



Received: 24 October 2017
Accepted: 18 March 2018
First Published: 26 March 2018

*Corresponding author: Maria A. Vlachou, Curriculum, Pedagogy and Assessment, UCL Institute of Education, London, UK
E-mail: blaxoumaria@yahoo.gr

Reviewing editor:
Sammy King Fai HUI, The Education University of Hong Kong, Hong Kong

Additional information is available at the end of the article

EDUCATIONAL ASSESSMENT & EVALUATION | RESEARCH ARTICLE

Classroom assessment practices in middle school science lessons: A study among Greek science teachers

Maria A. Vlachou^{1*}

Abstract: This study aims to examine the classroom assessment practices of five science teachers, alongside the teachers' own perspectives of them, within the rather poorly investigated Greek educational system. In Greece, student assessment at the level of middle education is based only on teacher-led assessment and not on external exams, as the international assessment paradigms prescribe. The aim of the study is to investigate the different purposes of classroom assessment and the principles of classroom assessment practices used to enhance student learning and report on student assessment in science. The findings of this study reveal that, although classroom assessment practices served both formative and summative purposes, participants focused more on the summative uses, without effectively using the assessment evidence to complete the learning loop, and thus meet the formative assessment requirements. Teachers appeared to use some formative assessment principles which are valuable in promoting student learning, but their approaches were more teacher-directed, while students appeared not to have any role in the assessment process. Results underline the fact that summative assessment has a leading role in classroom practices, even in cases where teachers are responsible for keeping a balance between formative and summative assessments.

ABOUT THE AUTHOR

Maria A. Vlachou is a chemistry teacher at Hellenic-American Educational Foundation in Athens, Greece where she teaches in the International Baccalaureate Diploma Programme. She holds a MA in Educational Assessment from UCL Institute of Education, London and she has undertaken many courses about inclusive education, critical thinking, science in education, teacher training and counselling theories in education.

Her research and academic interests lie in the field of educational assessment and teachers' innovative classroom practices. She is a workshop leader for classroom assessment practices and she has written several articles on educational assessment in Greek and international journals. She is a member of scientific committees of students' conference and she participated in the Scientific Committee of the Greek Chemistry Olympiad Competition in 2016 and 2017.

PUBLIC INTEREST STATEMENT

One of the critical issues in educational systems is the effective integration of formative and summative functions of assessment. Greece is one of the countries where classroom assessment (CA) in middle schools is used for both formative and summative purposes. Learning outcomes are not measured through external exams and are not used for accountability reasons. Therefore, the educational environment seems ideal for developing effective CA practices.

The aim of this study is to explore the CA practices of five science teachers used in middle schools in Athens. It was revealed that participants used a range of CA practices which put strong emphasis on summative uses, and students did not appear to have any role in the assessment process. The findings underline a lack of teacher assessment literacy in the given context and the paramount role of the professional development of teachers to maximise the benefits of assessment processes for students in science classes.

Subjects: Middle School Education; Assessment & Testing; Teaching & Learning

Keywords: classroom assessment practices; science education

1. Introduction

Assessment in science classrooms is directly connected with what counts as science and how learning takes place (Cowie, 2013). The goal of science education is that all students develop as knowledgeable and confident knowers, and they use science as part of their everyday lives now and into the future (Cowie, 2013; Harlen, 2007). Inquiry and argumentation-based teaching approaches are prominent in moving student learning beyond conceptual goals, and both of these approaches can be supported through Classroom Assessment, CA (Cowie, 2013; Pellegrino, 2012).

Science teachers use assessment in the classroom for a multitude of purposes, many of which are contradictory. The two main assessment purposes discussed in this article are to enhance learning—known as formative assessment (FA) and to summarise learning—known as summative assessment (SA). SA in classrooms is usually done at the end of a unit, a chapter or a learning experience and takes the form of tests that include questions based on the syllabus studied during that time. It is almost always a formal process and the results are expressed symbolically, as marks or letter grades identifying the gaps in student learning (Isaacs, Zara, Herbert, Coombs, & Smith, 2013). These assessment results are usually used to certify learning and to report to parents and students about student achievements and progress (Earl, 2013). On the other hand, FA is an alternative perspective to traditional testing in classrooms. FA is purposed as a cycle of events, where evidence and judgments of student achievements are used to indicate the next steps in teaching and learning—in terms of progress towards lesson goals. Feedback to the teacher and students is important for promoting and improving teaching and learning. Moreover, students have a central role in this procedure, as they become responsible for their learning and should take action on this, working together with their teachers (Harlen, 2012).

In the literature, SA is often contrasted to FA because of the different purposes, functions, uses or methods that the two sets of terms appear to have and the different time frames in which they are applied (Harlen, 2012; Newton, 2007; Pryor & Crossouard, 2008). However, as many authors argue, assessment information can be used both formatively and for summative purposes. The distinctions are not so sharp in practice (Black, 2013; Harlen, 2012; Wiliam, 2011). In the words of Wiliam and Black (1996, p. 538), “the terms applied not to the assessments themselves, but to the functions they served”. According to several studies though, it appears that the formative use of assessment is compromised by the dominant role of SA, and it is challenging for teachers to balance these two approaches of assessment when they are responsible for both (Earl, 2013; Gipps, 1994; Harlen, 2012). Nevertheless, in many educational systems around the world, teachers are mainly responsible for student FA and less for student SA, as SA judgments are based on external exams (Looney, 2011).

In the case of Greece, student assessment at the level of middle education is based only on teacher-led assessment and not on external exams. Teachers are responsible for collecting data from student learning and achievements through a variety of CA activities so as to provide data to students and parents giving an overall summary of the students’ learning.

2. Literature review

2.1. Classroom assessment: a combination of formative and summative functions

Classroom assessment is a process in which teachers and students gather evidence of student learning through several assessment practices. This evidence can be used for summative purposes, formatively, or for both. SA and FA can both contribute to student learning, but in vastly different ways.

SA is the predominant form of assessment in schools, as it has a long history in education and is widely accepted by parents and the public (Earl, 2013). Its purpose is to certify learning and report to parents about the students' progress and achievements at school. Such information is highly important, both for pupils, who get to know how their learning is progressing, and for parents, who have an active role in their child's learning (Power & Clark, 2000). Teachers' summative judgements on student work should be valid and reliable, based on clear criteria and performance standards and should reflect what they are taken to mean (Black, 1998). The most commonly used SA practice in classrooms is tests, while teacher judgements on evidence such as homework, lab work, attendance and effort can also be used summatively.

The purpose of FA is to enhance the pupils' learning and motivation (James & Lewis, 2012). Assessment happens in the middle of learning and is used to adjust instruction, teaching and learning to better meet student needs (Shepard, 2007; Wiliam, 2011). Teachers, learners and peers share roles and power in this procedure and work together to monitor student progress, evaluating their strengths and weaknesses and designing strategies and practices to promote learning (Bell & Cowie, 2001; Wiliam, 2011). The process of FA follows three main steps: eliciting evidence of learning, interpreting this and acting on it (Bell & Cowie, 2001; Heritage, 2010). FA promotes practices, such as sharing criteria with learners, the effective use of classroom discussion and questioning, feedback and self- and peer-assessment.

