Science Education International
Vol. 18, No. 3, September 2007, pp. 189-208

International Council of Associations in Science Education

Assessment for Learning in Teaching and Assessing Graphs in Science Investigation Lessons

OLGA GIOKA, (ogioka@edc.uoc.gr), *Department of Education, University of Crete, Greece*

ABSTRACT *The major literature review by Black and Wiliam (1998a) provided strong research evidence indicating that assessment for learning can produce significant gains in students' learning. The present study applied the main aspects of assessment for learning, as highlighted in the Black and Wiliam review, in a specific area of science education, that of graphs in investigation lessons. It is a qualitative study with ethnographic aspects aiming to look at the extent to which science teachers employ assessment for learning practices, when they teach and assess graphs. It involved close observation of nine science teachers for one school year, collection of marked copies of students' investigation reports, and two interviews with each of them. Two in-depth interviews were carried out with each teacher to look at how they teach and assess graphs, and how they help students. One interview (pre-observation), prior to the classroom observations and one upon completion of the observations (post-observation). The study found that only a few of the participant teachers implemented some aspects of assessment for learning practices. The contribution of the present study is that it exemplifies what good quality feedback in graphing skills looks like. It is argued that science teachers need professional support encouraging them to implement assessment for learning practices.*

KEY WORDS: *Assessment for learning, feedback, graphs*

Introduction

During the last years, there has been given a great deal of interest in research on and teachers' implementation of assessment for learning. Three literature reviews and research studies relate to this interest. The English literature review by Black and Wiliam (1998a) provided strong evidence regarding the impact of formative assessment on students' performance. Black and William highlighted the importance of peer- and self-assessment, the qualitative and informative feedback in marking, and the communication of clear learning goals and assessment criteria, and how these impact students' learning. In particular, the aspects of feedback, peer-and self-assessment constitute the 'assessment for learning' perspective of the present study. In terms of the curriculum, the study focuses on the topic of graphs within the broader area of science investigations in secondary education.

The review of the French literature by Allal and Mottiez-Lopez (2005) had a particular focus on the concept of "regulation" (how teachers orchestrate learning for and with students). Allal and Mottiez-Lopez (2005) emphasised not only the importance of feedback, but also how to taylor instruction to the needs of different students and the importance of providing them with skills and tools for self-assessment.

The review by Koller (2005) explored the German literature in educational psychology. This review put emphasis on how students respond to various forms of feedback, a key element in formative assessment. The findings pointed to the greater impact of feedback on individual progress toward learning goals, rather than on comparison with other students.

The purpose of the presented study was to explore the extent to which the teachers employ assessment for learning practices when they teach and assess. The study took into consideration the important aspects of formative assessment, as these were highlighted by Black and William (1998a), and attempted to investigate whether the teachers communicate their learning goals and their assessment criteria, and whether they provide feedback in marking and opportunities for self- and peer-assessment to their students.

In England, where the present study was carried out, the most common component of practical work for secondary education is investigations. In each investigation, students progress through four phases: prediction and planning, data collection, analysis of evidence and conclusions, and evaluation of the whole task. The graph itself plays an important role in the analysis of experimental results, drawing of conclusions, and the evaluation of an investigation.

Graphs Are Not Easy

Graphs present results, organise data, demonstrate patterns and relationships, and communicate scientific knowledge. A considerable body of research identified students' (in primary and secondary education) difficulties in drawing, interpreting, and using graphs. Students at all ages tend to hold alternative conceptions, make mistakes, and experience difficulties when they deal with graphs (Tairab & Khalaf, 2004). They have difficulties when they draw (Wavering, 1989; Berg & Smith, 1994) and interpret graphs (McKenzie & Padilla, 1986; Mokros & Tinker, 1987). Interpretation of graphs usually refers to students' ability to read a graph, and make sense of or gain meaning from it (Leinhardt, Zaslavsky, & Stein, 1990). In contrast to interpretation, construction refers to:

> *the act of generating something new ... building a graph or plotting points from data (or from a function rule or a table) ... In its fullest sense, construction involves going from raw data through the process of selection and labelling of axes, selection of scale, identification of units, and plotting'* (p. 12).

Leinhardt and her team (1990) claimed that:

> *construction is quite different from interpretation. Whereas interpretation relies on and requires reaction to a given piece of data (e.g., a graph, an equation, or a data set), construction requires generating new parts that are not given'* (p. 13).

The latter statement indicates that plotting a graph is more demanding and difficult than the interpretation of a given graph. Leinhardt and her colleagues (1990) put it as follows:

> *in terms of their relationships to each other, it can be noted that whereas interpretation does not require any construction, construction often builds on some kind of interpretation'* (p. 13).

The AKSIS (ASE-King's Science Investigations in Science) project team (1999) argued that the construction and use of graphs are two areas that students face difficulties. Over 75% of line graphs and bar charts were constructed incorrectly, and hardly any students used their graphs to make predictions for values they had not tested (by interpolation and/or extrapolation). The AKSIS study attributed poor competence in graphing to the fact that students are not explicitly taught how to draw graphs in school science, and do not use graphs as a tool to communicate results. Secondly, it was suggested that this was due to the fact that graphs are usually considered as an end-point of the activity, and they are not used for the interpretation of evidence (Goldsworthy, Watson & Wood-Robinson, 1999). Although the majority of students from 11-15-years old could read straightforward information from graphs, tables and charts, and could perform basic interpretation tasks, they experienced difficulties in more demanding interpretation tasks. The TIMMS study (1997) on Mathematics and Science showed that both primary and secondary students had difficulties in describing trends and identifying a pattern in the data (Harmon, Smith, Martin, Kelly, Beaton, Mullis, Gonzalez, & Orpwood, 1997).

