2009. From balkanisation to boundary crossing: Using a teacher learning community to explore the impact of assessment on teaching and learning in a disadvantaged school. Unpublished Ed.D. thesis, St. Patrick's College, Drumcondra, Dublin.

Lysaght, Z. Forthcoming 2012. The professional gold standard: Adaptive expertise through assessment for learning. In *Re-imagining Initial Teacher Education: Perspectives on Transformation*, ed. F. Waldron, J. Smith, M. Fitzpatrick, M. and T. Dooley. Dublin: The Liffey Press.

Lysaght, Z., M. O'Leary, and L. Ludlow. 2012. *The development of an instrument to measure teachers' assessment for learning classroom practices.* Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, April 13.

Nichols, P., J. L. Meyers, and K.S. Burling-Pearson. 2009. A framework for evaluating and planning assessment intended to improve student achievement. *Educational Measurement: Issues and Practice.* 28, no. 3: 14-23.

Natriello, G. 1987. The impact of evaluation processes on students. *Educational Psychologist* 22, no. 2: 155-175.

Nyquist, J. B. 2003. *The benefits of reconstructing feedback as a larger system of formative assessment: A meta-analysis.* Unpublished master's thesis, Vanderbilt University, Nashville, TN.

O'Leary, M., Z. Lysaght, and L. Ludlow. Forthcoming 2012. A measurement instrument to evaluate teachers' assessment for learning classroom practices. *The International Journal of Educational and Psychological Assessment.*

Pajares, M. F. 1992. Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research* 62, no. 3: 307-332.

Pellegrino, J. W. 2006. *Rethinking and redesigning curriculum, instruction and assessment: What research and theory suggests.* Washington, DC: National Center on Education and the Economy.

Professional Development Service for Teachers (PDST). n.d. *Using Excel to support analysis of standardised test scores*.
http://ppds.ie/index.php?option=com_contentandtask=viewandid=59andItemid=276 (accessed November 20, 2012).

Pophan, J. 2008. *Transformative assessment.* Alexandria, VA: ASCD.

Ryan, K., N. Gannon-Slater, and Culbertson, M. J. 2012. Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation* 33, no. 3: 414-430.

Schwartz, D. L., Bransford, J. D., and Sears, D. A. 2005. Efficiency and innovation in transfer. In *Transfer of learning from a modern multidisciplinary perspective*, ed. J. Mestre, 1-52. Greenwich, CT: Information Age Publishing.

Senge, P. 2006. *The fifth discipline: The art and practice of the learning organization. Revised Ed*. USA: DoubleDay.

Shepard, L. 2000. The role of assessment in a learning culture. *Educational Researcher* 29, no. 7: 4-14.

Thompson, M., and L. Goe, 2006. Models for effective and scalable teacher professional development. Paper presented at the annual meeting of the *American Educational Research Association*. San Francisco, CA.

Thompson, M., and D. Wiliam. 2007. Tight but loose: A conceptual framework for scaling up school reforms. Paper presented at the annual meeting of the *American Educational Research Association*. Chicago, IL.

Wiliam, D. 2011. *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Yorke, M. 2003. Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education* 45, no. 4: 477-501.

Table 1. Outcomes for the Reliability and Factor Analyses of the AfLAi Scales

	LISC	QCD	FB	PSA
Number of items	16	16	12	14
Alpha Reliability	0.92	0.89	0.83	0.88
Factor 1				
Eigenvalue	7.2	6.1	4.4	5.5
Percent of variance explained	45.1	38.6	36.6	39.5
Range of factor loadings	0.49 to 0.80	0.56 to 0.69	0.55 to 0.71	0.49 to 0.73
Average of factor loading	0.67	0.62	0.60	0.62

Table 2. Average ratings for the *Sharing Learning Intentions and Success Criteria* scale: Rank ordered by practices that are most and least embedded.