As mentioned earlier, even though SA and FA have different purposes, the distinctions are not so sharp in practice. Many examples by scholars in the literature show how summative tests and practices can be used formatively, and vice versa (Black, 2013; Harlen, 2012; Wiliam, 2011). For instance, a test at the end of a unit, when students mark their own work, and reflect on their understanding to see where more work might be needed, can be used as an FA instrument. Teachers can also collect evidence of students' achievements through FA, providing opportunities and then using them for summative judgement (Black, 1998; Harlen, 2005).

However, questions arise in the literature about the validity and reliability of the judgements made when using FA information for summative judgements. As Harlen (2012) argues, evidence gathered from FA practices is often "inconclusive and may be contradictory" (p. 95) and this variation can be problematic for SA, as the assessment criteria should be more specific. Harlen and James (1997) propose that classroom evidence can be used for both summative and formative reasons if a distinction is made between the evidence and its interpretation. The evidence used for SA should be valid and should adequately reflect the learning goals, while the judgements made should be reliable. In terms of FA, the evidence can be interpreted, so that the next steps of the students' learning can be decided upon.

Additionally, there are some implications in the literature that SA and FA are based on different value systems. FA tends to value aspects of learning, such as student motivation and autonomy, while SA focuses on students' achievements against specific success criteria and only motivates students when it has components of FA (Harlen, 2012). Moreover, two of the fundamental principles of FA are to maintain a safe learning environment and to build a strong relationship between teacher and pupil (Willis, 2011). Students should feel comfortable to express their ideas and to take risks in their learning, as well as to feel free to make mistakes and show their lack of understanding to teachers and peers (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Heritage, 2013). However, achieving this in practice is rather challenging, considering the fact that teachers are responsible for both student FA and SA since the relationship between teacher and student can become strained when teachers are seen as judges rather than as facilitators. In addition, students are unwilling to take risks in their learning when they know that they are being judged (Gipps, 1994).

Hence, it is not an easy process for teachers to keep a balance between SA and FA in CA. Teachers should have a deep understanding of the different purposes and functions of these approaches in order to achieve the right balance (Earl, 2013). The major role of CA and assessment in general

should be to contribute to student learning, and not only to evaluate learning (Christoforidou, Kyriakides, Antoniou, & Creemers, 2014). To make this happen, FA should have the main role in CA, while SA could be used only when there are requirements for summative judgments or when students and teachers want to see the final results of their work.

Finally, according to the literature, one of the boundaries of the implementation of the FA principles in CA practices is the lack of understanding among teachers about the theoretical underpinnings of FA and its connection with pedagogy and effective learning (Black & Wiliam, 2012). Moreover, in case that there are high-stakes external exams, the pressure teachers feel to improve student results in them lead teachers to avoid working on unassessed goals and skills that FA requires—such as student autonomy—while the demands of external exams focus more on the proficiency of student knowledge (James, Black, McCormick, & Pedder, 2007; Read & Hurford, 2010).

2.2. Classroom assessment and science subjects

Reviewing the literature on assessment in teaching and learning science, research studies have shown that CA and especially FA can improve students' understanding and learning in science subjects (Black & Harrison, 2004; Ruiz-Primo & Furtak, 2007). Many scholars suggest that questioning, dialogue and feedback play a key role in conceptual development in science education, and therefore CA is seen as a crucial component in this process (Bell & Cowie, 2001; Black & Harrison, 2004). Moreover, as Coffey, Hammer, Levin, and Grant (2011) argue, the role of assessment in science education is to emphasise student reasoning, so that students learn to assess ideas as participants in science, and do not resort to telling the teacher what they think she wants to hear. The purpose of such an approach should be to investigate pupils' ideas and misconceptions and promote their thinking (Black & Wiliam, 1998; Black & Harrison, 2004). Furthermore, science lessons provide many opportunities for classroom talk. Dialogue is vital for both teachers and pupils as they gather and use information about their educational progress (Black & Harrison, 2004; Harlen, 2007; Ruiz-Primo & Furtak, 2007). As Cowie (2013, p. 477) notes: "To sustain discussion teachers need to withhold judgement and encourage students to clarify, compare, challenge and defend their various views using evidence that can also be subject to critique". In addition, student feedback regarding their existing understanding of science can be related to scientifically accepted models, and help them to modify their thinking accordingly, a crucial step for conceptual development in science education and teaching (Bell & Cowie, 2001).

Several studies have been carried out exploring teachers' questioning strategies in science. In a project at King's College, London, teachers from six schools worked on FA practices in their science classrooms. Black and Harrison found that at the beginning of the project many teachers did not identify questions in classroom dialogue as an important instrument of assessment (Black & Harrison, 2001). Over the course of the study, however, teachers focused more on the construction, quality and different functions of their questions, and also on the importance of "wait time". Minstrell and Van Zee (2003) found positive gains in learning when science classroom conversations are led by students. However, asking "rich" questions and encouraging productive dialogue during a lesson does not seem to be an easy process for science teachers, even when they have prepared some of their questions in advance. As Coffey et al. (2011) point out, specific instances of classroom dialogue in the literature illustrate little consideration for the disciplinary substance of students' ideas and argumentation. This is in line with the study of Ruiz-Primo and Furtak (2007), where promoting dialogue and argumentation was not a practice used by participants.

Moreover, Black and Harrison (2004) report on research carried out to look at feedback on written work given by secondary teachers of science. They argue in favour of "comment only" marking, in which teachers give only written comments to students, without giving a grade or mark. These comments should be pertinent and relevant in order to help students move forward in their learning. According to their findings, written feedback takes teachers much longer, but, as they note, this process is worthwhile, as effective comments prompt pupils to move on and develop their thinking and learning (Black & Harrison, 2004).

Considering the aforementioned research, it seems that CA and especially FA practices could have a beneficial role in science education and on student learning in general, but they can be realised to their full potential only when teachers work on assessment principles and have a clear decision in mind about how they will use the information elicited to promote learning, rather than simply reporting on student achievements.

2.3. The Greek assessment policy in middle education

Middle or secondary education in Greece consists of two cycles: compulsory secondary education, which is provided by Gymnasium (age range 12–15); and post-compulsory secondary education (age range 15–18), which is provided by Unified (Eniaio) Lyceum and Technical Vocational Schools. According to the Presidential Decrees on student assessment (Presidential Decree, 2016a, 2016b), teachers assess students using a few assessment practices, thus: day to day oral tests; the pupil's involvement in the learning process; quizzes; and end-of-term tests given without warning during the school year and coursework. There are two terms in a school year and at the end of each term, teachers submit a numeric (0–20) grade for each student's achievements. Moreover, students sit internal written examinations at the end of the school year in which the test items have been designed and are marked by the students' teachers under specific instructions identified in the Presidential Decrees.

Furthermore, according to Greek policy documents about educational assessment (Presidential Decree, 2016a, 2016b), assessment is a process which focuses on identifying student achievement in a systematic and valid way against educational goals and purposes. The main purpose of assessment is the continuous improvement of teaching and the educational system in general, as well as enabling students and educators to be informed about the outcome of their efforts in order to achieve better outcomes. Former Presidential Decrees about student assessment (1994, 2014) described the purposes of both summative and formative classroom assessment, but without making a clear distinction between them. However, the more updated Presidential Decrees (2016a, 2016b) make a distinction between FA and SA, and explain briefly the different functions and purposes they serve.

