Objective and continuous assessment of student performance in the physics laboratory

Citation for published version:
Ganiel, U & Hofstein, A 1982, 'Objective and continuous assessment of student performance in the physics laboratory', *Science Education*, vol. 66, no. 4, pp. 581-591.

Total number of authors:
2

Published In:
Science Education

(article begins on next page)

Objective and Continuous Assessment of Student Performance in the Physics Laboratory

URI GANIEL and AVI HOFSTEIN
Department of Science Teaching, The Weizmann Institute of Science, Rehovot 76100, Israel

Introduction

The laboratory has long been used in science instruction to involve students in concrete experiences with objects and concepts.

With the advance of new curricula in the 60s and 70s which stress the process of science and emphasize the development of higher cognitive skills, the laboratory has acquired a central role. Some science educators have even gone so far to regard the laboratory as "... not just as a place for demonstration and confirmation but rather as the core of the science learning process" (Shulman and Tamir, 1973). Once the central role of the laboratory in science instruction is recognized, the problem of assessment of student performance in the laboratory becomes one of the utmost importance.

Types of Assessment of Laboratory Work

Several different styles of assessment of student laboratory skills have been utilized. The more common style is based on written evidence, namely written reports or items on paper and pencil tests (Kruglak, 1958, Grobman, 1970, Doran, 1978, Giddings & Hofstein, 1980, Lunetta et al., 1981). However, since practical work involves abilities, both manual and intellectual, which are in some measure distinct form those used in non practical work (Kelly & Lister, 1969), the ongoing assessment of students during practical sessions should not be overlooked. This argument is also supported by research evidence. It has been shown that the correlation between students' achievement in practical tests and their achievement based on written evidence is rather low. (Robinson, 1969, Buckley, 1970, Tamir, 1972, Ben Zvi et al., 1977). Thus, there is a need to develop special measures to assess not only what the student knows about the laboratory but also what he actually does in the laboratory.

Manual skills and resourcefulness in performing an experiment need to be assessed in an actual laboratory situation. Practical examinations were used in several research studies by Tamir (1974) in biology, and by Eglen and Kempa (1974) and Ben-Zvi et al. (1977) in chemistry. These practical examinations utilized systematic observations based on a list of specific criteria, as opposed to an open ended, subjective type assessment, which currently is very common among teachers and examiners.

Science Education 66(4): 581-591 (1982)

CCC 0036-8326/82/040581-11$02.10

During the school year, even those teachers who employ practical work as part of their teaching often tend to stay away from practical examinations since their implementation is time consuming and difficult. However, practical examinations are in use by several Boards of Examination and in final examinations in several countries (e.g., A-level practical examinations in the U.K. and the matriculation in Israel). In these examinations, students are usually examined by external examiners and not be their own teachers. Such examinations suffer from several drawbacks:

a. In many cases different examiners use different criteria to assess student performance.
b. Examinations are limited to those experiments that can be readily administered to students during a limited time period. This obviously restricts both the scope and validity of the assessment.
c. Since such examinations are difficult to implement, they cannot be conducted very often. Consequently the element of chance is rather dominant. This obviously increases anxiety of the student.
d. Because of administrative constraints, practical examinations will often be administered to a large group of students simultaneously. Consequently, the examiner will not be able to concentrate on observing each student systematically, and will have to rely in his assessment on the results of the experiment and on the written reports.

In an attempt to overcome these drawbacks, in recent years there has been a movement towards implementation of continuous assessment by the teacher in normal laboratory sessions. This has been formalized to some degree in the United Kingdom (University of London, 1977, J.M.B., 1979). In this system of continuous assessment, the teacher unobtrusively observes each student during normal lab activities and rates him or her on specific criteria. The assessments can be recorded for each student over an extended period of time. Normally only a few students will be carefully observed and rated during each activity.

Continuous assessment on several occasions throughout the year is necessary to adequately cover the variety of tasks and skills which comprise a total program of practical work.

