International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Science teachers' mission impossible?: a qualitative study of obstacles in assessing students' practical abilities

Per Sund

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
http://dx.doi.org/10.1080/09500693.2016.1232500

Science teachers' mission impossible?: a qualitative study of obstacles in assessing students' practical abilities

Per Sund

School of Education, Culture and Communication, Mälardalen University, Eskilstuna, Sweden

ABSTRACT

Science teachers regard practical work as important and many claim that it helps students to learn science. Besides theoretical knowledge, such as concepts and formulas, practical work is considered to be an integral and basic part of science education. As practical work is perceived and understood in different ways, comparing the results between classes and schools is difficult. One way of making the results comparable is to develop systematic inquiries to be assessed in national large-scale tests. However, introducing similar testing conditions in a laboratory environment is not always possible. Although the instructions and assessment guides for such tests are detailed, many obstacles need to be overcome if equality in the overall test situation is to be achieved. This empirical case study investigates two secondary school science teachers' assessments of 15–16 years old students in three separate groups in the practical part of a Swedish national test in chemistry. Data are gathered using two video cameras and three pairs of spy camera glasses. The results show that individual and independent assessments are difficult due to the social interactions that take place and the physical sources of errors that occur in this type of setting.

ARTICLE HISTORY
Received 18 March 2016
Accepted 31 August 2016

KEYWORDS
Practical work; formative assessment

Introduction

The assessment of students' subject knowledge and practical abilities has increased world-wide (Broadfoot & Black, 2004; Lundahl, Roman, & Riis, 2010). The aim of large-scale tests is to enable teachers to grade more equivalently at the national level (Lundqvist & Sund, in press; Sund, 2016). Theoretical tests are suitable for subject matter knowledge, such as concepts and formulas, for example in chemistry. The challenge lies in how to assess students' practical abilities. This is often done by theoretical tests with questions about students' experiences of and results from practical work (cf. Ofqual, 2015). Another approach is to develop tests that include practical and individual inquiries. In Sweden, the Swedish National Agency for Education has developed a national test that includes practical work. In that part of the test, teachers are expected to conduct a systematic inquiry using detailed instructions about how to assess students' achievements and the support of well-developed assessment guides. In theory, the Agency describes a setting that aims to support and create clear and uniform conditions. However, in reality, questions

CONTACT Per Sund per.sund@mdh.se

have emerged such as: Do teachers manage to assess students' practical abilities? Are the instructions capable of creating comparable laboratory environments and securing the equivalency of the assessments? What are the obstacles for teachers in the assessment of practical abilities?

Earlier research has shown that conducting practical tests is complicated (Abrahams & Reiss, 2012; Harlen, 1999; Ottander & Grelsson, 2006). This case study focuses on the obstacles that teachers encounter when assessing students' practical abilities during a test in a laboratory environment. Many of the things that are often taken for granted in an ordinary classroom environment can create a bias about how testing is 'usually' conducted. This type of environment bias can exist in, for example, the instructions for students how to conduct a systematic inquiry. The assessment guidelines presume that students choose laboratory equipment independently, although earlier research shows that their choices are influenced by different kinds of social group phenomena (Serder & Jakobsson, 2015).

An empirical study of testing in a laboratory environment can make this type of bias more visible and reveal the kind of obstacles that exist for teachers when they try to qualitatively assess students' practical performances.

Conducting empirical studies of practical work in the school laboratory (Högström, Ottander, & Benckert, 2010) is difficult. Gathering data in a confined space is complicated, especially when people are constantly moving around and the room is full of equipment. New technology, such as spy camera glasses, offers new advanced possibilities to study and assess students' practical work in great detail. One major advantage of using spy camera glasses in a laboratory environment is that they are very similar to the ordinary safety glasses (goggles) used in practical work, which means that the novelty value of wearing them is short-lived. Another advantage is that data can be collected with a number of cameras simultaneously, which makes it possible to 'cross-reference' key situations in practical work and make sound recordings of all the people involved in the specific situation.

Purpose

The purpose of this study is to investigate obstacles that prevent teachers to make individual assessments of students' practical abilities in science. The aim is to contribute to discussions about how practical work in general can best be assessed and what the limitations are. This is done using an empirical case study of two secondary school science teachers' assessments of students in three separate groups in the practical part of a Swedish national test in chemistry. Interestingly, it was the different ways in which these three groups chose the laboratory equipment at the beginning of the practical test that inspired the study of the teachers' assessments of students' independent practical work.

The study answers the research question: What kind of obstacles prevents teachers' equivalent assessments of students' individual abilities?

The study highlights a number of practical and social obstacles. These are discussed in relation to equivalent assessment and the alternative ways of assessing students' practical abilities in order to avoid some of the discerned challenges.

Background

Many researchers and educators regard practical abilities as crucial in science education (Abrahams & Millar, 2008; Harlen, 1999). However, in the broader science education community, there appears to be confusion about the actual definition of 'practical work'. This confusion makes discussions about the value of 'practical work' difficult and, at the same time, hard to assess. A number of terms are used to describe practical work, such as 'practical and enquiry skills', 'practical and investigative activities', 'independent enquiry' and 'experimental work' (Dillon, 2008). In the Swedish national test setting, the way science is learned and the scientific process are referred to as *systematic enquiry*. In this article, the term 'practical work' is used, rather than 'laboratory work' or 'experiments', to describe the kind of lesson activity that is in focus. An 'experiment', particularly in the philosophy of science, is generally taken to mean a planned intervention in the material world to test a prediction derived from a theory or hypothesis. However, many practical tasks in school science do not require this (Abrahams & Millar, 2008) and often 'follow a recipe' (Lunetta, Hofstein, & Clough, 2007).

In this study, it is the 'learning experiences in which students interact with equipment and materials or secondary sources of data to observe and understand the natural world' that are in focus (Hofstein & Kind, 2012).

