The Possibilities and Limitations of Assessment for Learning: Exploring the Theory of Formative Assessment and the Notion of "Closing the Learning Gap"

Ninomiya, Shuichi*

Black and Wiliam (1998a, 1998b) demonstrate that formative assessment is one of the most effective strategies for promoting student learning. Since the publication of their reviews, formative assessment has gained increasing international prominence in both policy and practice.

However, despite this early innovation, the theory and practice of formative assessment are currently at a crossroads. It is widely understood that problems emerge when formative assessment is being reduced to a mini-summative assessment or to a series of teaching techniques for coaching to improve grades and levels.

On the one hand, a serious threat to the effectiveness of formative assessment occurs when it is assimilated into larger accountability systems such as National Curriculum Assessment in England. On the other hand, a defense of formative assessment is offered by some researchers who suggest that the threat stems from misinterpretation of the evolved form of formative assessment.

In this paper, although I am alert to the rich potential of the evolved form of formative assessment, I suggest that the threat stems not from "misrepresentation" but that it exists in the original theoretical framework of Black and Wiliam and the early ARG definition of Assessment for Learning. I will illustrate that this type of formative assessment becomes "convergent assessment" (Torrance & Pryor, 1998) and identify the widespread notion of "closing the learning gap"(Sadler, 1989) as the mechanism of "convergent assessment". I also claim that formative assessment characterised by "convergent assessment" can lead to the practice of "criteria compliance" (Torrance, 2007). Together these claims point towards the theoretical problems of the evolved form of formative assessment and lead to a discussion of the main dilemmas for formative assessment: the kind of learning that is taking place, the effects of explicit learning objectives, the tension of accountability pressure and high-stakes summative assessments. Finally, by applying the suggestions above, this paper provides a critical analysis of recent assessment policy in Japan, emphasizing criterion-referenced approach in classroom assessment and proposes a pathway for developing formative assessment further.

Keywords: formative assessment; Assessment for Learning; Assessment Reform Group; closing the learning gap; criteria compliance

* Wakayama University
  e-mail: nshuichi@center.wakayama-u.ac.jp

1. Introduction

Assessment is vital to the education process. The most visible assessments are summative, measuring what students have learnt and creating accountability for student performance through testing. But assessment also serves a vital formative function through frequent, interactive assessments of student understanding, to identify learning needs and adjust teaching to meet the diversity of the modern classroom.

Over the past two decades, an emphasis on formative assessment has emerged in many parts of the world. In the context of education the term *formative* was introduced by Scriven (1967) to discuss curriculum evaluation and was later extended by Bloom (1969) to the improvement of educational processes including teaching. Whilst several researchers have worked to advance the theory of formative assessment (Sadler, 1989; Torrance, 1993; Torrance & Pryor, 1998), the most popular developments have been made by Black and Wiliam and the Assessment Reform Group (ARG) in UK. The landmark articles by Black and Wiliam (1998a, 1998b) demonstrate that formative assessment is one of the most effective strategies for promoting student learning.

However, despite these early innovations, the theory and practice of formative assessment have arrived at a "crossroads" (Torrance, 2012). For example, Marshall and Drummond (2006) explore the way in which teachers in the UK enact *Assessment for Learning* (AfL) practices, which is a UK programme of formative assessment. They conclude that some practices embody the "spirit" of AfL, whereas others conform merely to the "letter". Their research suggests that whilst the former can help prepare for future learning the latter is merely a tool for judging student performance. In more recent research Torrance (2012) points out the impasse in which formative assessment now finds itself. In spite of theoretical development and justification over many years, the practice "is often limited in terms of its scope and its utilisation of the full range of possible approaches associated with formative assessment". It tends "to involve fairly mechanistic forms of activity" (pp. 1–2).

According to Torrance the emerging problem is that formative assessment is being reduced to regular classroom tests, which are used for monitoring student's progress, or to a series of techniques for coaching to improve grade and test results. Many other researchers (Bennett, 2011; Klenowski, 2009; Hargreaves, 2013; OECD, 2013; Swaffield, 2011) also share this criticism.