3. The purpose of the study

Much of the existing literature on CA practices is focused on the effectiveness of FA practices on student learning. However, many scholars argue that the dominant role of high-stakes external exams—where they exist—and the accountability systems of some countries are boundaries that undermine the implementation of effective CA in classrooms (Gardner, 2012; Harlen, 2012; Vlachou, 2015; Willis, 2011). On the other hand, less is known about the implementation and combination of FA and SA practices in science lessons, when the student assessment is only teacher-based and the educational environment seems relieved of boundaries such as those mentioned previously.

However, the Greek educational system seems to favour the development of CA practices that are valuable in promoting student learning, as student assessment in middle schools is based only on teacher-based assessment and not on external high-stakes exams, and is only used for student qualification purposes. Hence, this study attends to investigate the existing CA practices of five science teachers in Athens, Greece, on a day-to-day basis, and the purposes and functions these practices serve, according to the participants' perspectives. Moreover, bearing in mind the existing literature on effective CA, it discusses whether these practices promote student learning in science.

4. Methodology and research methods

The research design of this study was based on a qualitative study, which is a commonly used research approach in education (Creswell, 2012), especially when the research seeks to get closer to teachers' practices and perspectives on CA through detailed interviewing and classroom observations (Norman & Yvonna, 2003). Moreover, qualitative research methods allow the researcher the flexibility not only to focus initially on issues decided upon in advance, but also to take advantage of factors that emerge in the data collection process, as the study progresses (Creswell, 2012).

The participants of this study were five Greek science teachers who worked in schools in Athens and the surrounding region. Teacher 1 was a male physics teacher who had been teaching for more than 22 years in schools. Teacher 2 was also a male physics teacher with 6 years teaching experience in schools. Teacher 3 was a male chemistry teacher, who had been working for 12 years in a school. Teacher 4 was a female biology teacher with 10 years teaching experience in schools. Finally, Teacher 5 was a male chemistry teacher who had been working as a teacher for 4 years. Teachers 1 and 2 held an MA in Science Education, Teachers 3 and 4 hold a PhD in Chemistry and Biology, respectively, and Teacher 5 held an MSc in Chemistry. According to the participants, no one had attended a pre-service or in-service training on Educational Assessment. Moreover, the sampling of the study was opportunistic; hence, the study does not intend to generalise the findings to the wider population (Cohen, Manion, & Morrison, 2011; Robson, 2002; Yin, 2009).

In order to collect direct data and get an idea of the CA practices that science teachers used in their normal classroom teaching, both lesson observations and interviews with teachers were carried out. More specifically, each participant was observed once and the researcher looked directly at the CA practices that teachers applied *in situ*, rather than relying on second-hand accounts. The researcher was a non-participating observer and she kept descriptive notes in a chronological fashion of the assessment activities carried out during the lesson (see also Appendix A). Notes also included the way teachers formed their questions and some dialogue between teachers and students. All the researcher's field notes were used in probes for the interview stage after the observation period (Creswell, 2012).

The observations were of paramount significance in this study, as the researcher not only observed a variety of different CA, but also the timings that teachers used in each practice, the framework used and its action fraction in the lesson. Moreover, interestingly, some of the observed CA practices were not identified as assessment practices by the participants, so they might not have been discussed in the interviews without lesson observations taking place.

Subsequently, semi-structured interviews with the teachers were conducted after the observations. This method is flexible and the researcher has the opportunity to gather in-depth data about the interviewees' ideas and viewpoints on their CA practices (Denscombe, 2010; Dowling & Brown, 2010; Robson, 2002). The interviews were 30–40 min in length and took place a few hours after the lesson observations or, in two cases, the day after. The researcher recorded all the interviews, having first obtained the written permission of the participants.

The interview protocol (see also Appendix B) included two parts with open questions. The first part involved general questions about the role and purpose of CA in science lessons and the second part included questions related to the CA practices used in lesson observations. As this study sought the participants' perspectives about CA practices in general and the role of their own practices in the lessons observed, they were asked to reflect on the purpose, functions, limitations and effectiveness of these practices. At this point, corroborating evidence from the classroom observations was used to shed light on their perspectives. Furthermore, teachers also had the chance to discuss practices that they did not use in the observed lessons, but did use frequently.

Data analysis started at the beginning of the study when the field notes from the observations were analysed to identify what issues to cover in the following interviews, so as to address the purpose of the study. The issues were based on the teachers' CA practices the researcher observed, the timing of each practice and their function in the lesson. The interviews were in Greek and the audio of all interviews was transcribed when the data collection process was complete, each recording being rewound several times to allow the transcript to be completed. Then, the transcripts were read many times, line by line, and analysed to form open codes (Birks & Mills, 2011; Creswell, 2012). Open codes are usually formed when data are repeated in several places in the transcripts, when a participant explicitly states that something is important or when the researcher finds something relevant in a published article about CA practices. Moreover, corroborating evidence from both observations

and interviews was used to document and shed light on the codes. Similar codes were highlighted in the same colour to easily categorise them. Once the data were analysed into open codes, two central categories emerged via axial coding to conceptualise the major principles of the study (Robson, 2002). These categories were, in brief: (1) teachers' conceptions of the purpose and role of CA in science lessons, and (2) the function of CA practices. Moreover, memos were written to illustrate the development of these categories and the inter-relationships between them. Finally, as the field notes and interviews were analysed, quotes representative of the findings were selected, translated into English and edited by a professional translator to ensure that they delivered the same meaning.

With regard to the ethics of this research, the instructions of Ethical Guidelines for Educational Research issued by the British Ethical Research Association (BERA, 2011) and the Statement of Ethical Practice for the British Sociological Association (BSA, 2002) were followed. Based on them, the anonymity of the participants and the schools that they worked for was respected, and the real names of teachers and schools were not used. In the data analysis part, the researcher names the interviewees as "Teacher 1", "Teacher 2" and so forth to protect their identities. Moreover, even though the sampling was mixed gender, all participants are treated as female in this study to simplify all pronoun descriptors to a single word.

Finally, it is essential to discuss the limitations of this study's methodology and research methods. To start with, the small number of participants in this study inevitably limits the diversity of CA practices and conceptions observed among middle school science teachers. In addition, it is not uncommon for a researcher's presence to affect the lesson during observations (Gillham, 2000). In this study, as participants were observed only once, teachers may have changed their teaching habits because of my presence. This potential drawback could have been minimised if the researcher had observed more than one lesson for each participant. However, the study intended to investigate participant perspectives about the purpose and the function of the observed practices in that particular lesson and not to generalise the findings. Moreover, without videotaping the observed lessons, full transcripts of the class dialogues are missing and there were limited ways to validate the researcher's observations. However, the sharing of observational data with the participants had a positive effect on evaluating the reliability of the data. Nevertheless, the findings of this study are at best a preliminary aggregation of practices and conceptions that identify CA practices and how assessment information is used in middle school science classrooms in Greece.