Many in the science education community and beyond regard the practical work that is carried out by students to be an essential feature of science education (Abrahams & Millar, 2008). It is often argued that practical work is central to teaching and learning in science and that good quality practical work helps students to develop their understanding of scientific processes and concepts (Dillon, 2008). There are many other reasons for including practical work in the secondary school science education. Students are motivated by the excitement of discovery (Lunetta et al., 2007), while many teachers believe that science education will consolidate theory and develop manipulative skills and knowledge about standard techniques (Högström et al., 2010). This type of work is understood to include the collection of data and analytical and evaluative theoretical skills, for example a logical and reasoning way of thinking. These arguments support ideas about how students' understanding of how scientific processes work can be developed by, for example, collaborative working and reproducible results (Watts & Wilson, 2013). In general, both teachers and students are positive about 'practical work' (Dillon, 2008).

All skills have to be used in some context and scientific process skills are *only* scientific if they are applied in the context of science. Otherwise, they are general descriptions of logical thinking which are used in many areas of human activities (Harlen, 1999). If these abilities are not properly developed and, for example, the relevant evidence is not collected, the concepts that emerge will not contribute to an understanding of the surrounding world. Thus, the development of scientific process abilities has to be a major goal of science education (Harlen, 1999), not just in terms of preparing future scientists to 'do' science, but to equip people to be 'scientifically literate', so that they are able to make scientifically informed decisions in their everyday lives about global issues.

There is very little evidence from research to suggest that practical work enhances the development of conceptual learning in science. This is generally summarised as: 'Although practical work is commonly considered to be invaluable in science teaching, research shows that it is not necessarily so valuable in science learning' (Clackson & Wright, 1992, p. 40). Hodson (1991) claims that it is necessary to introduce students to the relevant scientific concepts before they undertake practical work if the task is to enhance their learning of scientific knowledge. Abrahams and Millar (2008) question whether observations of a practical task can lead to the development of conceptual understanding. It has also been suggested that practical work could serve as a bridge between the previously taught scientific concepts and the specific observations made in practical work (Millar, Le Maréchal, & Tiberghien, 1999).

According to Abrahams and Millar (2008), despite the widespread use of practical work as a teaching and learning strategy in school science, and the commonly expressed view that increasing the amount of such work would improve science education, some science educators have raised questions about its effectiveness. For example, Hodson (1991) claims that practical work is ill-conceived, confused and unproductive. Osborne (1993) supports this view and states that what goes on in the laboratory contributes very little to children's learning of science. Very few teachers' lesson plans include explicit strategies that help students to make the link between observations and scientific ideas (Abrahams & Reiss, 2012). Teachers need to know what students can realistically be expected to *learn* from practical work and that ideas and explanations do not simply 'emerge' from data.

It is further suggested that teachers should devote more practical lesson time to ensuring that students understand and can use scientific terms that are relevant to the practical tasks (Abrahams & Millar, 2008). This is important, because science also has a key role to play in developing general skills, such as communication, critical thinking, problem-solving and the ability to use and evaluate evidence. Thus, an assessment of the development and achievement of these important outcomes must be included in the assessment of learning in science (Harlen, 1999).

The assessment of practical work is complicated and different ideas abound about the effect that this type of testing has and how it can be made fairer for students. Important elements of the test situation include the general settings, the physical context, the relation between abilities and knowledge and how realistic and interesting the task is for students. Although practical abilities in science are clearly valued and often referred to in the literature as being of central importance, there is a lack of clarity as to what these skills actually are and how they can be properly assessed (Reiss, Abrahams, & Sharpe, 2012).

There is a body of evidence that indicates that practical work is a holistic activity (Matthews & McKenna, 2005). According to Hodson (1991), the emphasis should be on full investigations. Attempts to break practical work down into a programme of discrete skills are likely to miss the essential nature of the activity – where many different skills interact with each other and with the students' knowledge base. Erickson and Meyer (2003) claim that their data show that it is not possible to separate the assessment of unique skills, such as observing and inferring, from the specific content of the task. The assessment of skills is influenced not only by the ability to use the skills, but also by knowledge about and familiarity with the subject matter for which the skills are used. Thus, what is assessed in any particular situation is a combination of skills and knowledge (Harlen, 1999; Roberts & Gott, 2006). The 'setting' of the task also influences performance, in that a school laboratory setting may signal that a particular kind of thinking and acting is required, which may not be the case in an everyday domestic setting. Further, student engagement is a challenge to 'authentic' assessment, which aims to provide assessment tasks that are ‘more practical, realistic and challenging’ than conventional tests (Roberts & Gott, 2006). If student engagement can be fully harnessed, for example by providing tasks that are perceived as interesting and important, the variations in the results when performing systematic tests in a laboratory are likely to diminish.

The National systematic inquiry

The Swedish National Agency for Education offers teachers possibilities to assess students’ different types of scientific knowledge through annual national tests in science. The test consists of a theoretical part (A) and a systematic practical inquiry part (B). In the theoretical part, students are offered possibilities to apply their scientific knowledge by taking part in discussions about social issues. In other words, students are expected to apply their knowledge and to explain the connections between nature, society and their everyday lives. In addition to applying theoretical knowledge, students are asked to conduct a systematic inquiry and demonstrate their practical abilities. The purpose of Part B is to offer teachers an opportunity to *measure*, as it is expressed in the teachers’ instructions, students’ abilities to carry out practical work. The teachers are observing, for example, student’s skills and accuracy in measuring volumes of liquids with different laboratory equipment and their skills to measure how much chemicals they have already added in the experiment. Part B consists of three parts, with a recommended time of 30 minutes for each part. Students plan their individual inquiries in the first part and then conduct them in the second part, with the support of an ‘open’ guide.

They are expected to answer some of the questions, but in general are given plenty of opportunities to do their own planning and take the initiative. The results of their inquiries are evaluated in the third part of Part B. The results are also calibrated into scales and tables and students are asked to explain in detail what happened during the inquiry. Finally, they are asked to suggest how the inquiry could be developed.