Theoretical Context of the Study

According to Black and William (1998a), the term 'assessment for learning' can be used to describe all those activities undertaken by learners and teachers for the purpose of assisting learners in finding out where they are in their learning, where they are going, and how to get there. Formative assessment or assessment for learning serves its formative function when the information is fed back to the learners, and the subsequent activities in which they are engaged, so that they can directly guide further learning and improvement. A key role for the teacher in formative assessment is the elicitation of evidence relating to students' current attainment, and the provision of qualitative feedback and guidance on what they can do to improve their learning. A key role for the learners is the understanding of the criteria that will be applied to their work, and the ability to assess their progress towards the learning goals. Black and William (1998a) stated that the essential elements of any strategy to improve learning through implementation of formative assessment would be the setting and communication of clear goals, the design of meaningful learning and assessment tasks, the communication of assessment criteria, the provision for good quality feedback (oral and written), and opportunities for self- and peer-assessment. This set of guiding principles underlying teachers' classroom strategies stems from an attention to the social discourse in the classroom and the socio-cultural theory of learning (Gipps, 1999; 2002).

Black and William (1998b) also provided strong evidence indicating that:

classroom formative assessment [when] properly implemented [becomes] a powerful means to improve student learning, [and that] summative assessment, such as standardized exams, can have a harmful effect' (p. 19).

They concluded that:

formative assessment is an essential feature of classroom work and that [its implementation] can raise standards' (p. 19).

Thus, the present study attempted to identify the kind of feedback that science teachers give to their students during science investigation and graphing lessons, and whether they use formative assessment practices, when they teach and assess graphs.

Methodology

Teachers and Students

The study was conducted over the period of one full school year. It was carried out in four schools in three different local educational authorities in the greater London area. Nine science teachers teaching upper secondary classes (GCSE, General Certificate in Secondary Education and A-level, University entrance exams) participated in the study. The teachers were well-qualified in science and had substantial experience in teaching and assessing GCSE and A-level coursework, and had attended relevant in-service courses. The four schools were representative of a range of different types of schools. Students in the observed classes came from a wide variety of social and ethnic backgrounds, and were predominantly from middle- and low-income families. The science teachers that participated in the study were volunteers, and the sample was a convenient one. Table 1 presents the four schools and teachers, their qualifications, and teaching experience, using pseudonyms. These pseudonyms and pseudonyms for students are also used throughout the paper.

Table 1
Teachers, Teaching Experience and Qualifications

Teacher's name	School	Teaching experience	Classes	Degree/ Qualifications
Miss Warren	Rosehill School	1 year	GCSE Science	1st degree in Environmental Science, PGCE in Science Education.
Miss Platt	Rosehill School	5 years	GCSE Science	1st degree in Secondary Science, PGCE in Science Education.
Mr. Michael	St. Thomas' RC School	20 years, Head of the Science Department	GCSE and A-level Chemistry	1st degree in Chemistry, PhD in Chemistry. No teacher training.
Mr. Moss	Riverview School	2 years	GCSE and A-level Physics	1st degree in Physics and Astronomy, PGCE in Physics Education.
Mr. Scott	Riverview School	8 years secondary and at teachers' colleges	GCSE Chemistry and A-level Biology	1st degree in Chemistry and PGCE in Science Education.
Mr. Hall	Riverview School	7 years and teaching at the University level	GCSE and A-level Biology	1st degree in Biology, PhD in Zoology, PGCE in Science Education
Mr. Jones	Riverview School	24 years, Head of the Science Department	GCSE and A-level Physics	1st degree in Physics, PGCE in Science Education
Mr. Smith	St. Margaret's RC School	28 years, Head of the Science Department	A-level and GCSE Physics	1st degree in Chemistry and Biology.
Mr. King	St. Margaret's RC School	24 years, Assessment Coordinator of the school	A-level and GCSE Chemistry	1st degree in Chemistry, PhD and post-doc in Chemistry. No teacher training.

Context of the Study: The Lessons

The lesson observations were conducted throughout one year and were evenly distributed across the teachers. They included approximately 210 one-hour science lessons that included science investigations. An investigation is a practical activity of a fair-test kind that engages students in collecting data, drawing and interpreting graphs, and using them to develop an explanation, reach some conclusions, and evaluate the whole activity. The choice of the topics of investigation was not under the researcher's control, because these were selected by the teachers according to the existing syllabus. Common investigation topics were: photosynthesis, rates of reactions, osmosis, the effect of temperature and enzyme concentration on enzyme activity, anaerobic respiration, and so on. From all the observed GCSE and A-level lessons, the study focused mostly on observing the investigation activity, and the lesson units in which students collect, analyse and interpret experimental data. Table 2 presents information about the investigation topics across schools and teachers.

Table 2
Investigation Topics across Schools and Teachers.

Teacher's Name	School	Investigation topics
Miss Warren	Rosehill School	Neutralisation, reactivity, Ohm's Law, osmosis, photosynthesis, rates of reaction, Hooke's Law
Miss Platt	Rosehill School	Neutralisation, reactivity, Ohm's Law, osmosis, photosynthesis, rates of reaction, Hooke's Law
Mr. Michael	St. Thomas' RC School	Rates of reaction, anaerobic respiration, photosynthesis, osmosis
Mr. Moss	Riverview School	Ohm's law, photosynthesis, osmosis, Hooke's Law , enzymes
Mr. Scott	Riverview School	Osmosis, rates of reactions, anaerobic respiration, enzymes
Mr. Hall	Riverview School	Enzymes, photosynthesis, osmosis, rates of reaction
Mr. Jones	Riverview School	Rates of reactions, Ohm's Law, photosynthesis, osmosis
Mr. Smith	St. Margaret's RC School	Photosynthesis, enzymes, reactivity, Ohm's Law
Mr. King	St. Margaret's RC School	Osmosis, rates of reactions, photosynthesis, anaerobic respiration, enzymes

Procedure

The study was interpretive, and the content of interest was the perspectives and actions of the participants within a socio-cultural context (Lincoln & Guba, 1985; Miles & Huberman, 1994; Erickson, 1998). It was a small-scale classroom-based research and exploratory in nature. The study involved the documentation of what actually goes on in the teaching process, through first-hand observation and in-depth interviews with the participating science teachers in a classroom situation. The qualitative and interpretive design of the study included:

- Classroom observations in which teachers were observed as they were teaching investigations and graphs.
- Two in-depth interviews with teachers to look at how they teach and assess graphs, and how they help students improve. One interview (pre-observa-

tion) was conducted prior to the classroom observations, and the other upon completion of the observations (post-observation).