They hold the view that the most serious threat to the effectiveness of formative assessment occurs when it is assimilated into larger accountability systems such as National Curriculum Assessment in England. This case involves wide-ranging accountability measures, which have financial and managerial consequences for schools under significant pressure to improve test performance. In such an accountability driven culture significant tensions exist between classroom-based formative assessment and the highly visible summative assessments, which contribute to school performance and accountability measures. As the result of the tensions, formative assessment is often reduced to a mini-summative assessment or to a series of teaching techniques for tests.

On the other hand, a defense of formative assessment is offered by some researchers that suggest that the threat stems from misinterpretation of the evolved form of formative assessment and AfL. This misinterpretation is illustrated in the distinction of the spirit and the letter of AfL in Marshall and Drummond (2006). It is further discussed as the "misunderstanding" or "misrepresentation" of "authentic" AfL in Klenowski(2009), Swaffield(2011) and Wiliam

(in Stewart, 2012). They are disappointed that formative assessment practice is used without a clear understanding of why, and they criticise national policy which leads to the employment of formative assessment in an accountability culture.

In this paper, contrary to the above, although I am alert to the rich potential of AfL, I suggest that the threat stems not from "misunderstanding" or "misrepresentation" but that it exists in the original theoretical framework of Black and Wiliam and the early ARG definition of AfL. I will illustrate that this type of formative assessment becomes "convergent assessment" by making reference to the argument of Torrance and Pryor (1998), and I identify the widespread notion of "closing the learning gap" (Sadler, 1989) as the mechanism of "convergent assessment". I also claim that formative assessment characterised by "convergent assessment" can lead to the practice of "criteria compliance" (Torrance, 2007). Together these claims point towards the theoretical weakness of the evolved form of formative assessment, and lead to a discussion of the main dilemmas for formative assessment: the kind of learning that is taking place, the effects of explicit learning objectives, the tension of accountability pressure and high-stakes summative assessments. Finally, by applying the suggestions above, this paper provides a critical analysis of recent assessment policy in Japan, emphasizing criterion-referenced approaches in classroom assessment and proposes a pathway for further development of formative assessment.

2. Essential Features and Characteristics of an Evolved Form of Formative Assessment

The distinction between summative and formative functions in education was first explored by Scriven (1967) in the context of curriculum evaluation. Bloom (1969) extended the distinction to include classroom teaching. For Bloom, the purpose of *formative evaluation* was "to provide feedback and correctives at each stage in the teaching-learning process" (p. 48), whereas summative evaluation was used to judge what the learner had achieved at the end of a course. Significant works have been undertaken elaborating Bloom's distinction, especially by Sadler (1989), Black and Wiliam (1996) and Harlen (1997, 2012). The essence of Bloom's distinction holds today; however the term 'formative evaluation' is often replaced by that of 'formative assessment', connoting a wider focus on teaching and learning rather than curriculum and programmes.

The main drivers of the evolution of formative assessment since the 1990s have been the Black and Wiliam reviews and the works of ARG in UK. Black and Wiliam (1998a, 1998b) undertake a wide-ranging literature review which "shows conclusively that formative assessment does improve learning," and that gains in student achievement are "amongst the largest ever reported" (1998a, p. 61). They also identify the main features of formative assessment as: sharing criteria with learners, developing classroom talk and questioning, giving appropriate feedback, peer and self-assessment.

Following the research findings of Black and Wiliam, the works of ARG make a great contribution in developing the theory and practice of formative assessment. In 2002 ARG formulated "Assessment for Learning", describing it as "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"(ARG 2002).

This is essential for promoting learner involvement in assessment and responsibility in

learning. ARG (1999) explains the reason why learners must be responsible for their learning:

> A particular point of difference with much present practice is the view of learning that the approach to assessment implies. Current thinking about learning acknowledges that learners must ultimately be responsible for their learning since no one else can do it for them. Thus assessment for learning must involve pupils, so as to provide them with information about how well they are doing and guide their subsequent efforts. (p. 7)

AfL promotes a different relationship between teacher and student than in traditional models of formative assessment. Traditionally, the teacher is responsible for assessment activity, but it is also necessary to take account of the role that learners play. This is demonstrable in the words of AfL definition "the process of seeking and interpreting evidence" (ARG, 2002) to be used by the teacher and student to decide the next learning steps. The student is more involved in assessment activities, sharing learning objectives and success criteria, understanding their achievements and what they need to do to improve their learning through discussion with the teacher. The relationship becomes a collaboration with both student and teacher playing their roles. It describes formative assessment not as an assessment event but as the process of interaction between them.