The Swedish national tests are classified as secret, and by law can only take place on the dates specified. The test material can sometimes be reused in part after some years have elapsed. This level of secrecy means that this case study cannot provide details about the practical work the students were asked to conduct. The descriptions of the inquiries are therefore vague, although this is not regarded as having a detrimental effect on the findings of the study. The practical part of the systematic inquiry, namely the second part of Part B, is in focus in this study. The aim is to examine three solutions and place them on a scale regarding one by the national school agency preselected physical character. Students have access to the necessary chemicals and laboratory equipment. The teachers mix three stock solutions according to the instructions provided, which the students are then expected to analyse. Students then add certain chemicals to the stock solutions in order to identify the differences between them.

The written instructions and the assessment guidelines for both the teachers and the students are detailed. The teachers’ instructions include information about the various parts of the test, the time frame, the safety equipment, the laboratory equipment, which chemicals are needed and how they should be prepared. In the information the students receive prior to the test, it is stated that: ‘this national test gives you the opportunity to show what you can master in chemistry’. There are three grading levels, described as E, C and A (highest). The assessment guidelines provide teachers with detailed descriptions about what to observe in the students' work. The lowest level concerns safety and the basic handling of the equipment and chemicals. The next level relates to the type of equipment chosen for the measurements. At the highest level, the teacher has to accurately observe how the students actually use the equipment and how accurately they read the measurement scales.

Methods

The study object concerns two teachers and their students when conducting a systematic inquiry in the Swedish national test in chemistry in Year 9 (15–16 years of age). The practical work is carried out by three separate groups – a total of 38 students. The two teachers assist the three groups by providing the chemicals and the equipment and are responsible for the assessment. The teachers both have university degrees as science teachers in Chemistry and Biology and they both have about ten years of teaching experience each at this level. These teachers are experienced and knowledgeable and are considered 'good' teachers by their peers. Data from the three groups engaged in the practical work were collected on three consecutive days, which meant that the conditions for the three groups were comparable.

The two teachers and the three groups had previously taken part in a research project focusing on language in the science classroom. The school itself was selected due to its overall results in science, and these teachers volunteered to participate in a study about science language during practical work. When the author started to prepare the recordings of the first lesson, one of the teachers informed that she has forgot it is the first occasion for her students to make the systematic inquiry part in the national test in science. The author has earlier participated in a research project regarding Swedish national tests in science and signed a secrecy agreement/obligation. After some hesitations and discussions about this agreement on secrecy, the author and the teachers together decided to continue and gather data.

Classroom research in general and especially data recordings are especially of ethical character. Two weeks before the classroom recordings were planned to be conducted, the teachers informed their students about the research study and the gathering of data. In this case, students' guardians are not involved since in a Swedish research context students aged 15 years and above are entitled to make their own decision about whether to attend the study or not. At the beginning of each recording session, all students and teachers were asked about if they agreed to be video recorded and they all gave their active consent to be part of the study.

This study follows the most recent recommendations from the Swedish Research Council for the Humanities and Social Science, Good Research Practice (2011) in collecting and handling data. Consequently, all the participants involved in the research project were informed that these ethical principles and regulations are followed. In addition, the participants were provided with the overall plan, aim and how the empirical material will be used. They were also informed about the right to, at any phase of the research process, terminate their participation without giving any reasons therefore. In these three recording sessions, no participants asked to be excluded in the video recordings.

The data are collected in each group using two mounted video cameras and three pairs of spy camera glasses for a period of 60 minutes. The recorded data amount to approximately 15 hours. One of the video cameras is mounted at the front of the laboratory to cover the area in which the chemicals and equipment are placed and offered to students. The other video camera is mounted at the back of the room. Spy camera glasses are glasses with plain glass lenses and a small camera and microphone (looking like two dots) on the frame. The spy camera glasses film everything the students look at, write and say. The quality of the pictures and sound is good. Three pairs of spy camera glasses are used on each occasion, with students volunteering and three chosen by random to wear them. Both boys and girls are wearing glasses in each group. From an ethical point of view, it is an advantage to let them individually volunteer to have their actions recorded.

The mounted cameras take overall pictures, while the glasses catch the details and conversations. The data are gathered in a test situation, although brief and quiet conversations are captured by the microphones mounted on the glasses. The spy camera glasses capture the activities of the individual students and those of their neighbours. The video cameras record what different students are doing, in what order, how they work with the equipment, who they talk to and where the teachers are in the laboratory. The spy camera glasses help to make many of the important quality details described in the assessment guides visible, such as the handling of the equipment.

The students volunteering to wear the glasses are informed about their function and asked to leave the glasses on their work benches if they leave the laboratory. From an ethical point of view, it is possible to imagine that the researcher can receive inappropriate information especially using spy camera glasses, for example, that the students' view 'freezes' on details of more personal character. In these data, there is no information which can be considered to be violating the integrity of anybody involved.

Analytical method

The data are studied repeatedly during the analytical work. It is important to compare the data from the different cameras collected in the same time period. Three specific analytical questions are formulated in order to answer the research question in more detail:

(1) In which situations do students' social interactions become visible?
(2) What prevents the teachers from making equivalent assessments?
(3) What kinds of physical factors interfere with an equivalent assessment?

The data are analysed using an interpretive and iterative methodology (Kvale & Brinkmann, 2009; Wolcott, 1994). Three categories and three sub-categories are discerned in the analysis. The analytical cycle is repeated until the categories remain firm.

The teachers' and students' expressions and actions during the practical work are identified by the author and divided into three categories of 'practical obstacles' observed in the test situations. The categorisation is later reviewed by other researchers in the larger research project in order to anchor the case study by co-coding. This procedure is used to ensure that the categories are accurately represented in the video-recorded material. This method is similar to grounded theory (Strauss & Corbin, 1998), in which an open approach to empirical data is important and does not force a matrix of predetermined categories on the data.