- Informal discussions with teachers, and,
- Collection of students' investigation reports, as these were marked by the teachers.

Classroom observations involved audio-recording the teachers' talk and note taking. Audio recording was carried out on a regular basis in each class through each teacher wearing a tape-recorder and a clip-on microphone that picked up the teachers' voice, and any student talking nearby. All recordings were transcribed in full. The teachers validated the observation notes and read the developed narrative describing their teaching.

Pre-observation interviews were semi-structured. Some items of the pre-observation interviews are shown in the Appendix I. Teachers were also asked to bring along any planning sheets, schemes of work they would possibly use, and a copy of the National Curriculum. They were asked to refer to these documents during the interviews.

In the post-observation interview, teachers were provided with copies of students' investigation reports and were asked to assess and compare them, and explain the assessment criteria they used. They were also asked how they would explain to the children what they had to do to improve their performance, and how they would plan to teach a lesson to help them improve their attainment and coursework report. All the interviews were audio taped and transcribed verbatim. The participating teachers checked the interview transcripts for accuracy; they made corrections or added some points prior to any analysis. This was a deliberate attempt to increase the reliability of the data from interviews that it is usually higher, when the informants are asked to check the interview transcripts (Guba & Lincoln, 1989).

Marked investigation reports were also collected to examine the kind of feedback that teachers provided to their students.

Data Analysis

The process of analysis of the teacher interviews and the observed lessons transcripts was characterised by action at different levels of detail, beginning with broad groupings, then identifying finer aspects of the data, and sorting them out into more specific categories. Thus, categories for analysis were not set a priori, but emerged from, and were grounded in, the data to identify potential themes for each participant (Miles & Huberman, 1994). The qualitative data in the lesson observation transcripts were analysed using a grounded theory approach (Glaser & Strauss, 1967). The field notes and the interview transcripts were analysed through the process of analytical induction to find patterns in the teachers' responses to the interview questions. An iterative procedure was used for uncovering the existing patterns. This entailed the continual examination and re-examination of the data to identify themes or patterns, which were then verified. The analytical process involved extensive listening to the audio-tapes, and reading and re-reading of the transcripts to gain a better sense of teachers' practices. From the transcripts, an initial set of categories was generated. This involved taking teacher responses in interview transcripts and trying to group similar statements together by describing what they had in common. The initial categories were then reviewed and revised, and the teachers' responses were finally categorised. When teachers' responses did not fit into the emergent categories, a new category would be defined. In addition, a second researcher reviewed the transcripts and agreed on the developed categories.

The second level of analysis consisted of an interpretive analysis of the data. It involved revisiting the data from the perspective of formative assessment. The purpose here was not to argue how effective the employed strategies were, but to examine whether and to what extent teachers employed formative assessment practices. For example, the aim was to identify instances of the formative assessment processes (i.e., instances of communication of the learning goals and the assessment criteria) in teachers' talk in the observed lessons. For this purpose, the teachers' actual practices were compared and contrasted with practices that were drawn out from the findings of research studies on formative assessment. The intention was to identify lesson segments and divide the transcripts into the categories drawn from the literature review on formative assessment. The theoretical framework reduced the problem of studying teachers' strategies of teaching and assessing graphing skills to looking into the:

- communication of learning goals
- communication of assessment criteria
- marking – written feedback, and
- opportunities for self- and peer-assessment.

This process included looking within and across teachers. To avoid misinterpretation of the results, it has to be noticed that the findings represent general trends for which the amount of supporting data substantially exceeded the amount of refuting data.

Triangulation

As themes and patterns began to emerge, data from lesson observations, the two interviews, and collection of students' marked reports were compared and contrasted to support, further elaborate, or suggest contradictions between what teachers stated in the interviews and what they actually did in the classroom. Triangulation was used throughout the analysis across data sources for each of the teachers (Lincoln & Guba, 1985). Triangulation is not seen as generating a unique knowledge of pure facts and truths, but as ensuring that the constructed knowledge constitutes a co-ordination of data and interpretation (Muralihdar, 1993). The fact that data was provided by different methods (interviews, observations, and collection of marked students' work) facilitated a more valid and reliable interpretation of the collected research evidence. These multiple data sources and methods of data collection allowed alternative perspectives to construct and shape the researcher's interpretations. Excerpts from the lesson transcripts and quotations from the interviews are also presented to illustrate the points made. Underlined words or sentences stand for words emphasised by teachers in the interviews or in the observed lessons. The interviews elicited statements that clarified what was happening in these classrooms, and provided further evidence supporting or diverging from the patterns noted during classroom observations.

Results and Discussion

The results of the analysis and their discussion are presented in different categories, as these emerged and were validated.

Students' Difficulties with Graphs

All the participating teachers were aware of the difficulties and the problems students experience with graphs. They referred to students' alternative conceptions associated with graphs, amongst which the most common was young children's conception that a straight line is always more correct. They also referred to difficulties and understandings of the line of best fit, that is, whether the best fit is a straight line or curve, and whether it goes through the origin 0(0,0). They explained students' difficulties by referring to the lack of coordination between Mathematics and Science Curriculum in secondary school. Better co-ordination of and collaboration between the two departments in a secondary school would help students with graphing skills from their early mathematics lessons. Three teachers (Mr. Scott, Mr. Hall and Mr. Moss) emphasised A-level students' difficulty in selecting suitable scales and units, writing an informative title, and pointing out anomalous data. The same difficulties were identified by the other teachers who also stressed students' difficulty to describe and explain the general trend, and the relationship between the variables by using scientific theory. Regarding the interpretation of graphs, teachers reported that students are not good at giving a quantitative statement concerning the relationship between the two variables by quoting figures from the graph. They have difficulties in calculating gradients, 'translating' the gradient in scientific terms, and giving it a scientific meaning. Finally, teachers said that it is often the case that students do not see a purpose for drawing graphs, and they rather consider that the drawing of graphs is an end in itself.