5. Findings

The presentation of the findings is organised according to the categories that emerged from the data analysis. An attempt was made to relate categories and connect ideas, aiming at an overarching conceptualisation of the CA practices and the use of assessment information. The implications of these data will be analysed in the discussion section.

5.1. Teachers' conceptions of the purpose and role of CA in science lessons

This category emerged from participants' responses in the first part of the interview, where general questions were asked about the role and purpose of CA in science lessons, but also about different purposes related to specific CA practices, as revealed later in the second part of the interviews.

To begin with, the data analysis revealed that there was a pattern across participant opinions about the main purpose of CA related to its summative aspect. All participants admitted that they used CA to gather data on student achievement to inform parents during teacher/parent meetings and to assess student performance with a numeric grade at the end of the term. Moreover, it was revealed from the outset of this research that, when participants mentioned CA practices, they mainly referred to practices that had quantitative outcomes and were used for summative reasons. Two teachers explained:

CA is very important for me, as I enable the students to quantitatively understand their weaknesses and strengths and inform their parents. I have to have something measurable. Our field is science and we measure everything, or at least we try. So I try with numbers and expressions to present students' achievements. (Teacher 3)

I use CA to gather grades for teacher/parent meetings and for the students' term assessment. (Teacher 5)

At the same time, the majority of teachers shared the idea that they used CA and assessment outcomes to evaluate the productivity of their teaching and to adjust their teaching plans accordingly. It is clear that teachers see their role in managing learning to be about helping students to understand the learning goal and identifying the gap between where the students are and the learning goal. Three teachers explained the following:

I want to be able to know to what extent the lesson has delivered. That is the most important part of learning, I believe. That is the main reason. [...] I am very interested in that. (Teacher 3)

The lesson is addressed to the students. [...] I should know whether the students have acquired the larger percentage [of knowledge] from what I taught. [I need to know] what percentage [of students] have understood, which of them have understood less and which have not understood anything. (Teacher 5)

CA is helpful as it allows me to understand how efficient I am, to determine my students' level and to adapt my lessons to their level. (Teacher 4)

Moreover, through the data analysis, it appears that when teachers identify the nature of the gap between where their students are in their learning and their learning goals, they adjust their teaching plans to bridge this gap. Teachers may identify the gap using practices that are analysed later in this paper. Most of them said that they covered the topics their students had not understood again, while only Teacher 1 mentioned that she tried to use different teaching methods or practices to ensure that students had understood. Teachers also mentioned feeling pressurised for time when they repeated parts of previous lessons. For example, in the observed lesson, Teacher 5 appeared dissatisfied with the answers she received from the students about the topic that had been covered in the previous lesson and spent a few minutes repeating the topic. She explained this as follows:

I try every time to adjust the lesson to the average students, moving the standard down if I see that there are many students whose understanding and preparation are insufficient. Today, I asked questions of six students and three of them had not understood the content taught, so 50% was enough ... Hence, I repeated things that I had covered in the last lesson [...], [doing] a very fast recap of [it].

Finally, one more pattern across participant opinions was the belief that one of the purposes of CA in their lessons was to assess student preparation for the current lesson. In most cases, students had been given homework to study from their textbook and their notes on the topic are discussed in the subsequent lesson. Teacher 2 explained this:

[CA] is a way to ensure that students kept in mind that they should study frequently and do their homework.

However, this purpose of CA was related only to certain specific CA practices, such as quizzes and oral tests. These practices are analysed later in this paper.

5.2. The function of observed CA practices

This category relates to researcher observations and teacher conceptions about the observed CA practices and practices they usually applied on a daily basis. These practices are clustered in three groups: (1) questioning—oral tests; (2) oral feedback; and (3) quizzes, scores and written feedback.

In each interview, the researcher asked the participants which CA practices they applied frequently and which ones they applied in the observed lesson. As mentioned earlier in this paper, most teachers focused only on practices whose outcomes were valued for summative purposes, such as oral tests and quizzes, and shared their perceptions about the rest of their CA practices only when asked about them. What is remarkable here, though, is that only two out of five teachers did mention their questions during the lesson as a form of CA practice and none mentioned oral feedback as a form of CA, even though these practices were used by most of the participants in the observed lessons.

5.2.1. Questioning—Oral tests

Not surprisingly, the most common CA practice used among the participants was questioning. The data analysis revealed that teachers separate their questions over their lessons into two groups. The first group includes questions which are asked at the beginning of the lesson and whose outcomes are valued more for summative reasons. The second group of questions are asked during the lesson and used more formatively.

As far as the first group of questions is concerned, interviewees responded to them as if it were an “oral test”. The researcher observed this practice in three out of five observed lessons, but all the participants said they used it frequently. At the beginning of the lesson, teachers asked two or three questions to approximately five students about the topic of the previous lesson. Students normally have to study the material taught in the previous lesson for homework. Two teachers out of three who used this practice kept notes about the students’ answers in their mark book because, as they said, students’ achievements in this area counted towards their term grades. The whole practice lasted 5–10 min in each lesson.

According to the majority of teachers, the purpose of this assessment method at the beginning of each lesson was to assess whether students had done their homework. A few teachers stated that through this practice they also assessed the students’ understanding of the topic taught in the previous lesson. Furthermore, interviewees said that this practice connected previous concepts with the current lesson, cultivated students’ metacognitive skills and was helpful for students who had not understood the taught material—or who were absent from the lesson—to learn from their peers. Two interviewees shared the following about this practice:

Students should study the previous lesson using the textbook and their notes. This is the core of their studying. I expect a consistency and continuity of what we have said in the class [previous lesson]. And, through my questions, I also expect them to delve into concepts that are in their notes because these are also questions based on critical thinking. (Teacher 2)

[I use the oral test] to see to what extent the student has studied for this specific lesson, so that s/he doesn’t miss the sequence of the learning process. (Teacher 3)

Concerning teachers’ questions both at the beginning and during the lesson, the question types used were varied. Teachers 3 and 4 employed mainly closed questions—where student responses were just in a few words. Teacher 1 used only open questions—where students had the opportunity to discuss their ideas further, and the other two used both types. Teacher 3 used questions, such as “What do you think about ...?” and Teacher 2 encouraged students to express their thinking in more depth by saying “Why do you say that ...?” and “What do you mean by that?”

Moreover, most expressed that they felt it was important to keep a balance between open and closed questions as teachers cannot check student understanding with closed questions only. Two teachers commented as follows:

There is a question where students should define a concept. I cannot assess students' understanding through this question. So, I come back with a second question: "Why?", "What did you mean by this?" I use these questions to check the students' understanding. Through this practice I understand if students have acquired only unproductive knowledge.
Teacher 2

I want each student to come up with her/his own arguments. I insist very much on the arguments. Even if they are not right, they should have an argument to express. (Teacher 1)

On the other hand, Teacher 5, who used only closed questions, seemed unsure about her practice.

In my view, I think I understand who has studied, who has studied less, who understands and who understands less, even by asking a short, simple question ... because this is the nature of my subject ... I don't know, maybe I am wrong.