Results

As the study is conducted in a test situation in which students are not supposed to talk to peers or teachers, it does not include examples of teachers' and students' utterances. Three descriptions of the situations of the three groups are presented below in an attempt to offer the reader a richer picture of the context. As these descriptions are interpretations of the reality of the situation, they are presented as initial results.

Group 1 (16 students + 2 teachers)

During this test, students regularly move around in the laboratory. The students are also involved in quiet discussions and look at each other's notes. Initially the teachers deal with the chemicals and the equipment and also take care of spillages and accidents with the glassware.

Using the instructions provided for this part of the test, the teachers tell the students what needs to be done. The students write a plan for the inquiry (Part I) but they will from now on use the prepared optional instructions. They assemble the safety equipment and put on safety aprons and safety glasses. The teachers prepare the necessary equipment, chemicals and solutions that the students will examine in accordance with the test instructions. Everything is then placed on the teacher's desk, with a table on wheels beside it. One girl, who has just finished putting on her safety apron and glasses, distributes the equipment – a set of test tubes and a test tube rack – to the other students. Very few of them use the test tubes during the test and instead mix the solutions in 100 ml beakers. Six to eight students stand beside a bench and pour the solutions into the beakers. The teachers stand beside the whiteboard and observe what is going on. When the students go to the teacher's desk to collect the test samples of the three stock solutions, they pour them directly from the 3000 ml beakers marked A, B and C into three smaller beakers. Some of the students also mark their small beakers A, B and C, but not all. If they accidentally pour too much of the test solution into the small beaker, they pour the excess liquid back into the stock solutions. The students pour the liquids back and forth between the vessels and spillages occur, which the teachers then mop up. Boxes of glassware are placed next to the solutions.

While one of the students wearing spy camera glasses is waiting for his turn to collect the chemicals, he asks the teachers *is this the way to do it?* The teacher next to him answers – *this is one way of doing it*. The student turns around to face the other teacher and asks – *is this the right way to do it?* The teacher does not answer, but when the student continues to look straight at him, he starts to smile. In a loud voice, the student then tells the teacher that *everybody is doing it wrong!* He looks first at the teacher and then at his peers around the table. They look up quickly at the teachers and then continue their activities.

The data from the spy camera glasses show that when the students collect the chemicals, they ask questions about different practical issues, for example: Is this A? How much do I need? Should I use this chemical? Are there any beakers left? It takes about ten minutes for all the students to collect the equipment and pour the test solutions. During the practical work, the students move back and forth between their benches and the desk to collect more equipment (pipettes, glass rods, beakers, measuring cylinders, etc.) and chemicals. After a while, when the students start to focus on their own independent work, the teachers start to move around the laboratory and talk individually to them. The students ask the teachers questions about what to do, the number of chemicals needed and the presentation of results in tables and figures. The teachers avoid giving direct answers or revealing information about the inquiries. During the test, the students chat to each other and sometimes walk over to their peers' work benches for more direct talks. This usually happens when the teachers are not immediately on hand.

After a while, two students collect two measuring cylinders located next to the small beakers on the desk. They are the only ones in the group to use measuring cylinders. The student who earlier expressed that everyone was doing things wrong is the only one in the group to use the measuring cylinder accurately according to the assessment guide. He measures all the solutions at his workplace with good precision. His bench peer also collects a measuring cylinder, but simply puts a pipette into it and nothing happens. Meanwhile, when the student makes the correct volume measurements, one of the teachers has temporarily left the laboratory and the other is talking to a student at the other end of the room.

One of the teachers has an observation protocol in his hand. He walks around, looks at and chats to the students, but sometimes leaves the room to make notes. Periodically, the two teachers walk around the room together, stop for a little while, talk to each other for about 30–40 seconds and then continue walking around for a few more minutes. By this time, the students are coming to the end of the practical part of the test and are writing and presenting their results in tables. When they have finished, they clear everything away and wash the equipment.

Group 2 (12 students + 2 teachers)

In general, this is a much calmer laboratory situation than that of Group 1. The students sit at their own workbench and occasionally discuss things with their peers.

The equipment and the stock solutions are placed on a desk and roll-table at the front of the laboratory. The solutions are exactly the same as those used by Group 1. It can be observed by the markings on the three large beakers and that the levels of stock solutions are now slightly lower. The students all wear safety equipment and one of the boys is working with three small beakers. He starts to transfer the solution from one of the big beakers to a small one with a pipette. Seven or eight students stand around the table with the chemicals on and chat while transferring the solutions. It takes several minutes to transfer the solutions with the pipettes. At frequent intervals they put the pipette they have been using on the table in-between the beakers and sometimes absentmindedly pick up a pipette that has been used for a different stock solution. When most of the students have finished, one girl starts to pour the solution from one of the big beakers into a smaller one. A similar thing happened in Group 1. The excess solution in the small beaker is then poured back into the big beaker containing the stock solution.

A boy in this group takes the laboratory equipment first. When he approaches the desk, the beakers are easily accessible. All the students in this group use beakers, although measuring cylinders are also available next to the solutions. One girl later uses a measuring cylinder to transport a chemical to add to the solution for examination, but does not use the cylinder to measure the volume. Five other students do the same thing and some use test tubes to transport additional chemicals from the desk to their workplaces.

While the students are engaged in their laboratory work, the teachers walk around the room and chat with them. Now and again the two teachers pause at the back of the room to observe and exchange information about the activities that are taking place.

It is calm and the only sound that can be heard is that of chinking glass as students stir the solutions in the glass beakers.

A girl says to one of the teachers – *I do not understand.* The teacher answers – *I am sure you do, you can*. The teacher moves towards her. A boy is sitting in front of the girl who tells the teacher that he has poured the same amount of chemicals into three small beakers [he should have added different amounts]. The teacher asks him how he knows this. He answers that he became aware of this when pouring the solution from three test tubes into the beakers with approximately the same amount of adding solution. The girl sitting behind him, wearing spy camera glasses, has done the same thing and overhears the conversation between the teacher and the boy. She asks the teacher – *can I start from the beginning?* The teacher looks at her watch and says – *you will have time enough*. The boy and the girl quickly wash the equipment and start all over again. The teachers once again stand at the back of the room and discuss what has happened. The two students in question are the last to leave the room, but have nevertheless managed to do everything in the allotted time. In general, the students in this group work more independently and are much calmer than those in Group 1.