All the teachers talked about the learning experiences they provide to students to offer them training and practice in drawing and interpreting graphs. Mr. Smith said that he uses the Excel program for students to obtain help with the drawing, which they sometimes find more difficult than the interpretation of graphs. This is in line with one of the main themes in the review of research studies on graphing (Leinhardt et al., 1990).

The Role of Graphs in Investigation Tasks

When teachers were asked about the main purpose of students drawing a graph, most of them replied that the purpose of graphs is to visualise experimental results and identify relationships between the two variables. They also stated that graphs are crucial in the analysis of results, in making quantitative statements, and identifying errors and anomalous results. Teachers also reported that students tend not to use the plotted graphs in the analysis of results. Instead, they use figures from the results table. Mr. Hall quite often referred to this tendency and said that he keeps on reminding students to quote figures from their graph and not from the Table of results. Mr. King also said that students may not see any purpose for drawing a graph. This is also documented in other research studies on students' graphs, such as the AKSIS Project (Watson, Goldsworthy & Wood-Robinson, 1999). Observation data showed that most teachers did not use the plotted graphs as a tool to create a 'rich' discussion in the classroom or to develop an argument (Roth

& McGinn, 1997). There were some teachers, like Miss Platt and Miss Warren, who talked about the development of graphing skills as such, but did not go further to say how graphs can be used in the analysis section and in the evaluation. For example, although Miss Warren and Miss Platt used computers for data-logging and graphing, they did not place much importance on graphs to develop students' deeper understanding of the phenomenon under investigation. This was also evident in Mr. Michael's teaching, whereby the lesson ended with the drawing of graphs.

On the contrary, Mr. Hall, Mr. Scott, and Mr. Smith emphasised the use of graphs in the analysis of results. This was also the case of Mr. King who was explicit about the purpose for plotting a graph. Mr. King explained that one of the principal reasons for drawing graphs is to use them in the evaluation section, that is, to identify anomalous points and evaluate the whole activity. In one of his lessons, Mr. King stated that *'So, it* [the graph] *can help by identifying points to make for your evaluation'.* And later he completed that *'you have to learn more generally what graphs do for you, and how you use them'.* In the observed lessons, Mr. King made extensive use of the plotted graphs, firstly, to develop understanding of the key concepts related to the investigation topic, and, secondly, to promote students' understanding of how to use the graph as a tool in the analysis and evaluation. The same teacher (Mr. King) emphasised that he would like to make full use of graphs even from earlier years science classes, but he could not because students usually do not have the appropriate background and experience from mathematics lessons.

The importance of the analysis of experimental results by using the plotted graph was frequently underestimated and on occasion bypassed altogether, because of the pressure of time (Watson et al., 1999). But, it is at this point that a critical understanding of experimental evidence takes on its full impact, because students have to use evidence and the graph as a tool to support their analysis and conclusions.

Emphasis on Exams

Results indicate that there was a strong emphasis on students' preparation for GCSE and A-level exams. Half of the participating teachers (Mr. Hall, Mr. Scott and Mr. Moss at Riverview School, and Mr. Michael at St. Thomas's School) emphasized that their main goal was to prepare students to succeed in their exams and obtain a high mark. Evidence from the observed lessons was strong regarding the tension to cover the exam syllabus and the pressure that exams exert on teachers. Observations showed that the assessment system has been used to motivate students to be prepared for the external exams as better as possible. Teachers, like Mr. Hall, Mr. Scott, Mr. Moss and Mr. Michael, referred to the 'exam-driven' system and to their teaching, which looks like 'a game,' because the first priority for students is to pass the exams with a good mark. Mr. Hall and Mr Scott provided the following explanations, respectively:

> *I am playing the game; I am trying to help them get the best mark possible'. 'It's very much exam-driven really'. 'We are very much here an exam-factory'*
>
> *The aim is your students to do well in the exams and get a good mark'.* And, *'I teach coursework differently from usual work. You show them the mark scheme written by the examiner. 'This is what you need for perfect marks.*

Consequently, teachers have to clearly and explicitly communicate the exam board criteria and not necessarily to deal more deeply with the appropriate quality of work. In getting students to be prepared to meet the exam requirements, teaching does not aim at a deeper understanding of scientific theory and development of investigative skills, but at a superficial sort of learning focusing on examination style questions. For example, Mr. Jones in talking about how he makes his day-to-day instructional decisions said that he always had '*exam questions in* [his] *mind*'. Another example is one question by Mr. Hall asking one student to correct his coursework by explicitly stating: '*How could you fit in with the criteria that the exam board wants?*' Obviously, by attempting to get students prepared for the exams seems to be happening at the expense of promoting their investigative skills. In Mr. Hall's words: '*You have to just do it, and move on really with the content*'. In contrast, Mr. King and Mr. Smith seemed to be interested in giving students a strong background in doing investigations, so that they are better equipped to cope with demands of subsequent years.

Since teachers have to cover the syllabus in a limited length of time, they do not have enough time to teach investigative skills at an appropriate pace, so that they support students with particular learning difficulties. Mr. King said that '*they are constrained by time, the National Curriculum, and the syllabus they should cover for the exams*'. In addition, they do not have the appropriate time to focus on graphing skills, and give the attention they would like to help students with alternative conceptions and common mistakes with graphs. Hence, teachers placed much importance on the teaching of theory, concepts, and facts. This finding supports the argument made by Hodson (1992) that the teaching of investigative science is rather a lost educational opportunity or, as a recent Ofsted report (2004) documented, scientific enquiry is not being systematically developed.