Moreover, in the majority of lessons observed by the researcher, there was no waiting time for student responses. A teacher mentioned that this was a limitation of the question and answer procedure, as there is less time for students to think than in written exercises or tests. On the contrary, another teacher who set back-to-back questions stated that: "*I ask questions all the time because I want them to be alert*" (Teacher 3). The lack of waiting time seemed to be connected to teacher frustration in two lessons when there were no hands up to answer the question, so the teachers answered their own questions.

Finally, although in oral tests most teachers did not use the "hands up" practice to choose which students to assess, in their questions, all of the teachers picked students who had raised their hands. One teacher stated that she did not want to "bother shy students", while another teacher said:

I do not ask only the good students that have their hands up. I would also ask someone from whom I do not expect a correct answer, but who raises her hand. I want her to have a role in my class. However, when there is no oral exam, I would respect the fact that someone does not raise the hand; the abstention that she has opted for, because she may be thinking about the question at hand. I am at the stage of scaffolding at the moment. (Teacher 3)

5.2.2. Oral feedback

In the observed lessons, teachers provided oral feedback in a variety of ways. The practice observed showed that most teachers corrected wrong answers—without further investigation of incorrect answers or by asking other students to give the correct answer when a student was struggling. However, Teachers 1 and 2 collected answers from different students and then formed the final answer or asked questions back to the students in order to get them to explain their thinking, if they had given wrong answers. In two lessons, where students had to work on activities both in groups and independently, teachers went to each group or individual, checking the progress of their work. Teacher 1 gave feedback in the form of questions, while Teacher 3 explained what students had to do next.

In terms of the teachers' views on the oral feedback provided, most stated that they tried to make students think more deeply and also gave the students hints on how to move on. Moreover, Teacher 3 said that it was difficult for her to give feedback to student answers at the beginning of the school year because she did not know the learning background of each student at that point and the effort the students had put into it. Furthermore, Teacher 1 pointed out that it was very important for her to investigate the misconceptions behind a wrong answer, so as to provide proper feedback.

When the researcher engaged interviewees in reflecting on the effectiveness of their feedback and to find out whether students acted on it, most could not give a definite answer. One teacher admitted that she was not sure, two said that it depended on whether the students were motivated or not and the rest of them mentioned that they used students' facial expressions to understand this.

In addition, the observational data also suggested that, in three out of four lessons, the teachers praised students and their work. Although the researcher had the feeling that these comments created a friendly atmosphere in the classrooms, two occasions needed further consideration. In Teacher 1's lesson, when students worked in groups, the teacher praised one group for their examples. That comment led to a student asking which group had found those examples, rather than wanting to know about what the good examples were. Moreover, in Teacher 3's lesson, she praised a student who had just finished the first step of an exercise. The student seemed surprised by this positive feedback and responded that she had not done anything yet.

In terms of positive feedback, the majority of interviewees stated that they praised their students only when there was a reason. However, Teacher 3 said that under some circumstances she tried to create reasons for praise, so she lowered the standard. Teacher 1 also said that she tried to praise each student, even for minor things that may be significant to them, while Teacher 5 stated that she did not praise students for things that they were obliged to do, such as to study. Moreover, three mentioned that positive feedback and praise helped them to build a good relationship with students, create a friendly environment in the classroom and motivate students, especially the low achievers. According to them, all this tended to promote learning. However, two appeared to be concerned about the drawbacks of this practice, as individual positive comments can bring out competition among students and some students may stop putting in an effort if praise is given away too freely.

5.2.3. Quizzes

Quizzes were also a very common CA practice mentioned by all of the participants, and observed in two lessons. This practice refers to test papers which include open and/or closed questions which students have to complete in 15 min or less. Participants said that they give students a quiz approximately every two weeks. They noted that both quizzes and oral tests had the same purpose—student SA—and that their outcomes were valued equally. Moreover, they said that they used quizzes when they wanted to assess all of the students and get a general picture of student achievement.

Moreover, according to the observational data, in three out of five visits to lessons, the researcher noticed that students were highly anxious about quizzes, asking their teachers whether or not they would have a quiz in this lesson or the following one. For example, when, at the end of Teacher 2's lesson she said to students that they should “thoroughly revise” certain topics, students started asking if this was because they would have a quiz during the next lesson and if it would be possible to cancel it? The teacher replied that they should not worry.

The researcher questioned two participants about why their students were so anxious about quizzes. Both said that students related quizzes to the final examinations and did not think of them as an alternative form of oral testing. A teacher stated:

I've told them many times that they write quizzes because if I assessed them all by means of an oral test, I would need a whole period. So they write a quiz for a few minutes. Also, they should be more careful with their writing. (Teacher 2)

5.2.4 Scores and feedback on student quizzes

Scoring students on their assessments was a common pattern across interviewees. The majority of the participants held the belief that scores and grades promote student learning only when students understand why they have gained a particular grade. Most teachers pointed out that students should reflect on the mark given by the teacher for each item and notice the weaknesses in their

answers and what is missing, so as to get better marks in subsequent tests. Most of the teachers referred to this as “descriptive scoring”. An interviewee also added:

Students want to find out about their numbers [grades]. But through descriptive scoring, I save myself the nagging and protesting because they see where they have lost marks; where they are weak. So they understand the areas of their weaknesses and why they got that grade. So that method prevents them from comparing their grades to their classmates because they understand that they have achieved the same grade for a different reason. (Teacher 3)

Furthermore, one participant shared a different concept:

Grades are important for students because they all want a better grade, so they study harder at home and try to recall what the teacher said. It helps students to promote their learning because they try harder. They are alert. It is cultivated in students from a very young age to want to get better and better grades. (Teacher 5)

As far as teacher written feedback on student quizzes and tests is concerned, three teachers out of five do not give written feedback. They said that they follow this practice because they either do not have time to write comments on student tests and quizzes or they think that it is pointless, as students do not keep their test papers (most schools do not allow students to take the tests and quizzes home). However, they stated that they spent enough time explaining the test items during the lesson. For instance, Teacher 1 said that she spoke individually with each student, explaining the weaknesses in their answers and asking them to correct their mistakes. Moreover, Teacher 5 said that there was no need to write comments on student work, as most of the questions were closed. Furthermore, the teachers who gave written feedback to students stated that they underlined the weaknesses in student answers, wrote model answers and also wrote individual comments for each student. Teacher 4, who said that she did not often give written feedback, mentioned that she thought that questions about the students’ answers were more important since students could understand what was missing in their answers.

6. Discussion

As Stobart (2008) argues, the best first question to ask in any assessment is “What is the principal purpose of this assessment?” (p. 14). According to this research, the main purpose of the majority of CA practices is summative, while the information elicited by these practices is usually used formatively. Participants appeared to have felt pressure to collect qualitative data on student achievements through assessment practices, to present these data at teacher/parent meetings and to use them to calculate final-term grades. Nevertheless, some of them said that, through the CA practices, they evaluated the extent to which the taught material had been understood, and then adjusted their teaching to bridge the gap between the point where the students were in their learning and the intended learning outcome. This perspective appears to be one of the main principles of FA, as the “spirit” of FA is evidence of the assessment being used to adjust instructions, consequently improving teaching and learning to better meet student needs (Shepard, 2007; Wiliam, 2011).