Stobart (2008) suggests useful distinctions between three kinds of formative response to assessment information, referred to in Allal and Lopez's work (2005). The distinctions, "interactive", "retroactive" and "proactive", help us to identify the essential features of the evolved form of formative assessment from original Bloom's idea. The "proactive" response leads to future changes in teaching. Teachers use assessment information to modify their future teaching in response to evidences from their current students. This would be similar to curriculum evaluation. The "retroactive" is the "test and remediate model" following Bloom and remains dominant in the USA. This is where a formative response is conducted after a phase of teaching, often using a test. It works to identify learning difficulties of students and to improve teaching. The "interactive" response is the focus of those practitioners who follow AfL. This is based on interactions of the learner with the other components in the classroom. The result is shared with students so that they may reflect or improve their learning, through feedback and dialogue with the teacher.

The central difference is whether the focus is on improving teaching or learning. For both the "proactive" and "retroactive" approaches, the teacher is the principal agent as the teaching is adjusted. In contrast the "interactive" approach focuses on student learning. The practices carried out by the teacher (such as giving feedback, clarifying criteria, dialogue with student) can eventually be adopted by the student so that they take on these practices to help to become autonomous learners. In the "interactive" approach, dialogic reflecting and decision-making about learning steps are vital.

The traditional model of formative assessment focuses on teacher improvements of teaching and curriculum. The key function is to provide information for the teacher to judge and reflect on teaching and curriculum. The evolved form of formative assessment aims to improve teaching and learning by focusing on the learning process, particularly on the dialogue between the teacher and student, and it helps the student to become an autonomous learner. It is crucial in the evolved form of formative assessment for the teacher to discuss with students, who reflect on their achievements so that they can identify their next learning steps. These features lead us

to understand that formative assessment occurs within the teaching and learning process rather than subsequently to them. It inspires the recognition that the interrelation between teaching, learning and assessment is not a distinctive event, but is a more integral and symbiotic process than traditional models would suggest.

3. Assessment for Learning and the Assessment for Learning Strategy

In the UK there is an emerging fear that formative assessment practice sometimes leads to mechanistic forms of activity, where the emphasis is on improving performance rather than on learning. This is in contrast to the original ambitions of the pioneers who were engaged in developing the model (Klenowski, 2009; Stewart, 2012). The problem has emerged as AfL has been adopted and rolled out as part of large national initiatives; such has been the case in England with the Assessment for Learning Strategy (AfLS). The strategy states that good assessment for learning makes:

- An accurate assessment – knowing what the standards are, judging pupils' work correctly, and making accurate assessments linked to National Curriculum levels;
- A useful assessment – identifying barriers to pupil progress and using that information to plan and discuss the next steps in learning. (DCSF, 2008, p. 5)

The Strategy describes how "assessment for learning" aims to provide "an accurate" judgment of student work linked to National Curriculum levels and "useful" information to identify student progress and to "plan and discuss" the next learning steps. It includes some familiar features of the 2002 ARG definition by involving students in the assessment process to reflect on their achievements and to seek and decide the next learning steps.

However, some researchers are keen to point out that this is not AfL but rather a mini-summative assessment or a formative use of summative assessment (Swaffield, 2011; Stewart, 2012). Wiliam (in Stewart, 2012) blames the introduction of the national AfLS as contributing to the general confusion about the meaning of AfL. He complains that, "the problem is that government told schools that it was all about monitoring pupil progress; it wasn't about pupils becoming owners of their own learning" and he concludes that it "failed" because it overlooks the "basic ideas that we had been advocating"(in Stewart, 2012).

The AfLS was also criticised by Swaffield (2011), who reiterated that AfL aims to improve student learning, by encouraging their active engagement with assessment as well as learning, discussing and sharing their attainment, and most importantly their growth in becoming more autonomous learners. However, Swaffield asserts that the AfLS fails to understand the essential features of AfL and perpetuates a "distorted view" of AfL which has led to its ideas being "discredited and rejected" (p. 447). As the AfLS describes good assessment for learning as being "accurate" and "reliable" (p. 444), the focus is on properties of summative rather than formative assessment. Moreover, it emphasises monitoring student progress and does not contribute to developing student responsibility. She concludes that "assessment for learning was seen as being about the use of tests" (p. 444) in order to help the teacher judge student progress in comparison with the National Curriculum levels, and she maintains that this is a "misrepresentation" (p. 444).