Teachers' Marking: Written Feedback

When teachers talked about marking, they made a distinction between two different strands: marking class and homework, and marking of coursework. For the first case, marking of class and homework, as well as of test and exam questions, is simple and straightforward. When a single-word answer is required, it is most likely that there is only one 'right' answer. Teachers correct answers by putting a tick or a cross for 'right' answers. When longer answers are required, there is usually a maximum mark for each question, and it is the teacher who allocates it. For the case of coursework, all teachers said that they mark it against the exam board criteria. However, only three teachers mentioned that marking of coursework, although based on the exam board criteria, is subjective, since teachers quite often have to use their professional judgement about the intermediate marks. All the participant teachers underlined that marking is time and effort-consuming.

Most teachers gave ticks and grades, corrected mistakes in science reports, and wrote some comments either for the graph or for the whole write-up as follows:

> *Graphs need to be labelled', 'Try drawing a line of best fit', 'Label axes and give your graph a title', 'Well done! You have drawn a good line of best fit.*
>
> *This is good, but do you think you could change the design of your experiment or improve it in any way?*', '*Well done. This is a good write-up. Try to be clearer in your method*'; '*You have not completed this work - why?*

Miss Warren and Miss Platt would give merits, smiley faces, and praise to their students. In the case that students should make corrections or changes, Miss Warren wrote: '*Changes required*' and '*This needs to be re-written*'. Stickers and stars may have been given for motivation reasons, but they did not help students understand how to 'close the gap' and what they need to do to improve. Perhaps the two teachers (Miss Platt and Warren) made extensive use of praise and awards, because their students were of average or low achievement.

Mr. Jones reported that marking mainly serves summative assessment functions or it is used for assigning a mark, and that when he marks a piece of coursework:

> *he has in mind the assessment criteria for the benefit of moderator, not of the child. Has the student met the criteria? This is for the moderator rather than the student*

Mr. Smith and Mr. King talked about science report marking as a source of information on students' learning needs and difficulties. As both teachers explained, information from science report marking readily informs teachers' subsequent planning and teaching, or what to do 'next', which is at the heart of formative assessment practices. Mr. Smith underlined the fact that young children do not read written comments in their books. In fact, the quality of dialogue and discussion in feedback sessions is important, and most research indicates that oral feedback is more effective than written feedback for younger students (Boulet, Simard & Demelo, 1990). Therefore, Mr. Smith, instead of giving written comments about how to improve and what they need to do, organised feedback sessions where he talked to students individually about their work, 'scaffolded' their thinking by appropriate questions to stimulate thinking, and gave them time to work on the provided oral feedback. In such discussions (class or individual), Mr. Smith told students what they had achieved and what they had not, with specific reference to the nature and the cognitive demands of a piece of work. Students also talked about their learning difficulties. The teacher did not give them complete solutions and answers, but he prompted them with questions to encourage them to think. Giving feedback involves making time to talk to students, and to teach them to be reflective about the quality of their responses and coursework. Mr. Smith emphasised that the focus is not on marks and levels, but on the effort and the progress students have to make. For example, he guided them, as follows:

> *Have you done your best? Are you making progress?'. Students should be aware of what they should be achieving or aiming at improving when developing the next piece of work. Work itself maybe introduced as 'This is to develop ....'. 'You are doing this, because I want you to be better at ...'. This is what will help you get better at ...*

Mr. King emphasized that '*the feedback he gives to students is a reflection about what he feels that would be expected of that individual child for a particular task*'. Thus, Mr. King considered both the demands of the task, and the 'potential' of each individual student in order to give feedback and assign a mark. Mr. King also provided specific guidance on what needs to be done and how it needs to be done. For example, in one A-level piece of coursework, the teacher suggested that the graph has to be used to prove the direct proportional relationship in the analysis section:

> *At this point you needed to use data from your line of best fit to <u>prove</u> the direct proportion you claim exists. You have numerous examples in your KS3 books to guide you on this. You have explained why a longer wire will have more resistance, but why is it a directly proportional relationship?*

With feedback in the form of a question, Mr. King encouraged his GCSE students to think about how to improve the explanation by them. Consider also how the teacher encouraged them to refer back to previous year's work on the same topic. In another case, Mr. King acknowledged the progress made by the student in graph work and gave some guidance about improvement needed. In doing so, he started with students' strengths, such as, '*the axes and scales used are correct, the points are plotted accurately*' and he went on to mention weaknesses: '*But the line of best fit is not entirely suitable*'. At the end of his comments, the teacher raised a question to get the student to think about how to improve his graph: "*I have added a more suitable one, can you see why it is better?*"

From a similar line of thinking, Mr. Hall reported that he gives to his A-level students only comments without grades. He believes that if students obtain both comments and grades, they ignore comments and they attach more importance to grades. Such students' tendency is again well-documented in the literature (i.e., Butler, 1988). According to Butler (1988), written comments are more beneficial than grades, or than both comments and grades. Thus, Mr. Hall's comments and written feedback in the form of questions had a specific formative function, and were intending to encourage students to think about what they had to change and correct in their draft work. Suggestions for changes and improvements act as 'scaffolding' tools in the learning process.

For the case of a graph questions, such as, '*What have you missed on that graph?*' and '*What's missing?*' were used to encourage them to make their graph perfect. Similar questions were stated for the interpretation of graphs: '*You've drawn your graph this way, how can you interpret it quantitatively?*' For the construction of graphs: '*How could you do it better? How could you fit in with the criteria that the exam board want?*' For the title of the graph: '*Is there any more detail you can put in your title?*" For the scale of the graph: '*Why have you done that so small? If you doubled the scale, it would be probably better. Draw accurately and read off properly*'. For the analysis section, to get students to think about how to make a quantitative statement by using figures from the graph: '*How can you make this quantitative?*' '*How can you use this graph?*' or : '*If you increase light intensity, how much does the rate of photosynthesis increase? If you double light intensity, what happens to the rate of photosynthesis?*', '*Can you explain the anomalous results?*'

On the whole, from the participant teachers' marking strategies, Mr. Hall, Mr. Smith, and Mr. King gave more constructive feedback, that is, feedback pointing out how the work could be improved by informing the students about the learning goals, the assessment criteria, and the intended quality. In addition, the same three teachers provided students with time to respond to marked work. Students were also provided with time to read and act upon comments on marked work by reworking it and handing it in again. The three teachers were "formative" in their approach, because they found out what students knew or did not know, and they were then allocated time and effort on the areas that required improvement rather than simply ensuring coverage of a topic. The formative assessment was practised by providing lesson time for students to improve their first attempt in the supportive environment of the class and with the teacher at hand to check their progress (Sadler, 1989; 1998).