As Wiliam (2006) argues, FA is only applied when evidence of student learning is used to adapt the teaching to meet student learning needs. If teachers do not use the evidence to adjust their teaching, then they are not using FA. However, as he points out, teachers often say that they gather evidence to take action to help pupils, but he found that the data were never employed and the teaching never changed direction. In this research, participants tended to re-teach the lesson, or part of it, in the same way as before, when they were not satisfied with student outcomes in an assessment. Only one participant appeared to change her teaching practices to achieve student understanding. As Bennett (2011) argues, effective FA—and therefore effective CA—requires quality inferences and instructional adjustments. He notes that teachers should understand the different kinds of errors—such as slips, misconceptions and a lack of understanding—and adopt the proper instructional action. Hence, re-teaching the lesson is not a panacea for all student errors.

With regard to the observed CA practices, participants used practices that are required by Presidential Decrees (1994, 2016a, 2016b), such as quizzes and oral tests. However, the most common CA practice among participants was questioning, but it was not named as such. According to the participants, it was mainly used to make some students, though not all participate in the lesson and scaffold new learning. Participants may not view this practice as assessment, as it was not used for summative purposes. As Bell and Cowie (2001) note, much of the information gathered through informal interactions is not recognised by teachers or students as having a potential assessment function. Moreover, during the data collection, it was difficult to find a pattern of the type of questions that the participants used in their lessons, as one of them used only open questions, two used only closed questions and the rest were mixed. Nevertheless, all of the participants agreed that the open questions were more fruitful in science lessons, except for one who thought that closed questions were more suitable for science, though she appeared not to be sure about this. Black et al. (2003) wrote about the King's College project where teachers working with FA practices realised that they compromised learning by asking simple, closed questions, where recall rather than thinking provided the answers. However, as Earl (2013) and Wiliam (2006) argue, it is challenging and complex for teachers to rethink and change habits and practices that they are used to employing.

Arguably, just one participant noted that she could not elicit evidence of student understanding through asking questions about the terminology of concepts. Instead, she followed up with open questions to assess the students' understanding of the concept. This idea fits with the idea that questioning is an essential component of both CA and science learning, as its purpose is to investigate pupils' ideas, understanding and misconceptions, and to promote thinking (Black & Harrison, 2004). Black and Harrison (2004) argue that, although teachers do sometimes need to set closed questions to assess pupils' knowledge, it is often the case that science questions should ask the pupils to "delve deep into their conceptual learning" (p. 6) and to ask "rich" questions. This idea also complies with the reconceptualisation of science education that puts argumentation, model building, and explanation at the centre of learning science (Cowie, 2013). Moreover, Coffey et al. (2011) argued that an open question "provides teachers with more data and sparks deeper student thinking" (p. 1113). He also points out that questions should seek information from the students' ideas and reasoning, rather than recalling terminology.

Moreover, as noted in the data collection part, the participants did not give enough time to students to think through their answers and, at one extreme case, a participant explained that she asked back-to-back questions to keep students alert in the classroom. Some participants did not engage so much in classroom dialogue, as their feedback was focused more on correcting students' answers or answering their own questions. Black and Wiliam (2010) argue against these practices, saying that they discourage pupils from even trying to think of a response, as they know that the answer will come along in a few seconds from either the teacher or a quick-thinking classmate. Hence, there is no point in them trying, knowing that they cannot respond quickly enough and being unwilling to risk making mistakes in public. In this case, the lesson can keep going, but it will be beyond the understanding of most of the class. On the contrary, as scholars argue, when teachers put the emphasis on dialogue, rather than on the current answers, teachers can achieve a free-flowing exchange of ideas and elicit evidence of learning from more students (Black & Wiliam, 2010; Coffey et al., 2011). It is obvious that, working with a class of more than 20 students, it is difficult for teachers to elicit evidence of learning from each student. However, with classroom dialogue and other practices, such as group activities and group discussions, teachers can meet the learning needs of all students (Bell & Cowie, 2001). Furthermore, these activities operate within a constructivist framework of learning and are important in addressing the problem of only a few students actively participating in the classroom (Pryor & Crossouard, 2008).

As far as the oral tests are concerned, it is revealed that all the participants used this practice to elicit information about the student preparation, namely about whether students had studied at home after the previous lesson. The majority of teachers highlighted that studying at home is an important practice and that they assess it at the beginning of each lesson with quizzes and oral

tests, a common practice among the participants. These practices are also noted in the Greek Presidential Decree (1994), where it is pointed out that the test questions should be relevant to the examples in the students' textbook. However, the purpose of these practices appears to be to assess student memory skills rather than their learning and understanding. According to Bloom's taxonomy (1969), just asking students to recall information does not promote the development of complex mental skills, such as application, analysis, synthesis and evaluation. This purpose appears to be far from the purpose of CA, where the focus is not on recalling information or terms, but on seeking information about the students' understanding and thinking (Black et al., 2003; Coffey et al., 2011). Moreover, in terms of science education, scientific inquiry should not only elicit evidence of a student's understanding of central issues and fundamental concepts in each subject, but also of the student's higher order inquiry skills. These skills can under no circumstances be prompted and assessed by questions that simply ask for a recall of information from previous lessons.

Furthermore, in the observed lessons, a few participants gave oral feedback to students, mainly correcting their mistakes, while others sought further explanations about the students' responses by questioning them back. Arguably, asking questions back to students reflects the principles of CA, as it is a way of eliciting information and scaffolding learners' thinking (Black et al., 2003). However, corrective feedback seems to focus only on the quality of their speaking or work, rather than helping students to identify the nature of their learning gap and provide them with information and guidance on how to close it (Hattie & Timperley, 2007; Heritage, 2010). In other words, student errors should not just be dismissed by correction but should be investigated, as they reveal information about how students learn, and this can be used to improve teaching approaches (Pryor & Crossouard, 2008; Wiliam, 2011). In addition, there is a distance between comments, instructions, corrections and feedback. Teacher comments can be characterised as effective feedback only when the learner acts on them and improves her performance (Hattie & Timperley, 2007; Wiliam, 1998). However, in this study, participants were concerned about the effectiveness of their feedback, as students did not appear to take full advantage of it. Furthermore, many authors argue that, when feedback is given with grades, the latter undermines feedback, as students do not look beyond the numbers (Black & Wiliam, 2010; Brookhart, 2007; Earl, 2013; Lipnevich & Smith, 2009). Grades alone cannot reveal the study needs of students, and therefore asking students to reflect on their grades is ineffective.

In addition, the majority of participants appear to give positive feedback to students to help cultivate a friendly atmosphere in the classroom and to motivate students, especially low achievers. According to the literature, praise is the most common form of feedback, but it has little impact on learning as it focuses on the students' ego, rather than on the task and the learning objectives (Black et al., 2003; Hattie, 2008). It is also related to social comparison with peers (Earl, 2013), as shown by the quotation from an observed lesson when a student wanted to know which group received praise from the teacher, not which task was being praised.

Last but not least, according to the findings, students appeared to have a peripheral role in the CA practices, without taking responsibilities in the learning assessment process through self-assessment and peer-assessment practices. This teacher-centred assessment is a more didactic approach, with the teacher being responsible for transmitting knowledge and assessing the outcomes, while students remain passive recipients (Stobart, 2008). The teachers appear to think that their role is to be responsible for student understanding—and many hold the view that learning occurs when students listen to them. This approach embraces a rather convergent view of assessment, where only teachers gather evidence through assessing what learners know, understand or can do, so as to give quantitative feedback and adjust their subsequent teaching (Pryor & Crossouard, 2008).