Wiliam and Swaffield also claim that the AfLS fails to understand essential features and that it is a misinterpretation that leads the teacher to utilise it as summative rather than formative assessment. They contrast the AfLS with "authentic" AfL, which is intended to improve student learning by focusing on active engagement using assessment and student development towards becoming more autonomous learners.

4. The Assessment for Learning Strategy under Scrutiny: Misinterpretation or Inherent Weakness?

The defenses of AfL offered by both Wiliam and Swaffield as a challenge to the AfLS is very persuasive. Their employment of "authentic" AfL coincides with the essential features of the evolved form of formative assessment, which are illustrated above. Their claims point to a clear difference between the AfLS version of assessment for learning and the original version of AfL.

However, such critiques, which reject the AfLS as 'inauthentic', have failed both to account for its foundation in AfL and to accept responsibility for the hidden weaknesses and problems in AfL. Therefore, even though the AfLS may not be considered "authentic" AfL, it is still possible to recognise it as a kind of formative assessment. This is demonstrable in the above quotation drawn from the AfLS, which emphasises not only evaluating student achievements accurately but also discovering and determining their next learning steps through discussion of their progress and weaknesses. This process is the key feature of AfL and it also can be found within the AfLS. I suggest that we have to reconsider AfL in order to analyse why the AfLS tends to become a mini-summative assessment in practice. In my view, the fault is not in practitioner misinterpretation but in the ARG definition of AfL and the original theory of Black and Wiliam (1998a, 1998b, 2009).

The conceptual framework of formative assessment proposed by Torrance and Pryor (1998) that include "convergent assessment" and "divergent assessment" is very useful in diagnosing the hidden problem of AfL. In "convergent assessment" students are involved as recipients. It "aims to discover if the learner knows, understands or can do a predetermined thing" (p. 153). In "divergent assessment" students are also involved as initiators. It also "aims to discover what the learner knows, understands or can do" (p. 153). In applying these concepts we can position the AfLS version of assessment for learning as an example of "convergent assessment", and the "authentic" AfL as an examples of "divergent assessment".

5. Theoretical Issues and Problems of Assessment for Learning: the Notion of "Closing the Learning Gap"

The point at issue is that AfL tends towards "convergent assessment" and opportunities for "divergent assessment" are reduced. Although AfL has a potential to develop forms of both "divergent assessment" and "convergent assessment", the theory has tended towards "convergent assessment".

This issue is apparent in the work of Black and Wiliam. In 2007, Wiliam and Thompson (2007) provided a theoretical grounding for the evolved form of formative assessment, which

relied upon Ramprasad's (1983) process of effective feedback. They proposed it as consisting of three key processes, which are described in the noteworthy work of Black and Wiliam (2009) entitled "Developing the Theory of Formative Assessment". These three key processes are:

- Establishing where the learners are in their learning
- Establishing where they are going
- Establishing what needs to be done to get them there (Wiliam and Thompson, 2007, p. 63).

Feedback is used to refer to information provided by the teacher to the student about their work. However, simply understood, feedback does not involve any mechanism to ensure that the information improves learning. Such a form of feedback might simply be 'telling' the student a score or grade or whether their answer is correct or incorrect.

Ramaprasad (1983) uses an approach borrowed from engineering when he suggests that feedback should not simply give information about current work in comparison with the reference level, but also provide students with the information to improve the quality of subsequent work. Only when information possesses both properties of comparison and steps to improve can it play the role of feedback. Sadler (1989) developed these ideas further and proposed the feedback function as part of a widely accepted model of formative assessment. He points out that it is insufficient for teachers merely to give information about whether answers are right or wrong, and teachers need to give students a task-based comment linked to clear performance standards or attainment targets for improving learning. Sadler identifies three conditions for effective feedback, where "the learner has to (a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action which leads to some closure of the gap" (p. 121).