Communication of Learning Goals

Observations showed that most teachers (i.e., Mr. Michael, Miss Platt, Miss Warren) presented the learning goals and objectives in the form of key scientific terms, at the beginning of a new module or a lesson unit, and summarized them at the end of the lesson, in the plenary session. For example, Miss Warren and Miss Platt would write the learning goals on one side of the board, read them out at the beginning of a lesson unit, and, at the plenary session, they would summarize the key points of a lesson by reminding students of the learning objectives. But, it may be difficult for students to understand what they would learn when many new key terms are introduced to them for the first time. However, in the first interview, nearly all teachers insisted that the learning goals had always been clearly defined. Students needed more than teachers' writing learning goals on one side of the board or copying the goals down at the beginning of each lesson unit. Only a few of the teachers referred to the learning goals during the course of a lesson and expressed the learning goals in a comprehensible way for students. This was the case of Mr. Smith and Mr. King, who apart from presenting learning objectives in an appropriate language for students, they quite often referred to them as they were teaching. For example,

> *That's the first most important analysis jotting that it's going to go down. Because you are trying to explain why ... on here and comment ... that concentration between there and there, the results show that the potato gained weight. Whereas, the concentration from here to here in the results table show that the potato lost weight* (from a lesson transcript).

Communication of Assessment Criteria

Communication of the criteria applied in assessing students' attainment was a clearer and more common case in GCSE and A-level (upper secondary) classes. This mainly happened, because students should be aware of the exam board criteria and requirements in order to succeed in the external exams. Obviously, it is rather appropriate to differentiate between informal assessment criteria and exam board criteria, because teachers may communicate the criteria and the intended quality of student performance, but this mainly occurs for examination purposes. A characteristic case was Mr. Hall who was always explicit about the quality of expected coursework. Mr. Hall 'modeled' the quality of an answer, of an explanation or the whole investigation report by being specific about the components of each part. He also guided and 'scaffolded' his students to achieve a high level of attainment in order to enable them to obtain a high mark in the exams. This was in contrast with Mr. King who also modeled the quality of different sections in coursework, but, at the same time, his principal aim was to develop investigative and good writing skills in students. In doing so, Mr. King did not refer to the requirements of the external exams, but he emphasized and reminded students of 'the reader' who is going to look through their science reports and should understand them.

> *So, when you come to these things you have to think about your explaining it to somebody who has not done this work. They will only understand it if you write it properly and clearly, which is exactly the way your introductions and predictions have been done.*

Mr. King assisted students by encouraging them to reflect on how they could improve it, so that similar or better quality could be reached next time, in the next investigation, in a similar or different topic context. Mr. King gave appropriate details for the required quality of coursework informed by and based upon the task demands. He shared the assessment criteria with students, as indicated in the following example from a GCSE lesson relating to osmosis. In this specific example, the teacher explicitly explained that "*the success of your work depends on what you do put in both the analysis and then how fully you actually explain it by using scientific ideas about osmosis...*". Mr. Smith also communicated the intended quality, but not as explicitly as Mr. King, while the rest of the teachers were not committed to communicating the assessment criteria.

Assessment criteria and learning goals need to be expressed in a language that students understand, so that they really understand the goals they are pursuing and the quality of work they are aiming to. But, in the actual practice, assessment criteria were often defined in a 'formal' language that may be vague or ambiguous to students. For example, with regard to the interpretation and evaluation of experimental results, one criterion was for students to be able to '*complete a calculation, and to do so in an efficient and economical manner*' (from a student's handout). Some teachers, for example Mr. Smith and Mr. King, used a student-friendly language or a version that they had developed by themselves. The exam board criteria version may not be so clear even to older students, so that teachers need to use a friendly-student language or version. As Mr. Smith stated *'they [students] are given the criteria in students' language as nearly as possible'*. Secondly, communicating goals and assessment criteria to students means sharing them with students in the learning process, and continually referring to and reflecting on the possible learning outcomes when students work on a task. This was the case of Mr. King who shared the learning intentions with students by being clear and explicit about what students should demonstrate in carrying out an investigation task, and by being specific about the content when the class was going to write the science coursework report.

Self- and Peer-Assessment

All teachers, except one, reported that they did not give opportunities to students to mark their own or peers' coursework. Opportunities for self- and peer-assessment were very rare, and when such opportunities occurred, teachers would give students the opportunity to mark only end-of-unit tests or simpler exercises, which had blanks to be filled in, and asked for one 'right' answer that was usually provided by the teacher or by the textbook. In self- or peer-assessment of textbook exercises, students used a checklist and they put a tick next to the correct answer or the 'right' missing word. A general explanation offered by almost all the teachers related to the students' tendency to undervalue self- and peer-assessment, and give more importance and value to marks given by the teacher. Teachers explained that marking coursework is a '*high demanding task*', and, secondly, coursework 'counts' in the overall mark in their exams. Therefore, what Mr. Michael did was to take out marks from previous years' students' coursework and let his current students mark it by using the exam board criteria. Mr. Michael commented that his classes were very good at it. Another teacher, Mr. Hall, insisted that since coursework is a '*demanding task*', he gave to students only exercises and questions from the textbook or exercises that he developed by himself for students to answer and assess. Students marked their homework or class work every other lesson. When Mr. Hall's students did self- or peer-marking, not only they had to correct wrong answers and decide on grades, but more importantly they had to write comments and suggest improvements. Mr. Hall explained that this sort of peer-assessment helped them understand what they need to change, re-write or correct in order to improve answers and obtain a better mark.