Group 3 (10 students + 2 teachers)

Initially, one of the teachers shows the students where the chemicals are and tells them what they are used for. The other teacher asks them to read the instructions carefully before beginning the practical work. Safety glasses and aprons are placed on the first row of benches near the door, with boxes containing small beakers near the window. In general the laboratory is calm, apart from when two boys bang into each other and almost knock over the three stock solutions to be examined. All the materials and chemicals (apart from the small beakers) are available on the desk and roll-table used earlier by the other two groups. As some of the stock solutions have already been used by the other groups, the levels in the beakers have gone down.

The first student to collect equipment is a boy who picks up a measuring cylinder, weighs it in his hand and goes back to his bench. This move results in everyone in the group taking a measuring cylinder back to their work place, although in the end no one actually uses them. The measuring cylinders are placed in an accessible place on the desk, whereas the test tubes are in racks on the roll-table beside it. None of the students in this group pick up or use the test tubes. A second student picks three beakers from the box by the window and begins to mark them with the letters A, B and C. Some of the other students then discover that many of the beakers are similarly marked and put them in alphabetical order. From now on the students use only the small beakers to collect the stock solutions and the additional chemicals, which leads to a shortage of beakers. The students, therefore, start to look for new ones in the sink and on the drying rack.

One student starts to transfer the stock solution to a small beaker with a pipette, but realises that a peer opposite him has started to pour the solution into the small beakers. He stops using the pipette and also starts to pour the solution into his beakers. He then realises that his beakers are wet, probably because they have been used for earlier experiments, and pours the excess liquid onto the floor. He tries to dry the beakers with his fingers and then wipes his wet fingers on his jeans. Many of the other students also pour the stock solutions into small beakers and pour the excess back into the big beakers. Three to four students stand beside the desk and pour liquids into different containers. The teachers, standing only a few metres away, observe these activities.

A group of five students who are engaged in practical work talk quietly together and show each other their solutions and written notes. They communicate without the teachers being aware of it. A teacher standing a few metres away with an observation protocol observes their solutions and makes notes.

The other teacher asks a girl about her success so far. She says that she is uncertain about the result and where it should be written in the papers. The teacher points to the papers and shows her where to write the different amounts of added solutions. The girl, who is wearing spy camera glasses, has not added the different amounts in a controlled way. She looks at her notes, writes something in the test papers and completes the task.

After a while, a student wearing spy camera glasses realises that he has done something wrong, because his solutions do not change in the expected way. He adds more chemicals but still nothing happens. He has mixed the same amount of adding chemical to the stock solutions on his bench. One of the teachers sees the student searching through the instruction sheet and asks – *What are you looking for?* The student replies: *I am looking for what has gone wrong*. The teacher asks – *Why do you say that?* The student responds by saying – *I added too much of this chemical at the beginning* and shows what he has done. The teacher asks – *If you were able to re-do it, what would you do?* The student explains what he would do and asks the teacher if he has time to do it again. The teacher says that he has and the boy manages to accomplish the task before the other students have finished. The teacher approaches the researcher filming in the room and tells him that is difficult not to let students try again if their first attempt has failed. The teacher says that he does not like students to feel cheated and that it is important to be honest and tell them that they are either right or wrong. The teacher is aware that their dialogue has been recorded by the student's spy camera glasses.

Summary of the results

The analysis of the data resulted in the identification of three main categories of obstacles for the teachers in their attempts to equally assess students' independent practical work abilities: (1) students' interactions with each other, (2) teachers' interactions with the students and (3) systematic physical sources of error.

(1) Students' interactions with peers

In the three different practical work situations, the students' social interactions become obstacles for the assessment of their independent work.

(a) During the selection of laboratory equipment

On all three occasions, a group phenomenon arises in connection with the selection of equipment. The equipment that is selected first differs in the three groups. The students who choose the equipment first set the trend for the rest of the group. In general, the first student to choose equipment does not make a reflected choice, but instead grabs the item that is closest to him or her.

(b) During the handling of the chemicals

The students communicate even though it is a test situation. They talk in low voices and confirm each other's comments, often with glances and smiles.

They discuss how they should pour the stock solutions into the smaller beakers. At all three sessions, the students crowd around the table with the chemicals on and use the equipment in similar ways. If a student in one group starts to use a pipette to transfer the solution into the small beakers, the other students soon copy this action. In another group, a student starts to pour the solution into a small beaker, which then leads to everyone else doing the same.

(c) Working at their benches

On many occasions students find themselves at the same stage in the test or laboratory instructions. They take small breaks, look at each other and start the next step more or less simultaneously. They drip solutions and stir the mixed solution in the small glass beakers with glass rods. A rhythmic sound of tinkling glass can sometimes be heard when the solutions are stirred, especially when many students are doing it at the same time. Another example of the students' parallel work is when they observe the changes in their peers' solutions. The students also work through the laboratory instructions in a parallel way and thereby influence each other.

(2) Teachers' interactions with students

Three different types of problems emerge in the data when it comes to teachers making equivalent assessments of students' practical work abilities. One is of a physical nature, while the other two are the consequences of interaction between teachers and students.

(a) It is difficult for teachers to see exactly how students handle the equipment

As a school laboratory is often a confined space furnished with high benches, it is not always easy to see what individual students are doing and how they are working. Students also constantly move back and forth to collect chemicals and equipment.

The result is that teachers are often prevented from making long observations of students' performances. Lots of different parts of the inquiry need to be observed, which means that teachers need to be in close proximity to the students. For example, they may need to look at a scale on a measuring cylinder to determine the accuracy of the measurement. Sometimes students take different pieces of equipment to their work bench that they then do not use, simply because 'everybody else' selected them at the beginning of the practical test. Sometimes the teachers are active, and ask students what they are doing and about their plans for next step. However, in these brief moments, the teachers are only able to see parts of the investigation.