Sadler proposes that it is vital that the student does not passively receive the feedback but actively engages in these three processes with the feedback from the teacher. This is Sadler's significant contribution to the role of feedback, and through this he extends the notion of "closing the learning gap" within the model of formative assessment.

As represented by Wiliam and Thompson's three key processes and the 2002 ARG definition, AfL has integrated "closing the learning gap" as a conceptual mechanism to improve teaching and learning. Formative assessment with the notion of "closing the learning gap" becomes "convergent assessment". This is because the mechanism to improve teaching and learning is theorised by monitoring student progress in comparison with the reference level and informing what the student needs to get there.

The point at issue is that, when it is applied to practice, the three key processes of deciding "where the learners are in their learning, where they are going and what needs to be done to get them there" are very often interpreted as: what level is the student, what is the next step in the sequence in order to remedy where they fall short, and how to identify the next target. Thus, we are able to see that, in this case, success in learning is identified as attaining the target, and the teacher's role is to encourage the student to master small and detailed aspects of the curriculum. As a result, the process of formative assessment becomes increasingly mechanistic: judging student performance in relation to the target and telling the student how to reach the target rather than by discovering what they are learning.

The interactive process, as the key feature of AfL, tends to reduce decoding the learning

objectives and success criteria into phrases that the student can understand or small steps which the student can achieve, and neglects the discovery and construction of knowledge through exploring subject contents. The student is allowed only procedural autonomy, in which they are encouraged to understand their attainment level and the next steps they are expected to take. This practice is in danger of removing the challenge of learning, the development of student autonomy, and becomes a series of coaching techniques by which to meet targets and to improve academic levels.

Torrance (2007) has coined the precise term "criteria compliance" (p. 282) to capture these practical problems. He suggests that "criteria compliance" has come to replace the teaching and learning experience, because assessment procedures have begun to completely dominate the school setting. Formative assessment as underpinned by the mechanism of "closing the learning gap" becomes "convergent assessment" and therefore reveals the problems and weaknesses inherent in the theory.

6. A Critical Analysis of the Criterion-referenced Approach After the 2001 Assessment Policy Change in Japan

Despite having different traditions and approaches to assessment, in Japan educators are faced with a similar problem of "criteria compliance" as in England. This is caused by a combination of formative and criterion-referenced assessments, which are promoted by recent education policies.

From the end of the Second World War to the late 20th century, a norm-referenced approach had been dominant in Japan, but this began to shift from the 1970s to a criterion-referenced approach. The 2001 Japanese educational assessment policy proposed that teacher's summative assessment should be based on criterion-referenced assessment (MEXT, 2001). This recommendation was the final move of thirty years of gradual change, and it completed the shift to a nationally implemented criterion-referenced approach. Notwithstanding some minor updates in 2010, the policy remains largely unchanged since its inception in 2001.

However, the 2001 policy had a significant weakness. It failed to link teaching and learning to integrated classroom assessment, suggesting that the only assessment of student performance was in reference to the attainment criteria of the National Curriculum. As a result, teaching and assessment did not work in conjunction with each other and the function of assessment had become summative rather than formative.

In order to compensate for this weakness and make assessment formative, the Japanese government has recently emphasised the importance of the integration of teaching and assessment. For example the Basic Plan for the Promotion of Education in 2008 (MEXT, 2008) and the 2010 report of the Central Council for Education (2010) recommended that assessment in each school should be implemented within the PDCA (Plan-Do-Check-Act) cycle in order to improve the school curriculum as well as classroom teaching.

The PDCA cycle is well utilised in management methodology. It is a management tool used for the control and continuous improvement of processes and outcomes. In respect to classroom practice, it means improving teaching and learning process by reviewing targets or goals that are expected as the outcome in the National Curriculum. Criterion-referenced assessment plays a core role in the PDCA cycle. Because the starting point of the PDCA is to set clear objectives

and targets, the clearer the description of the objectives and targets, the easier the PDCA cycle is implemented by reviewing them.