Mr. Smith said that self- and peer-assessment are not used extensively, because they are very time-consuming and they lead to disputes among students. An additional problem mentioned by the same teacher was that students do not give the appropriate concentration on that, and it is usually difficult for the teacher to keep all students moving through at the same pace. However, the teacher said that students can assess their partner's work when it is a simple test of questions with straightforward answers, like multiple-choice questions, provided that the teacher has already allocated in advance the maximum marks for each question.

In contrast with all the previous teachers, Mr. King explained that he intended to get students involved in marking coursework and to offer them opportunities for peer- and self-assessment. At the same time, he found it difficult to involve students in investigations that are usually '*high-level tasks*'. Therefore, the teacher developed in advance a checklist for each particular investigation topic, so that students could use it when assessing their own work or their peer's work. To illustrate the process, the teacher referred to a recent case with two different GCSE classes. He gave them one checklist for the same investigation topic. Both groups were asked to write comments for improvement and give grades to one another. Mr. King underlined that this practice proved to be very useful for students' learning. It helped them improve not only their attainment in the specific topic but in the next investigation too. Mr. King emphasised that students were honest in marking and deciding about grades. Obviously, Mr. King's checklist 'worked' not only as a guide for marking, but mainly as a guide for what students need to include in each section of the investigation coursework (planning and prediction, obtaining evidence, analysis and explanation, evaluation), or as a guide for the intended quality of students' work. This practice proved to be extremely useful since self- and peer-assessment are usually followed by a whole classroom discussion, where students have thus the opportunity to explicitly discuss the assessment criteria and the quality of the investigation report. The students and the teacher can also discuss likely discrepancies or mismatches between comments and grades they had received from other students and their own. After the discussion, students can work on suggestions and comments and they can produce a second version taking into consideration the highlighted points. However, Mr. King explained that self- and peer-assessment practices seem to be more effective for higher ability students, since writing an investigation report is a very demanding task that presupposes good writing skills, closely related to how well one can communicate his/her ideas to others.

Conclusions Implications for Science Education

The present study provided a detailed insight into the formative assessment practices focusing on the particular topic of graphs in science investigation lessons. The findings add to the relatively limited literature of assessment for learning in science education. For example, only a few of the participating teachers employed some aspects of formative assessment practices. Three teachers gave more constructive feedback, that is, feedback indicating how students can improve their work and only one teacher organized opportunities for self- and peer-assessment in the teaching of investigations. Only Mr. King was somehow committed to formative assessment practices, and was informative in his teaching and written feedback. He was also 'scaffolding' and modelling the learning process in terms of the cognitive processes involved. Mr. King was also the only teacher who targeted the development of self- and peer-assessment skills in students. This finding is consistent with earlier research studies reporting that the characteristics of high-quality formative assessment are not well understood by teachers and that formative assessment is weak in practice (Crooks, 1988;, Black, 1993). High-quality formative assessment is a complex process, as illustrated by research evidence (Black & Wiliam, 1998a). In particular, studies in science education (Black, 1993; Daws & Singh, 1996; Wiliam, Lee, Harrison, & Black, 2004) revealed that formative assessment practices are weak in the actual science teaching.

The content focus in a particular school subject, that of science, and in teaching the particular topic of graphing in investigation lessons is one of the strengths of the study. The study exemplified what good quality of feedback in graphing skills looks like. The analysis of teachers' feedback focused on the extent to which teachers gave feedback by communicating the intended quality and by making specific reference to the cognitive demands of investigation and graphing tasks. Good quality feedback is specific to the task and informed by the nature and the cognitive demands of an investigation activity.

Teachers do not use formative assessment practices, because these practices place high demands on teachers. Effective use of formative assessment requires that the teachers need to have clear objectives and assessment criteria, whilst maintaining the flexibility needed to respond to student needs, as they emerge from a particular investigation or a piece of work. Experience is crucial to having both the knowledge and confidence to follow this path of working. It also means that the teacher has to regularly mark, comment and review students' work, in order to keep up-to-date with their development as learners, and guide them towards the next step(s) that are necessary for improving their attainment. The implementation of formative assessment strategies is a very demanding and time consuming process, and some teachers may not be willing to work on this basis. For example, the implementation of peer- and self-assessment practices places many demands on the teacher. Before students get involved in self- and peer-assessments, they need to have deeply understood the learning goals, the assessment criteria and the quality of expected work. For the particular case of science investigations and graphs, before peer-assessment of coursework occurs, teachers have to get students through the teaching and communication of learning goals, assessment criteria and quality of written work, so that they know the legitimate text they should produce. Instructing students to get involved in self- and peer-assessment is a long and demanding process during which teachers should also teach students how to reflect on their learning, and take ownership of and responsibility for their own learning. In peer-assessment, students share assessment criteria and learning goals, and they contrast their ideas among one another, because they have to justify their decisions, comments, feedback, and the grades they will assign. The main idea is that people learn together, and from one another, by sharing ideas and attempting to become and function as learning community. The realisation that the sharing of ideas and practices is a critical element in the process of learning science, and that learning proceeds through a process of social construction between the teacher and students and among students. Self- and peer-assessments are also beneficial, because they feed-forward as they help students to internalise the required quality of work and consider it in the next investigation. The emphasis of self- and peer-assessment must be on the teaching and improvement of attainment rather than on 'right' or wrong answers, so that self- and peer-assessment becomes really a learning process.

Additional factors relating to self- and peer assessment also depend upon the school environment in which teachers have been working, such as how supportive the school or the science department is. Teachers need support from the science department and the school in general, and a positive school ethos and the science department's ethos are always necessary to promote teachers' professional development. Teachers need to share experiences and learn from each other, and receive advice and guidance from the head teacher, the head of the department, or the senior management team. They also need to share good practices and examples of effective strategies in staff meetings. The implementation of formative assessment depends to a large extent on a sustained program of professional development that will put emphasis on assessment for learning strategies in teaching and assessing graphs.