These findings are consistent with other research studies in science education which have shown that, even after professional development programmes on FA, teachers find it difficult to include self- and peer-assessment practices in their everyday routines (Jonsson, Lundahl, & Holmgren, 2015; Wylie & Lyon, 2015). This concurs with Black et al. (2003) findings that teachers' beliefs and classroom practices are influenced by what they teach, and specifically by how the taught material is

interpreted in the school curriculum. According to them, comparisons among teachers of English, science and mathematics, show that science and maths teachers adopt a more “delivery”-focused teaching approach, without necessarily ensuring that students learn with understanding, as the focus is more on delivering conceptual models and specific goals.

However, interaction and collaboration between teachers and students, and among students, should be a central feature of CA. Teachers, learners and peers should share roles and power, working together to monitor student progress and to take action to promote learning. Self- and peer-assessment strategies appear to develop metacognitive skills in students, which motivate them and enhance their sense of ownership of their learning (Black & Wiliam, 2010; Black et al., 2003; James et al., 2007). Especially in terms of peer assessment, students appear to understand better through the language of their peers, and peers’ work can act as a stimulus to improving their own work (Clarke, 2005; Isaacs et al., 2013).

7. Conclusion

Considering the previous analysis, it was revealed that participants use a range of CA practices which put strong emphasis on summative uses. Teachers’ understanding of the CA process is that assessment is largely teacher-led and directed, while students do not appear to have any active role in the assessment process. Moreover, it has been revealed that the participants mainly adopted CA practices to gather evidence of student learning and to identify the nature of their learning gap, but they did not provide students with information and guidance on how to close it. Hence, CA is not considered as a multi-purpose process that can extend over time, occurring before, during, and following instruction (Campbell, 2013).

The findings are consistent with recent studies on assessment in science education which reveal that the most challenging aspect of CA is using the assessment evidence to inform the next instructional steps and also activate students in the assessment process (Jonsson et al., 2015; Wylie & Lyon, 2015). To make this happen, a shift of mind-set from CA as a final indicator of achievement to CA as learning along a developmental process of improvement is required. However, this shift appears difficult in practice, even in countries that have large-scale implementation programmes in applying FA principles on CA practices. According to the literature, the barriers mainly constitute three factors: the accountability system of each country; the dominant role of national examinations in many grades of schooling; and the insufficient assessment literacy of teachers (Hayward, 2015; Hopfenbeck, Petour, & Tolo, 2015; Ratnam-Lim & Tan, 2015; Vlachou, 2015).

In the case of Greece though, there is neither an accountability system for teachers related to student outcomes nor an extended focus on national examinations (OECD, 2011). Hence, the main barrier in the implementation of effective CA practices among the participants seems to be the lack of teacher assessment literacy, as there are no courses on Educational Assessment in pre-service programmes for new teachers nor in the professional development continuum programmes run by the Greek Ministry of Education. This underlines the paramount role of the professional development of teachers. Although teachers could develop and improve their assessment practices through their own experiences (Vogt & Tsagari, 2014), multiple studies have explored how continuing professional development for teachers, along with classroom-based professional learning and inquiry to foster effective assessment, are essential for promoting CA practices (DeLuca, Valiquette, Coombs, LaPointe-McEwan, & Luhanga, 2016; Pedder & James, 2012).

As stated above, this study does not aim to generalise the findings to the wider population. However, the findings are at best a preliminary aggregation of practices and conceptions that detect how CA practices are used in middle school science classrooms in Greece. I hope the findings in this paper encourage policy-makers and researchers in Greece and internationally to invest in large-scale policy programmes on FA and CA in Greece, as there is plenty of room to improve and investigate the effects of FA and CA principles on student learning in a framework that differentiates from the international assessment paradigms and their barriers.

Funding

The authors received no direct funding for this research.

Author details

Maria A. Vlachou¹

E-mail: blaxoumaria@yahoo.gr

¹ Curriculum, Pedagogy and Assessment, UCL Institute of Education, London, UK.

Citation information

Cite this article as: Classroom assessment practices in middle school science lessons: A study among Greek science teachers, Maria A. Vlachou, *Cogent Education* (2018), 5: 1455633.

References

- Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht: Kluwer Academic.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. doi:10.1080/0969594X.2010.513678
- BERA. (2011). *Ethical guidelines for educational research*. London: Sage.
- Birks, M., & Mills, J. (2011). *Grounded Theory: A Practical Guide*. London: Sage.
- Black, P. (1998). *Testing: Friend or foe? Theory and practice of assessment and testing*. London: Falmer Press.
- Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 167–178). Thousand Oaks, CA: Sage.
- Black, P., & Harrison, C. (2001). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review*, 82(301), 55–61.
- Black, P., & Harrison, C. (2004). *Science Inside the Black Box: Assessment for Learning in the Science Classroom*. London: nferNelson.
- Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London: King's College London School of Education.
- Black, P., & Wiliam, D. (2010). Kappan classic: Inside the Black Box: Raising standards through classroom assessment: Formative assessment is an essential component of classroom work and can raise student achievement. *Phi Delta Kappan*, 92(1), 81–90. <https://doi.org/10.1177/003172171009200119>
- Black, P., & Wiliam, D. (2012). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and Learning*. London: Sage.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it into practice*. Berkshire: Open University Press.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In H. G. Richey & R. W. Tyler (Eds.), *Educational evaluation: New roles, new means*. Chicago, IL: University of Chicago Press.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice*. New York, NY: Teachers College Press.
- BSA. (2002). *Statement of ethical practice for the British Sociological Association*.
- Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment*. Thousand Oaks, CA: Sage.
- Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation*, 40, 1–11. doi:10.1016/j.stueduc.2013.11.006
- Clarke, S. (2005). *Formative assessment in the secondary classroom*. London: Hodder & Stoughton.
- Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48(10), 1109–1136. doi:10.1002/tea.20440
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. London: Routledge.
- Cowie, B. (2013). Assessment in the science classroom: Priorities, practices, and prospects. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. London: Sage.
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*. doi:10.1080/0969594X.2016.1244514
- Denscombe, M. (2010). *The good research guide: For small-scale social research projects*. New York, NY: McGraw-Hill Education.
- Dowling, P., & Brown, A. (2010). *Doing Research/Reading Research: Re-Interrogating Education*. Oxford: Taylor & Francis.
- Earl, L. M. (2013). *Assessment As learning: Using classroom assessment to maximize student learning*. London: Sage.
- Gardner, J. (2012). Assessment for learning: A compelling conceptualization. In J. Gardner (Ed.), *Assessment and learning*. London: Sage.
- Gillham, B. (2000). *Case Study Research Methods*. London: Bloomsbury Academic.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *Curriculum Journal*, 16(2), 207–223. doi:10.1080/09585170500136093
- Harlen, W. (2007). Formative classroom assessment in science and mathematics. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice*. New York, NY: Teachers College Press.
- Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. N. Gardner (Ed.), *Assessment and learning*. London: Sage.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365–379. doi:10.1080/09585170500136093
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hayward, L. (2015). Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 27–43. doi:10.1080/0969594X.2014.984656
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.
- Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action*. Cambridge, MA: Harvard Education Press.
- Hopfenbeck, T. N., Petour, M. T. F., & Tolo, A. (2015). Balancing tensions in educational policy reforms: Large-scale implementation of Assessment for Learning in Norway. *Assessment in Education: Principles, Policy & Practice*, 22(1), 44–60. <https://doi.org/10.1080/0969594X.2014.996524>
- Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). *Key concepts in educational assessment*. London: Sage. <https://doi.org/10.4135/9781473915077>