One example of the problem of assessment is a specific situation that was observed through a student's spy camera glasses when using a measurement cylinder. At that particular moment, one of the teachers had temporarily left the room and the other was positioned at the opposite end of the room. As the teachers were unable to observe this particular part of the test, they were unable to grade the activity. According to the data, this is the only occasion in which this practical ability was demonstrated in all three groups.

(b) It is difficult for teachers to avoid scaffolding

When the teachers move around the room for observation and assessment purposes, they rarely have time to ask students open and analytical questions. Instead, the students often ask the teacher practical questions, such as where to find more beakers and where waste chemicals should be disposed. Sometimes students ask questions related to the test, which means that the teachers have to be careful not to give hints or scaffold in a particular direction.

Sometimes the teachers respond to questions with a glance or a smile to confirm that the student is on the right track. On some occasions, students' questions relating to the instructions or about where they should write their answers or tables lead the teachers to give more information than they may have intended about the expected results.

(c) It is difficult for teachers to avoid in situ assessments

A couple of times during the data collection, the teachers approached the author and said that they had sometimes had to instantly 'mark' the task in hand verbally. They did not say that something was right or wrong, but that after some discussion the teacher and the student had agreed that the student should start again. If a student was completely lost in the inquiry, it was not possible to adjust or scaffold him or her in a better direction. For example, they found it difficult to respond to a question about adding more chemicals if too much had been added from the beginning. The point of no return had been reached. The teacher could, of course, simply fail the student; but if the student realised that something was wrong, then a decision had to be made. Students often solved this by asking whether they had time to begin again, and if there was, the teacher was supportive. In test situations, it is natural that teachers want their students to pass.

(3) Systematic physical sources of error

Three categories of physical sources of error are noted as obstacles for an equivalent and independent assessment. The two first categories relate to the handling of chemicals and equipment, while the third is about the access to material resources that can vary between schools. The first is an obstacle to equivalent assessment because it could lead to wrong results in the inquiry. The second makes the assessment dependent on others, while the third can be an obstacle for schools conducting the practical test.

(a) Incorrect analytical behaviour reduces equivalency

Prior to the systematic inquiry, the teachers' instructions are detailed but do not include all the possible practical variations. They may or may not talk about the need to mix new stock solutions and check the other chemicals before every test. In all the three groups, each student pours the excess solutions back into the stock solutions that the teachers mixed before the test began. In most cases, the students pour the A-solution back into the A stock solution, but sometimes the A-solution is poured into the B-stock solution. As a result, the stock solutions for groups two and three are contaminated already after the first test occasion.

(b) Marking by pen left on the equipment reduces independency

According to the assessment guide, in order to pass a particular part of the test, the equipment has to be carefully marked so that the stock solutions can be analysed and a record kept of the results. When groups two and three start their work, all the small beakers in the storage box are marked by pen. Students are also guided by the kind of equipment that has been used by other groups and is still wet.

(c) Possible shortage of equipment decreases equivalency between schools

A systematic inquiry and equivalency require that every student has access to a well-functioning, spacious laboratory and the specified equipment. If the space in a laboratory is limited and crowded, reaching the highest analytical quality set out in the assessment guides will not be possible and students will, therefore, be disadvantaged. For example, in the three test situations outlined above, there was a shortage of small beakers, which meant that the students needed to look for other types of vessels in which to transport and mix the chemicals. They therefore started to wash used beakers and used test tubes and measurement cylinders for the chemicals. It can be ascertained that if schools are unable to provide the necessary equipment, students will be unable to perform at the highest quality levels.

Discussion

From the Swedish science teacher's perspective, conducting a systematic inquiry can be described as 'mission impossible'. Even though teachers plan and work in accordance with the detailed instructions provided, in the test situation different kinds of obstacles for assessing students' practical abilities arise and have to be negotiated. The teachers cannot always see or have the time to capture all the details outlined in the assessment guidelines. In addition, it is difficult for teachers and students to avoid interacting socially and communicating when moving around the laboratory, which makes individual independent assessment difficult or almost impossible.

According to Reiss et al. (2012), practical skills can be directly assessed, for example by a teacher observing and assessing a student carrying out a titration, or indirectly assessed skills, for example by a teacher assessing a report written by a student who has undertaken such a task. If the intention is to determine students' abilities to undertake specific practical tasks, then direct assessment is more appropriate. Conversely, if the intention is to determine the understanding of a skill or a process, indirect assessment is more appropriate (Reiss et al., 2012). Earlier research recommends indirect assessment, mainly due to time constraints and costs (Dillon, 2008; Roberts & Gott, 2006). This study shows that direct assessment is complex.

The practical test situation is complex

Many readers might think that the teachers in question could have managed the test situation differently, for example by being more explicit and telling the students that talking to each other was forbidden during the test. A teaching approach like this could have avoided some of the obstacles to equal assessment. However, if the two teachers had focused on creating a silent test environment, the time set aside for observations and the assessment of practical abilities may have been eroded. At the end of the day, it is up to each individual teacher to determine the rules and environments for the conduct of tests. The obstacles identified in this study are not mentioned or commented on in the instruction materials. One way of understanding the lack of information about classroom management is that there may be a 'bias' in the test situation. The Swedish National Agency for Education expects a practical test in a laboratory will be conducted in a similar way to a written test in an ordinary classroom. It is difficult to create a practical work test situation if there is no communication. In a laboratory environment, students are constantly moving around and often do similar things simultaneously. In order to offer students a test situation in which they can work individually, independently and show their real individual abilities, a suitable room needs to be available, such as individual cubicles with a desk and the necessary equipment to choose from. The cubicles would also need some kind of camera surveillance so that teachers could see what was happening in detail. Such a setting would, of course, be costly. However, in the setting studied here, the confined laboratory environment and the small number of students (12–16) create a sense of solidarity. They are all in the same stressful test situation, which makes it almost inevitable that they will start to interact socially and communicate.