These policies demonstrate the view that assessment becomes formative only when it is integrated in the PDCA cycle and can provide information for the teacher to improve teaching. Therefore, the aim of formative assessment in Japan is interpreted as 'the integration of teaching and assessment'. It is widely accepted that the PDCA cycle and 'the integration of teaching and assessment' are key parts of formative assessment. For example, Kajita (2010) describes this PDCA assessment cycle as "setting objectives, the educational activities to achieve the objectives, criterion-referenced assessment, and the feedback assessment results have on teaching and learning" (p. 17, translated from Japanese by author). This improvement cycle is the generally accepted model of formative assessment in classroom practice. According to a survey about the consciousness of instruction and assessment, about 80% of teachers are positive toward criterion-referenced assessment, because they are able to focus more of their attention on individual student progress compared with norm-referenced assessment (MEXT, 2009).

However, this is not a wholly positive picture, because the implementation of criterion-referenced assessment and the PDCA cycle has led to increasing government control over educational aims and objectives. The objectives in the National Curriculum have been changed from general guidelines to targets that are expected and measured. Schools and teachers have thus been incorporated into the target management system as implemented by the criterion-referenced approach.

Some researchers have recognised the risks inherent in the criterion-referenced approach. Sato (2000) is highly critical of it when he states that criterion-referenced assessment "risks destroying the flexibility, creativity and diversity of the practice" (p. 17, translated from Japanese by author). Tanaka (2002) also points out the significant tensions existing between the setting of attainment targets and the quality of teaching and learning. He summarises research findings on "evaluation for achievement" in Japan, which develops Bloom's idea from the 1970s, and suggests that when the success of learning is identified as attaining targets, "all educational encouragement and learning are reduced to techniques, and assessment is also reduced to the function of checking whether attainment targets are achieved or not" (p. 73, translated from Japanese by author).

Regardless of this criticism, formative assessment underpinned by the criterion-referenced approach has been carried forward and strengthened throughout all educational processes over the last decade. As a result, there has been a greater tendency towards "criteria compliance" in Japan.

The notion of processing used by the criterion-referenced approach is similar to the idea of Assessment for Learning that has been applied in the UK. Although both approaches emphasise the idea that assessment should be formative, they are intrinsically committed to "convergent assessment", which is underpinned by the mechanism of "closing the learning gap". As far as targets and goals are presented as fixed objectives, criterion-referenced assessment and the PDCA cycle are implemented for improving teaching and learning processes by reviewing targets or goals that are expected as the outcome. No matter what targets are set by the criteria or standards, this practice serves a similar function to "closing the learning gap". Consequently, teaching and learning tend to be reduced to a series of techniques used to meet targets. Therefore, there remains a risk of "criteria compliance" and a danger of eliminating the flexibility, creativity and diversity of learning.

The conceptual framework of formative assessment of Japan is still dominated by "proactive" and "retroactive" approaches. The focus is on teachers' improving their teaching and curriculum by using assessment results, rather than students' learning process. In order to move beyond the mechanism of "closing the learning gap", we need to explore the possibilities of an "interactive" approach. It is commonly assumed in arguments supporting the evolved form of formative assessment that it is vital to promote student engagement in the assessment process and to develop their responsibility and autonomy. Without this emphasis on student engagement and responsibility, learning often becomes superficial and outcomes are reduced.

7. Conclusion

Since the publication of Black and Wiliam's influential reviews, formative assessment has gained increasing international prominence in both policy and practice. Its proliferation has been accompanied by different interpretations of formative assessment and further theoretical developments have ensued. Subsequently, the theory and practice of formative assessment are at a crossroads and ripe for examination.

It is clear that the definition of formative assessment has been developed by focusing on the learning process. This has been called Assessment for Learning in the UK and is defined by interactions between teacher and students in the classroom as well as by placing more emphasis on the responsibility of the learner. As some scholars insist, this conceptual development leads us to the recognition that formative assessment is characterised by "interactive" and dialogical process, and by an emphasis on promoting student responsibility and autonomy in their own learning.

Nevertheless, the theory of formative assessment has also contributed to the notion of "closing the learning gap", which possesses a significant mechanism to improve teaching and learning by monitoring student progress in comparison with a reference levels and continuous feedback. Formative assessment understood in this context becomes "convergent assessment". Perhaps because of its easy alignment with accountability and performance measures, this form of formative assessment has become a prevalent feature of educational policy around the world.