The study attempted to capture some of the complexities of assessment for learning and explain the processes underpinning formative assessment. The findings are useful indicators pointing out that further research on formative assessment in science education and teacher development is needed for putting the theory of formative assessment into improved practice.

References

ALLAL, L., & MOTTIER LOPEZ, L. (2005). Formative assessment of learning: A review of Publications in French. In Centre for Educational Research and Innovation (2005) (CERI) *Formative Assessment Improving Learning in Secondary Classrooms* (pp. 241-264). *Paris: OECD.*

BERG, C.A., & SMITH, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education, 78(6),* 527-554.

BLACK, P. J. (1993). Formative and Summative Assessment by Teachers. *Studies in Science Education, 21,* 49-97.

BLACK, P., & WILLIAM, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5 (1),* 7-75.

BLACK, P., & WILIAM, D. (1998b). *Inside the black box. Raising standards through classroom assessment.* London: King's College London.

BOULET, M-M., SIMARD, & DEMELO, D. (1990). Formative evaluation effects on learning music. *Journal of Educational Research, 84 (2),* 119-125.

BUTLER, R. (1988). Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58,* 1-14.

CROOKS, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58(4),* 438-481.

DAWS, N., & SINGH, B. (1996). Formative assessment; to what extent is its potential to enhance students' science being realized? *School Science Review, 77(281),* 93-100.

ERICKSON, F. (1998). Qualitative research methods for science education. In B. J. Fraser and K. G. Tobin (eds.), *International Handbook of Science Education* (pp. 1155-1173). Dordrecht, Netherlands: Kluwer.

GIPPS, C. (1999). Socio-cultural aspects of assessment. In A. Iran-Nejad and P.D. Pearson (eds.), *Review of Research in Education* (pp. 355-392). Washington, DC: American Educational Research Association.

GIPPS, C. (2002). Socio-cultural perspectives on assessment. In G. Wells and G. Claxton (eds.), *Learning for Life in the 21st Century* (pp. 73-83). Oxford: Blackwell.

GLASER, B. G., & STRAUSS, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

GOLDSWORTHY, A., WATSON, R., & WOOD-ROBINSON, V. (1999). Improving investigations from the AKSIS Project. *Education in Science, 183,* 22-23.

GUBA, E., & LINCOLN, Y.S. (1989). *Fourth Generation Evaluation.* Beverly Hills, CA: Sage.

HARMON, M., SMITH, T., MARTIN, M., KELLY, D., BEATON, A., MULLIS, I., GONZALEZ, E., & ORPWOOD, G. (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study (TIMMS).* Boston: Center for the Study of Testing, Evaluation and Education Policy.

HODSON, D. (1992). Redefining and reorienting practical work in school science. *School Science Review, 73(264),* 65-78.

KOLLER, O. (2005). Formative assessment in classrooms: a review of the empirical German Literature. In Centre for Educational Research and Innovation (2005) (CERI) *Formative Assessment. Improving Learning in Secondary Classrooms* (pp. 265-275). *Paris: OECD.*

LEINHARDT, G., ZASLAVSKY, O., & STEIN, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60(1),* 1-64.

LINCOLN, Y. S., & GUBA, E. G. (1985). *Naturalistic inquiry.* London: Sage.

MCKENZIE, D. L., & PADILLA, M. J. (1986). The construction and validation of the test of graphing in science (TOGS). *Journal of Research in Science Teaching, 23(7),* 571-579.

MILES, M., & HUBERMAN, M. (1994). *Qualitative data analysis (2nd edition).* Thousand Oaks, CA: Sage.

MOKROS, J. R., & TINKER, R. F. (1987). The impact of microcomputer-based labs on children's ability to interpret graphs. *Journal of Research in Science Teaching, 24(4),* 369-383.

MURALIHDAR, S. (1993). The role of multiple data sources in interpretive science education research. *International Journal of Science Education, 15*(4), 445-455.

OFFICE FOR STANDARDS IN EDUCATION (2004). *The Key Stage 3 Strategy: Evaluation of the third year.* London: Office for Standards in Education.

ROTH, W-M., & McGINN, M. K. (1997). Graphing: Cognitive ability or practice? *Science Education, 81,* 91-106.

SADLER, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

SADLER, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in Education, 5(1),* 77-84.

TAIRAB, H. H., & KHALAF Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs? *Journal of Biological Education, 38*(3), 127-132.

WATSON, J. R., GOLDSWORTHY, A., & WOOD-ROBINSON, V. (1999). Practical science investigations in England and Wales: A National Survey. In M. Komorek *et al* (eds.), *Proceedings of the Second International Conference of the European Education Research Association (ESERA)* (pp. 146-148). Kiel, Germany: Institute for Science Education (IPN).

WAVERING, M. J. (1989). Logical reasoning necessary to make line graphs. *Journal of Research in Science Teaching, 26,* 373-379.

WILIAM, D., LEE, C., HARRISON, C., & BLACK, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education, 11*(1), 49-65.

Acknowledgements

This study was part of the author's PhD at the University of London South Bank. I would like to thank all the teachers and their science departments for enabling the school-based work to take place, and of course the children. I wish to thank my supervisor, Professor Stephen Lerman, for his support and help and acknowledge the valuable feedback from Professor Paul Black. The study was financially supported by a PhD scholarship from the Hellenic State Scholarships Foundation (IKY).

APPENDIX I
Examples of Items in the Pre-observation Interview

- How would you describe the sort of difficulties that your children experience with graphs?
- What sort of difficulties do children experience with the interpretation of graphs?
- Why do you think students have such difficulties?
- What do you see as the main purpose of children drawing a graph?
- Tell me what particular aspects of children's graphing you will be targeting at improving. Where are you trying to go?
- How do you think you can get children there?
- What graphing skills will you teach in this investigation(s)?
- Before you get children to work on a task or a project or a piece of coursework, what kind of information do they need to get started?
- How will you find out how well learning targets have been achieved?
- Tell me how you mark children's coursework.
- Do you let children know how you will make judgements about their performance or coursework?
- Do children have the opportunity to mark their own coursework?