Even if the teachers managed to create a silent test environment, students' social interactions would probably continue, for example when choosing equipment, conducting the inquiry and working in parallel. These social interactions would also be obstacles to an individual and independent assessment.

In this environment of solidarity, it is difficult for teachers to avoid verbal scaffolding and giving students more information than they perhaps need. This may be unintentional, and could be due to the fact that teachers feel the need to care for their students in a vulnerable situation like this. For example, this is clearly visible when students have misjudged their inquiries and are in danger of failing. The students themselves realise that they are not doing the same things as their peers and are not on the same parallel track. In situations like this, teachers are often put in the position of having to give a student instant feedback – is it right or wrong? The teachers thus 'mark' this situation by agreeing with the student that it would be better to start again. Earlier research shows that teachers often show that they care for their students because they want them to be as successful as possible (Sund & Wickman, 2008, 2011). The results of this case study show that the teacher's care for his or her students resulted in agreements to redo the inquiry and pass the test.

The assessment guidelines can also be problematic when it comes to qualitatively and unequivocally ranking students' practical abilities. The progression of the qualitative abilities described in the guidelines can be quite different in students' practical work. Sometimes, a student might demonstrate abilities that warrant a high grade, for example when stirring the chemicals. However, in the next moment, the same student might not use the measuring cylinder in the way the assessment guidelines indicate. Thus, the assessment of a student can vary between the highest and lowest levels depending on what is being examined.

In the study, the complexities of social interaction in a practical test are made visible in many different ways. In general, the systematic inquiry focuses on one basic chemistry concept that students may or may not be familiar with. Earlier research claims that conceptual understanding is important when judging practical abilities in that way and that the possible lack of conceptual understanding can become an obstacle (Harlen, 1999). In other words, there has to be a full conceptual understanding of the investigation being conducted in order to perform well. In the three groups, the conceptual understanding appears to have been weak, in that all the students poured the excess solutions back into the stock solutions prepared by the teachers. According to the recommended chemical and analytical practices, this is prohibited. It is therefore obvious that these groups had not learned to conduct practical work in an analytical way. The pouring back of chemicals is not mentioned in the assessment guidelines, although in a systematic investigation, and with a basic knowledge of analytical work, this is an important aspect of chemistry.

Alternative ways of assessing practical work

The results inspire us to look for alternative ways of assessing students in laboratory tests and how to avoid some of the obstacles indicated in the study. One way of making social interaction a fruitful part of the test would be to allow the students to conduct different inquiries in the same room. Many of the obstacles to teachers' equivalent assessments could thus be avoided. Instead, students' communications could be seen as an important support for each other in their individual tasks. In this type of setting, being allowed to talk chemistry would aid understanding and learning.

Another way of supporting students and encouraging them to talk chemistry would be to allow them to do the practical exercises in small groups. This would also mean teachers having only four to five groups to communicate with and assess, instead of observing up to twenty or so individuals. The individual and independent part of the assessment would be lost, but the teacher would gain a good overall picture of the students' knowledge about chemistry, and especially their practical work. Given that employers value skills such as teamwork, using practical work in science to assess students' collaborative and individual skills may be more appropriate (Reiss et al., 2012). Teachers' discussions with small groups of students about their methods of enquiry would also be helpful for reflection purposes (Harlen, 1999). However, this could result in a loss in the standardisation and comparison of individual results. Questions about the purpose of assessment remain, such as what is the main point of the assessment and what about standardisation and learning?

There are other ways of standardising the assessment of students' practical abilities than national tests. For example, in the UK the content requirements for the general certificate of secondary education (GCSE) in science (Ofqual, 2015) specify that students must carry out a minimum number of practical activities (8 in single science subjects and 16 in combined science). Although students are not assessed on these activities, they need to document them and schools have to inform the examination board that they have enabled students to carry out the full range of practical work specified. At the end of the science course, students are required to demonstrate their understanding of scientific activities in written examinations relating to the required list of activities, and 15% of the marks are dedicated to this. This type of assessment puts a great deal of responsibility on individual schools to both fulfil the criteria and assess students' performances. However, in the UK this change towards a more continuous assessment, or formative assessment, is controversial for stakeholders like the Nuffield Foundation (Foundation, 2015), who are concerned that scientific quality could be reduced by less centralised standardisation.

Science process skills are central and must be included in a formative assessment (Harlen, 1999). There are also different views about how to assess students' practical work properly. The creation of a one-off, separate test like the Swedish national test to assess practical abilities affects the reliability of the assessment. The problem of the length of time could be tackled by embedding the tasks in regular everyday activities, although this also has drawbacks in terms of standardisation. The combination of continuous assessment and well-designed practical tasks for direct observation could be a good compromise for the assessment of practical skills (Harlen, 1999). The annual Swedish national test aims to be holistic and embrace the whole process from planning, conduct and evaluation, although the practical and social obstacles encountered in the laboratory environment risk lowering the validity. The obstacles to individual and independent assessment highlighted in this study show that, at present, teachers assess the practical abilities of an entire group of students.

Conclusions

It is easy to argue that the results of this empirical case study may have low validity, although it is possible to make some generalisations (Flyvbjerg, 2006). The two teachers in question do not represent every teacher's conduct in this particular part of the national test. However, the results can be generalised in the sense that all students in Sweden carry out similar practical work and that obstacles can arise in such situations. These obstacles may vary between different teachers and schools, but are probably present to some extent. Many science teachers in Sweden will probably recognise some of the obstacles.

Nevertheless, despite the limitations of a case study, there is still reason to believe that different types of obstacles in the assessment of practical work need to be investigated further. It is not difficult to see that the setting of this studied systematic inquiry creates an artificial and social test situation for all the actors involved. One thing is certain – more and new attempts will be made to measure school activities in the near future. A useful and reflective question to ask is: Do we value what we measure or do we measure what we value? (Biesta, 2009)?