Sharing Learning Intentions and Success Criteria (LISC)	N	Mean	Std. Dev.
5. Child-friendly language is used to share learning intentions with pupils (e.g., "*We are learning to make a good guess (prediction) about what is likely to happen next in the story"*).	472	5.26	0.91
3. Pupils are reminded about the links between what they are learning and the *big learning picture* (e.g., "*We are learning to count money so that when we go shopping we can check our change"*).	471	4.78	1.03
9. Success criteria are differentiated according to pupils' needs (e.g., *the teacher might say, "Everyone must complete parts 1 and 2...; some pupils may complete part 3"*).	471	4.72	1.14
1. Learning intentions are shared with pupils at appropriate times during lessons (e.g., *Halfway through the lesson, the teacher might say*: "*Remember, we are learning to distinguish between 2D and 3D shapes"*).	475	4.68	1.05
12. Pupils are reminded of the learning intentions during lessons.	472	4.58	1.04
10. Samples of work are used to help pupils develop a *nose for quality*.	470	4.54	1.19
4. Pupils are provided with opportunities to internalise learning intentions by, for example, being invited to read them aloud and/or restate them in their own words.	472	4.35	1.27
2. Learning intentions are stated using words that emphasise knowledge, skills, concepts and/or attitudes i.e., what the pupils are learning NOT what they are doing.	454	4.34	1.13
11. Assessment techniques are used to assess pupils' prior learning (e.g., concept mapping…).	465	4.20	1.21
6. Success criteria related to learning intentions are differentiated and shared with pupils.	468	4.06	1.29

Sharing Learning Intentions and Success Criteria (LISC)	N	Mean	Std. Dev.
13. Learning intentions are available throughout lessons in a manner that is accessible and meaningful for all pupils (e.g., *written on the black/whiteboard and/or in pictorial form for junior classes*).	471	4.03	1.41
14. Pupils' progress against key learning intentions is noted and/or recorded as part of lessons	473	3.88	1.23
7. Pupils are involved in identifying success criteria.	470	3.80	1.20
16. Pupils are given responsibility for checking their own learning against the success criteria of lessons.	474	3.44	1.16
15. Pupils demonstrate that they are using learning intentions and/or success criteria while they are working (e.g., *checking their progress against the learning intentions and success criteria for the lesson displayed on the blackboard or flipchart, for example*).	470	3.41	1.20
8. Prompts are used to signal learning intentions and success criteria with pupils (e.g., *using WALTS and WILFs in junior classes*).	459	3.29	1.52

Table 3. Average ratings for the *Questioning and Classroom Discussion* scale: Rank ordered by practices that are most and least embedded.

Questioning and Classroom Discussion (QCD)	N	Mean	Std. Dev.
3. Questions are used to elicit pupils' prior knowledge on a topic.	463	5.44	0.74
2. Assessment techniques are used to facilitate class discussion (e.g., b*rainstorming*).	471	5.03	0.89
1. When planning lessons, key, open-ended questions are identified to ensure that pupils engage actively in lessons (e.g., "*If we put a coat on our snowman in the school yard, do you think the snowman last longer?*").	467	4.94	0.99

4. During lessons, hinge questions are used to determine pupils' progress in lessons (e.g., "*We have been learning to sort 3D shapes that stack and roll. Now, if you were given a choice, would you build a tower with spheres or cubes?*").	469	4.79	1.06
13. Pupils' incorrect responses are used to guide teaching and learning (e.g., *a pupil is asked to explain why he/she gave a particular answer*).	474	4.70	1.07
5. Assessment techniques are used to activate pupils /get them thinking during discussions and/or questioning (e.g., *using think-pair-share or talk partners*).	468	4.44	1.19
10. The pace of class discussions is deliberately slowed down to encourage pupils to think before responding (e.g., *using wait time*).	473	4.44	1.17
14. Pupils are asked to evaluate their peers' responses to questions (e.g., "*Fiona, do you agree with what Regina has said and why?*").	471	4.37	1.10
15. Pupils can explain to others what they are learning (e.g., *if a visitor came to the classroom, pupils could articulate what they are learning in terms that identify the knowledge, skills, concepts and/or attitudes being developed*).	474	4.32	1.04
9. Questioning goes beyond the one right answer style (*where the focus is often on trying to guess the answer in the teacher's mind*) to the use of more open-ended questions that encourage critical thinking.	473	4.31	1.05
6. Assessment techniques are used to encourage all pupils to engage with questions (e.g., *no hands up, names out a hat, etc.*).	469	4.29	1.36
11. Pupils are asked to explore their own ideas with others, using *think-pair-share*, for example.	472	4.09	1.21
16. Pupils are asked to explain why they are undertaking particular tasks (e.g., *the teacher might ask*, "*Why are we completing this worksheet/what are we learning by doing it"?*).	474	3.94	1.11
7. Pupils are encouraged to share the questioning role with the teacher during lessons (e.g., *the teacher routinely invites pupils to question their peers' contributions to discussions*).	472	3.83	1.18
12. Individual answers to questions are supplemented by pupils taking an answer 'round the class so that a selection of responses from the pupils is used to build a better answer.	469	3.81	1.35

8. Assessment techniques are used to encourage questioning of the teacher by pupils (e.g., *using hot-seating or a Post-Its challenge*).	473	3.37	1.15

Table 4. Average ratings for the *Feedback* scale: Rank ordered by practices that are most and least embedded.