The issue is that formative assessment as "convergent assessment", tends to cause alienated learning and assessment practices. Such an educational system is concerned not with learning as much as with meeting targets, and teaching is reduced to a series of techniques for coaching and practice through which to meet targets and to improve proficiency levels. Formative assessment becomes distorted by mechanistic activities to close the learning gap in order to achieve targets and thus, "criteria compliance" comes to replace teaching and learning. I have shown how this issue is faced in countries as culturally and politically diverse as England and Japan, which despite their different traditions and approaches to school evaluation and assessment fall into this contemporary problem.

In order to develop formative assessment further, it is necessary to explore an alternative theory to "closing the learning gap". Some alternative possibilities can be found in the definition of AfL by ARG and in the "authentic" AfL of Swaffield, which in her application tend towards "divergent assessment" rather than "convergent assessment".

In contrast to the notion of "closing the learning gap", Swaffield is concerned not with feedback from the teacher that is used to move student learning towards educational targets, but

with the interactive process between the teacher and student to promote ongoing learning. For instance, the former emphasises promoting student understanding of pre-determined objectives, but the latter emphasises re-constructing the objectives in collaboration with students. The latter underlines the importance of involving the student in an exploratory and dialogical process.

Recently researchers have demonstrated and elaborated on this idea as a further development of formative assessment (Crossouaurd, 2011; Dann, 2014; Hargreaves, 2013; Sadler, 2009). They suggest that the formative assessment process is not simple and mechanistic, such as those assumed in the notion of "closing the learning gap", but more complex and dynamic. Sadler (1989, 2009) stresses that task quality and assessment criteria cannot be learned "by precept" (1989, p. 135) and students can understand these through practical example, in which teacher and student discuss their work critically. Crossouaurd (2011) also suggests examining the complexity associated with the way assessment criteria are understood and brought into classroom discussions. Ultimately, an evolved form of formative assessment should be developed by focusing on complex and dynamic dialogical processes in the classroom.

These suggestions of the complexity and dynamics provide a starting point for exploring formative assessment that is more "divergent" while also offering a different perspective on "convergent assessment". As Sadler (2009) and Crossouaurd (2011) point out, there are limits to what can be learned through giving and receiving preset criteria for assessment. The student can understand these criteria through a dialogical process with multiple exemplars, in which teacher and student discuss and reflect critically together. In short, the understanding of educational standards and criteria cannot be accomplished without such a dialogical process. Therefore the theory of "convergent assessment" should be reconsidered from the perspective of involving the student in an exploratory and dialogical process. This is important because many countries have been pursuing what is now often termed standard-referenced assessment rather than criterion-referenced assessment as the framework of "convergent assessment" after PISA programme. The standard-referenced approach demands a qualitative judgment on the part of the teacher and the student in contrast with dichotomous criterion such as correctness. The exploratory and dialogical process promotes the understanding of task quality and success criteria and confirms the qualitative judgment.

This more evolved form of formative assessment has a rich potential to improve classroom practice and wider educational outcomes, but it needs further development with particular regard to the key concepts of learner's autonomy and the promotion of dialogical process with others. It is my firm belief that rather than relying on "closing the learning gap" and meeting targets, we need to develop formative assessment by focusing on exploratory and dialogical processes through the vitality of authentic relationships between teachers and students in the classroom.

References
Allal, L. & Lopez, L.M. (2005). Formative assessment of learning: A review of publication in French. In Organisation for Economic Co-operation and Development (OECD), *Formative Assessment: Improving Learning in Secondary Classrooms,* Paris: OECD publishing.

Assessment Reform Group (ARG) (1999). *Assessment for Learning: Beyond the Black Box*: University of Cambridge School of Education.

Assessment Reform Group (ARG) (2002). *Assessment for Learning: 10 principles.*

Bennett, R. (2011). Formative assessment: A critical review, *Assessment in Education*, 18(1).

Black, P. & Wiliam, D. (1996). Meaning and Consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5).

Black, P. & Wiliam, D. (1998a). Assessment and Classroom Learning, *Assessment in Education*, 5(1).

Black, P. & Wiliam, D. (1998b). Inside the Black Box: Raising Standards Through Classroom Assessment*, PHI DELTA KAPPAN,* 80(2).

Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment, *Educational Assessment, Evaluation and Accountability*, 21.