There may, of course, be a true desire among teachers to measure individual and independent practical abilities. This leads to further questions about measurement: Why is it done (learning, standardisation and tax payers' guarantee) and for whom (student, teachers, education system and other societal stakeholders)? From the findings of this study, the desire to standardise and compare grades between classes, schools and municipalities can be questioned. Attempts at standardisation seem to be ineffective and often have low validity. If students' science learning is prioritised, rather than measurement, good assessment methods with which to discern students' needs for teaching and other support in the learning of science can be sought. The teachers in this study are very relaxed and gentle in the practical test situations and communicate with their students when necessary. Based on their everyday long-term experience, they know which practical work abilities their students should have acquired. Even so, they said that they would appreciate in-service training on the assessment of science education.

Acknowledgements

I would like to thank the members of the research project SONAT (language and science education) at Malmo University for their important financial and scientific support. The research group of Studies of Meaning-making in Educational Discourses, SMED, at Orebro University has also provided invaluable help in the structuring and formation of the scientific results. Financial support for the study was provided by Mälardalen University and Södermanland's County Administration Board, both of which are based in Sweden.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, *30*(14), 1945–1969.

Abrahams, I., & Reiss, M. J. (2012). Practical work: Its effectiveness in primary and secondary schools in England. *Journal of Research in Science Teaching*, *49*(8), 1035–1055.

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, *21* (1), 33–46.

Broadfoot, P., & Black, P. (2004). Redefining assessment?: The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice*, *11*, 7–27.

Clackson, S. G., & Wright, D. K. (1992). An appraisal of practical work in science education. *School Science Review*, *74*(266), 39–42.

Council, S. R. (2011). *Good research practice*. Stockholm: Swedish Research Council.

Dillon, J. (2008). *A review of the research on practical work in school science*. London: King's College.

Erickson, G., & Meyer, K. (2003). Performance assessments tasks in science: What are they measuring. In B. J. Fraser, & K. G. Tobin (Eds.), *International handbook of science education* (Vol. Part 2, pp. 845–865). Dordrecht: Kluwer Academic.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245.

Foundation, N. (2015). Plans to remove direct assessment of practical skills from GCSE science have 'great potential for damage' [Press release].

Harlen, W. (1999). Purposes and procedures for assessing science process skills. *Assessment in Education: Principles, Policy and Practice*, *6*(1), 129–144.

Hodson, D. (1991). Practical work in science: Time for reappraisal. *Studies in Science Education*, *19*, 175–184.

Hofstein, A., & Kind, P. M. (2012). Learning in and from science laboratories. In B. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 189–207). London: Springer.

Högström, P., Ottander, C., & Benckert, S. (2010). Lab work and learning in secondary school chemistry: The importance of teacher and student interaction. *Research in Science Education*, *40*, 505–523. doi:10.1007/s11165-009-9131-3

Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage.

Lundahl, C., Roman, H., & Riis, U. (2010). Tidigt ute med sena betyg – sent ute med tidiga! Svensk betygspolitik i ljuset av internationell betygsforskning och betygssättning i Europa [Early with late rating – late with early! Swedish assessment policies in light of international assessment research in Europe]. *Pedagogisk forskning i Uppsala 157*. Uppsala: Uppsala University.

Lundqvist, E., & Sund, P. (in press). Selective traditions in group discussions – teachers' views about good science and the possible obstacles when encountering a new topic. *Cultural Studies in Science Education*.

Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Teaching and learning in the school science laboratory. An analysis of research, theory and practice. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 393–431). Mahwah, NJ: Lawrence Erlbaum Associates.

Matthews, P. S. C., & McKenna, P. J. (2005). Assessment of practical work in Ireland: A critique. *International Journal of Science Education*, *27*(10), 1211–1224.

Millar, R., Le Maréchal, J. F., & Tiberghien, A. (1999). 'Mapping' the domain: Varieties of practical work. In J. Leach & A. Paulsen (Eds.), *Practical work in science education – recent research studies* (pp. 33–59). Roskilde/Dordrecht: Roskilde University Press /Kluwer.

Ofqual. (2015). *Assessment of Practical Work in New Science GCSEs – Summary*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/408513/2015-03-03-assessment-of-practical-work-in-new-science-gcses-summary.pdf

Osborne, J. (1993). Alternatives to practical work. *School science review*, *75*(271), 117–123.

Ottander, C., & Grelsson, G. (2006). Laboratory work: The teachers' perspective. *Journal of Biological Education*, *40*(3), 113–118.

Reiss, M., Abrahams, I., & Sharpe, R. (2012). Improving the assessment of practical work in school science *Leading Education and Social Research*. London: Institute of Education University of London and The University of York.

Roberts, R., & Gott, R. (2006). Assessment of performance in practical science and pupil attributes. *Assessment in Education: Principles, Policy & Practice*, *13*(1), 45–67.

Serder, M., & Jakobsson, A. (2015). "Why bother so incredibly much?": Student perspectives on PISA science assignments. *Cultural Studies in Science Education*, *10*, 833–853. doi:10.1007/s11422-013-9550-3

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand oaks, CA: Sage.

Sund, P. (2016). "Discerning selective traditions in science education: a qualitative study of teachers' responses to what is important in science teaching." *Cultural Studies of Science Education*, *11*(2): 387–409.

Sund, P., & Wickman, P. O. (2008). Teachers' objects of responsibility – something to care about in education for sustainable development? *Environmental Education Research*, *14*(2), 145–163.

Sund, P., & Wickman, P. O. (2011). Socialization content in schools and education for sustainable development. I. A study of teachers' selective traditions. *Environmental Education Research*, *17* (5), 599–624.

Watts, A., & Wilson, F. (2013). *The assessment of practical science: A literature review Cambridge assessment*. Cambridge: Research Division, Assessment Reserach and Development Cambridge Assessment.

Wolcott, H. F. (1994). *Transforming qualitative data: description, analysis, and interpretation*. Thousand Oaks, CA: Sage.