- James, M., & Lewis, J. (2012). Assessment in harmony with our understanding of learning: Problems and possibilities. In J. Gardner (Ed.), *Assessment and learning*. London: Sage.
- James, M., Black, P., McCormick, R., & Pedder, D. (2007). Promoting learning how to learn through assessment for learning. In M. James, R. McCormick, P. Black, P. Carmichael, M. J. Drummond, A. Fox, ... R. Procter (Eds.), *Improving learning how to learn: Classrooms, schools and networks*. London: Routledge.
- Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of assessment for learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, 22(1), 104–121. doi:10.1080/0969594X.2014.970612
- Lipnevich, A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology*, 15(4), 319–333. doi:10.1037/a0017841
- Looney, J. W. (2011). *Intergrading formative and summative assessment: Progress toward a seamless system?* OECD Education Working Paper No. 58. <https://doi.org/10.1787/19939019>
- Minstrell, J., & Van Zee, E. H. (2003). Using questioning to assess and foster student thinking. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday Assessment in the Science Classroom* (pp. 61–74). Arlington, VA: National Science Teachers Association Press.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170. doi:10.1080/09695940701478321
- Norman, K. D., & Yvonna, S. L. (2003). Introduction: The discipline and practice of qualitative research. In K. D. Norman & S. L. Yvonna (Eds.), *Collecting and Interpreting Qualitative Materials*. London: Sage.
- OECD. (2011). *Improving efficiency in primary and secondary education, education policy advice for Greece*. Paris: Author. Retrieved from http://www.oecd-ilibrary.org/education/education-policy-advice-for-greece_9789264119581-en
- Pedder, D., & James, M. (2012). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning*. London: Sage.
- Pellegrino, J. W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching*, 49(6), 831–841. doi:10.1002/tea.21032
- Power, S., & Clark, A. (2000). The right to know: Parents, school reports and parents' evenings. *Research Papers in Education*, 15(1), 25–48. doi:10.1080/026715200362934
- Presidential_Decree. (1994). *Students assessment in high school (gymnasium)*. In Greek_Government (Ed.), (Vol. 409). State Printing. Retrieved from <http://www.et.gr/index.php/2013-01-28-14-06-23/search-laws>
- Presidential_Decree. (2014). *Students assessment in unified (eniatio) lyceum*. In Greek_Government (Ed.), 68/2014. State Printing. Retrieved from <http://www.et.gr/index.php/2013-01-28-14-06-23/search-laws>
- Presidential_Decree. (2016a). *Students assessment in high school (gymnasium)*. In Greek_Government (Ed.), (Vol. 211). State Printing.
- Presidential_Decree. (2016b). *Students assessment in unified (eniatio) lyceum*. In Greek_Government (Ed.), (Vol. 74). State Printing.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1–20. doi:10.1080/03054980701476386
- Ratnam-Lim, C., & Tan, K. (2015). Large-scale implementation of formative assessment practices in an examination-oriented culture. *Assessment in Education: Principles, Policy & Practice*, 22(1), 61–78. doi:10.1080/0969594X.2014.1001319
- Read, A., & Hurford, D. (2010). 'I know how to read longer novels' – developing pupils' success criteria in the classroom. *Education 3-13*, 38(1), 87–100. doi:10.1080/03004270903099843
- Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. Hoboken, NJ: Wiley.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Education*, 44(1), 57–84. doi:10.1002/tea.20163
- Shepard, L. A. (2007). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Abingdon: Routledge.
- Vlachou, M. A. (2015). Does assessment for learning work to promote student learning? The England paradigm. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 88(3), 101–107. doi:10.1080/00098655.2015.1032194
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. doi:10.1080/15434303.2014.960046
- William, D. (1998). *The validity of teachers' assessments*.
- William, D. (2006). *Assessment for Learning: Why, what and how*. Retrieved from http://www.dylanwilliam.org/Dylan_Williams_website/Papers.html
- William, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- William, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5), 537–548. doi:10.1080/0141192960220502
- Willis, J. (2011). Affiliation, autonomy and assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 399–415. doi:10.1080/0969594X.2011.604305
- Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, 22(1), 140–160. doi:10.1080/0969594X.2014.990416
- Yin, R. K. (2009). *Case study research: Design and methods*. London: Sage.

Appendix A

Observation schedule

Date:..... Time:..... School:.....

Teacher:..... Subject:..... Year:.....

Time	Descriptive notes	Reflective notes

Appendix B

Interview schedule

Background questions

- (1) What is your name?
- (2) How long have you been teaching in schools?
- (3) Which science subject(s) do you teach, at which level and to which Year?
- (4) Tell me a few things about your educational background. In which subject was your Bachelors' degree? Do you have a Masters' or PhD?
- (5) Have you ever taken a course or professional development in educational assessment?

Part 1: General questions about assessment practices in science

- (6) In your view, could you please describe the role/purpose of CA in your own teaching practice?
- (7) How important is it for your teaching to be able to make frequent judgments about your students' understanding and learning about the topic you teach?
- (8) How easy or difficult is it for you to balance summative and formative assessment purposes?
- (9) Which CA practices do you usually use to assess student learning on a daily basis? Do you apply some routines? What are they? Could you please describe them and their purposes?
- (10) What has your teaching experience taught you about the effectiveness of assessment practices?

Part 2: Questions about participants' own practices

I may observe: sharing and understanding learning criteria, adjusting teaching, questioning, providing verbal feedback, self- and peer-assessment activities.

- (11) Which CA practices did you use in the observed lesson?
- (12) What was the purpose of this assessment practice? Why did you apply this practice at that particular time?
- (13) Why do you think that this practice fits its purpose?
- (14) How often and on which occasions do you use this practice?
- (15) What did you discover about student learning through this practice?
- (16) Are the outcomes of this practice enough for you to have a general picture of student understanding/learning? Are there limitations?
- (17) What use did you make of the outcomes of these assessment practices?
 - (i) Did you change your teaching plan because of the outcomes?
 - (ii) If so, how did they affect student performance? Are the outcomes used for summative purpose?
- (18) How were the assessment outcomes communicated to students?
- (19) In your view, does this practice take student learning further? If yes, can you say how?
- (20) I noticed that when you questioned your students, you usually ... Why did you form your questions or your response in this way?
- (21) Do you use specific criteria to evaluate student responses or do you try to find the best answer among student responses?
- (22) Do you think that students took action on your comments or your targets in their answers?
- (23) How important is it for you to give feedback to students? In your opinion, how does this practice promote student learning?



© 2018 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