Feedback (FB)	N	Mean	Std. Dev.
1. Feedback to pupils is focused on the original learning intention(s) and success criteria (e.g., "*Today we are learning to use punctuation correctly in our writing and you used capital letters and full stop correctly in your story, well done John").*	473	4.82	0.98
5. Teacher-made tests are used diagnostically to identify strengths and needs in teaching and learning (e.g., *identifying common mistakes in the addition of fractions).*	472	4.82	1.04
4. Teachers' praise of pupils' work (e.g., *"that's excellent; well done*"), is deliberately and consistently supplemented with feedback that specifies the nature of the progress made (e.g., *"Well done Kate, this paragraph helps me to visualise the characters in the story because of the adjectives you use*").	473	4.70	1.08
6. Diagnostic information from standardised tests is used to identify strengths and needs in teaching and learning (e.g., *common errors in the comprehension section of the MICRA-T are identified and used in teaching*).	465	4.56	1.32
10. When providing feedback, the teacher goes beyond giving pupils the correct answer and uses a variety of prompts to help them progress (e.g., *scaffolding the pupils by saying*: "*You might need to use some of the new adjectives we learned last week to describe the characters in your story*").	471	4.53	1.04
8. Feedback focuses on one or two specified areas for improvement at any one time (e.g., *in correcting written work, punctuation errors may not be marked if the primary focus of the writing is on the use of adjectives*).	469	4.30	1.20
3. Written feedback on pupils' work goes beyond the use of grades and comments such as "well done" to specify what pupils have achieved and what they need to do next.	467	4.29	1.26
2. Assessment techniques are used during lessons to help the teacher determine how well pupils understand what is being taught (e.g., *thumbs up-thumbs-down and/or two stars and a wish*).	468	4.23	1.27
12. Pupils are provided with information on their learning on a *minute-by-minute, day-by-day* basis rather than end of week/month/term.	472	4.09	1.30

11. In preparing to provide pupils with feedback on their learning, the teacher consults their records of achievement against key learning intentions from previous lessons (e.g., *the teacher reviews a checklist, rating scale, or anecdotal record that s/he has compiled*).	468	3.72	1.29
7. Pupils are involved formally in providing information about their learning to their parents/guardians (e.g., *portfolios or learning logs are taken home).*	471	3.64	1.32
9. *Closing-the-gap*-feedback is used to focus pupils' attention on the next step in their learning.	467	2.96	1.57

Table 5. Average ratings for the *Peer-and Self-Assessment* Scale: Rank ordered by practices that are most and least embedded.

Peer- and Self-Assessment (PSA)	N	Mean	Std. Dev.
3. Lessons on new topics begin with pupils being invited to reflect on their prior learning (e.g., *pupils complete a mind map or concept map or brainstorm a topic*).	464	4.42	1.20
4. Pupils are provided with opportunities to reflect on, and talk about, their learning, progress and goals.	472	3.93	1.16
11. Pupils use each other as resources for learning (e.g., *response/talk partners who comment on each others' work and discuss how it can be improved*).	470	3.59	1.24
5. Pupils assess and comment on each other's work (e.g., *they are taught how to use the success criteria of a lesson to judge another pupil's piece of work*).	470	3.43	1.16
10. When pupils have difficulty in their learning, they are encouraged to draw on a range of self-assessment strategies and techniques to help them overcome the problem (e.g., *they consult with an exemplar on the bulletin board*).	471	3.38	1.26
6. Pupils are encouraged to use a range of assessment techniques to review their own work (e.g., *a rubric, traffic lights, thumbs up/down, two stars and a wish*).	471	3.37	1.19
7. A visual record of pupils' progress is maintained to celebrate pupils' learning and show areas of/for development (e.g., *a bulletin board displaying progression in story writing over a term*).	467	3.18	1.35
9. Assessment techniques are used to create an environment in which pupils can be honest about areas where they are experiencing difficulty (e.g., *talk partners are used to facilitate conversations between pupils about the challenges they face in their learning*).	464	3.08	1.12
8. Time is set aside during lessons to allow for self- and peer-assessment.	467	3.04	0.95