Bloom, B.S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: New roles, new means: the 68th yearbook of the National Society for the Study of Education (part II)*, 68(2). Chicago, IL: University of Chicago Press.

Central Council for Education (2010). *Jidouseito no gakushuuhyouka no arikata ni tsuite* (About the evaluation of the student's learning): Ministry of Education, Culture, Sports, Science and Technology.

Crossouard, B. (2011). Using formative assessment to support complex learning in conditions of social adversity, *Assessment in Education*, 18(1).

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice, *Assessment in Education*, 21(2).

Department for Children, Schools and Families (DCSF) (2008). *The assessment for learning strategy*, Nottingham: DSCF Publication.

Hargreaves, E. (2013). Inquiring into children's experiences of teacher feedback: Reconceptualising Assessment for Learning, *Oxford Review of Education*, 39(2).

Harlen, W. & James, M. (1997). Assessment and Learning: Differences and relationships between formative and summative assessment, *Assessment in Education*, 4(3).

Harlen, W. (2012). On the relationship between assessment for formative and summative purpose. In J. Gardner (Ed.), *Assessment and Learning (second edition)* : SAGE.

Kajita, E. (2010). Tashikana gakuryoku no tameno hyouka no arikata (The assessment for certain academic achievement). In E. Kajita (Ed.), *Tashikana gakuryoku no ikusei to hyouka no arikata* (The promoting and assessment for certain academic achievement): Kaneko syobo.

Klenowski, V. (2009). Assessment for Learning revisited: An Asia-Pacific perspective, *Assessment in Education*, 16(3).

Marshall, B. & Drummond, M.J. (2006). How teachers engage with Assessment for Learning: Lessons from the classroom, *Research Papers in Education*, 21(2).

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2001). *Syougakkou jidou shidouyouroku no kaizen ni tsuite* (The Record of Achievements of the Primary School).

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2008). *Kyouiku shinkou kihon keikaku* (The Basic Plan for the Promotion of Education).

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2009). *Gakushuu shidou to gakushuu hyouka ni tai suru ishiki chousa* (The survey about the consciousness of instruction and assessment): Nihon shisutemu kaihatsu kenkyusyo.

Organisation for Economic Co-operation and Development (OECD) (2013). *Synergies for Better Learning: An International Perspective on Evaluation and Assessment,* Paris: OECD Publishing.

Ramaprasad, A. (1983). On the definition of feedback, *Behavioral Science*, 28.

Sadler, D.R. (1989). Formative assessment and the design of instructional system, *Instructional Science*, 18.

Sadler, D.R. (2009). Indeterminacy in the use of preset criteria for assessment and grading, *Assessment and Evaluation in Higher Education*, 34(2).

Sato, M. (2000). Jissen no juunansei to souzousei to tayousei wo hakai suru kikennsei (The risk of destroying the flexibility, creativity and diversity of the practice). In *Sougou kyouiku gijutsu* (Universal educational technique), 55(13): Shougakkan.

Scriven, M.S. (1967). The methodology of evaluation. In R.W. Tyler, R.M. Gange & M. Scriven (Eds.), *Perspectives of curriculum evaluation,* Chicago: Rand McNally.

Stewart, W. (2012, July 13). Think you've implemented Assessment for Learning? *Times Educational Supplement.*

Stobart, G. (2008). *Testing Time: The uses and abuses of assessment,* London/New York: Routledge.

Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning, *Assessment in Education*, 18(4).

Tanaka, K. (2002). Toutatsudo hyouka no riron to jugyou no kaizen (The theory on evaluation for achievement and the improvement of lesson). In M. Endo & M. Amano (Eds.), *Toutatsudo hyouka no riron to jissen* (The theory and practice of on evaluation for achievement): Shouwadou.

Torrance, H. (1993). Formative Assessment: Some theoretical problems and empirical question, *Cambridge Journal of Education*, 23(3).

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning, *Assessment in Education*, 14(3).

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment, *Oxford Review of Education*, 38.

Torrance, H. & Pryor, J. (1998). *Investigating formative assessment,* Buckingham: Open University.

Wiliam, D. & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In *The future of assessment: shaping teaching and learning*, edited by Dwyer, C.A.D: Bloomington.