14. Pupils have ready access to exemplar materials showing work at different levels of achievement across a range of subject areas (e.g., *pupils use examples of collage on the Art display board when advising peers on how to improve their work*).	465	2.97	1.11
1. Pupils are given an opportunity to indicate how challenging they anticipate the learning will be at the beginning of a lesson or activity (e.g., *by using traffic lights).*	469	2.95	1.10
13. Pupils use differentiated success criteria to self- and/or peer-assess (e.g., *pupils can distinguish between what must be achieved to be successful on a task and what might be done to gain extra credit*).	466	2.91	1.10
2. Pupils are encouraged to record their progress using, for example, learning logs.	472	2.82	1.00
12. Time is set aside during parent/guardian-teacher meetings for pupils to be involved in reporting on some aspects of their learning (e.g., *pupils select an example of their best work for discussion at the meeting*).	471	2.48	1.03

Table 6. *How the AfLAI Scales Compare.*

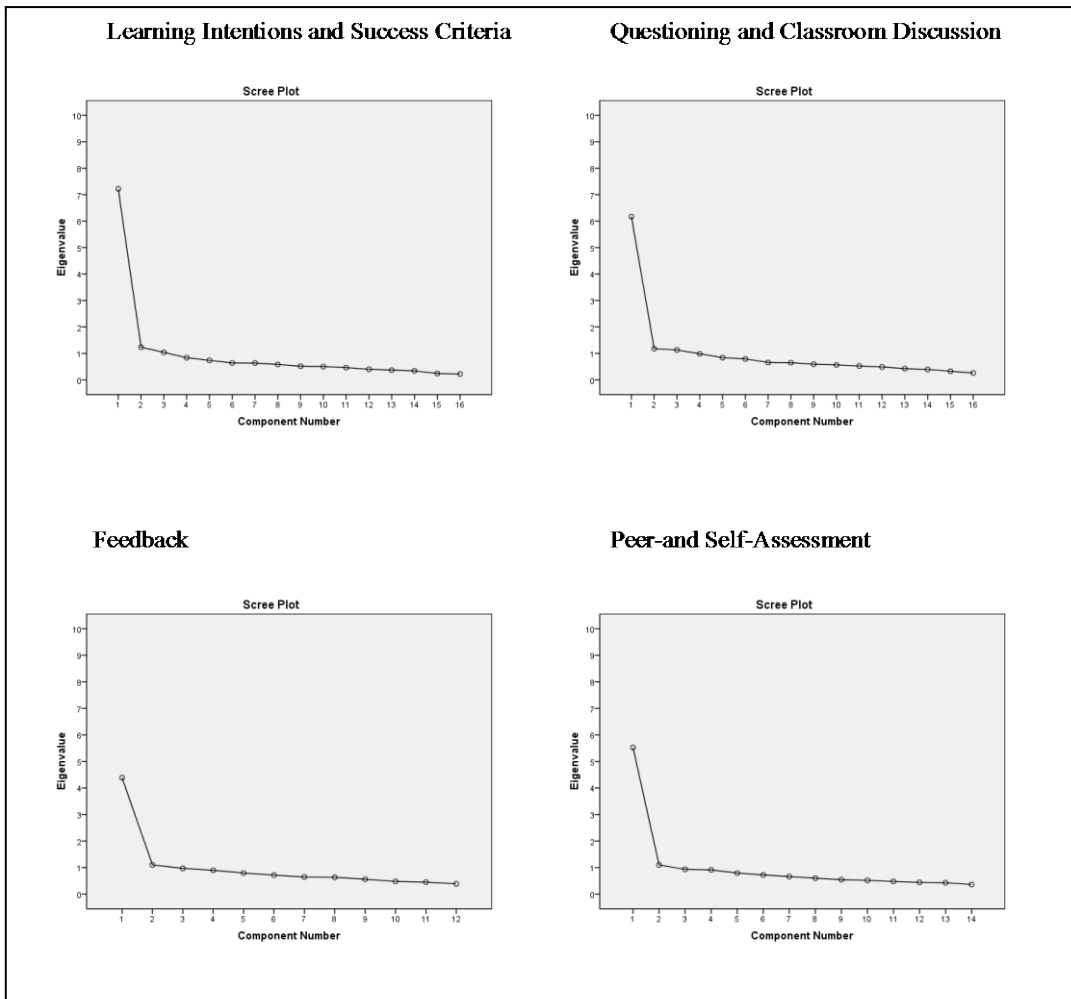|	Mean	Std. Dev	
Questioning and Classroom Discussion (QCD)	4.4	0.68	Emerging
Sharing Learning Intentions and success Criteria (LISC)	4.2	0.80	Emerging
Feedback (FB)	4.2	0.73	Emerging
Peer- and Self-Assessment (PSA)	3.3	0.74	Sporadic

Figure 1. Scree plots from factor analysis of for the four AfLAi scales