TABLE I
Means and Standard Deviations of Percentage Weights Suggested by Teachers ($N = 25$) for the 5 Components of the Assessment List

Component No	Mean	Std. Dev.	Weight used in Assessment
1	17.3	2.95	15
2	25	1.47	25
3	10	2.95	10
4	24.2	2.41	25
5	23.5	3.19	25

Purpose of Study

The present study was undertaken in order to develop and evaluate an instrument for the continuous assessment of the performance of high school students in the physics laboratory. Our aim was to improve the quality of this assessment. By improvement we mean an increase in *objectivity*, in the sense of achieving a better consensus among teachers when evaluating a particular student. Secondly, we hoped to increase their *precision*, when assessing different students who may differ in their abilities in various components of the performance in the physics laboratory. Clearly, in such a situation, the personal preferences and biases of each teacher play a major role, and different teachers may tend to emphasize different components in their assessment. By convincing all teachers to use the same criteria, one expects to increase the precision of their assessment.

Research Method

Subjects

150 physics teachers in Israeli high schools (grades 9–12) were contacted and asked to participate in the experiment. All were experienced teachers, who have taught physics in high school for a number of years and performed part of their instruction in the laboratory. About half of these teachers agreed to participate and out of these, 25 were randomly chosen. These constituted the group with which this project was undertaken.

Research Instruments

Video films and student's reports: As a preliminary stage, we decided to create a situation in which all the participants would be able to observe and assess the same laboratory exercises, performed by the same students. In order to achieve this aim, it was important to satisfy the following criteria as closely as possible:

(1) All teachers should watch the performance of the same student performing a lab exercise.
(2) Every teacher should be able to watch the student closely.
(3) The student should work in an environment similar to that of a regular lab session in school.

We therefore chose to make video films, each of which showed a student performing a physics experiment in a school laboratory. While the films did not fully simulate the situation in school, where students normally work in pairs or even in larger groups, this method was judged adequate for our requirements. Three films were made of experiments taken from the regular physics course (Physics Experiment for High School, 1975) taught in Israeli high schools (11th and 12th grades). We chose experiments with which all teachers were familiar, and which contain many elements which could be assessed (Appendix I). The students' teachers were consulted, so that each of the students performed an experiment he had not done before, but for which he had the necessary theoretical preparation. Each student received the relevant physics text book and the lab guide which describes the experiment and explains what he was expected to do. All the equipment necessary was available on the bench, and a teacher was on hand to answer any questions the student wanted to ask before he started to work. The student was asked to talk while working and explain what he was doing, much in the same manner as he would talk to his partners in a regular lab session in school.

After performing the experiment and recording his data, the student was asked to write a lab report. These reports, in the students' handwriting, were photographed and later used by the teachers in their assessment.

Assessment Scheme

The central part of the investigation consisted of testing the use of a common list of criteria for assessment by all the teachers.

We surveyed the literature for similar lists, both in physics and in other natural sciences (Mathews, 1969, Tamir & Glassman, 1971, J.M.B., 1979). A questionnaire was distributed to Physics teachers in Israel asking for their opinions about the importance of various components of the work in the physics laboratory.

After analyzing the answers received, we interviewed a selected group of experienced teachers and discussed the opinions reflected in these answers. Utilizing all the above information, a list was made up, which consisted of the following five components:

1. Construcing the experimental setup and other manipulative skills.
2. Observing and measuring.
3. Ordering and organizing work.
4. Organizing and processing data (including graphs).
5. Drawing conclusions and critical discussion.

Under each of these five headings, there is a more detailed list, which contains the component-skills belonging to each heading. The detailed list is given in Appendix II.

Components 1, 2, and 3 refer to practical skills, i.e., those which have to do with manipulation and direct performance during the experiment. Components 4 and 5 refer to cognitive skills which connect the laboratory performance to theoretical thinking and drawing of conclusions, and depend on the ability to process information, digest it, and utilize it in the learning process.

The laboratory experience should combine all these elements in order to be a worthwhile activity in the science teaching process, and it is therefore imperative that all components be assessed.

Procedure

The teachers who participated in the project met twice. The first meeting consisted of two parts. First, the teachers watched two films (films A and B). Before viewing the films, the participants received the instructions which were given to the student whose performance they were about to watch. Teachers were also shown the equipment and apparatus used by the student in the experiment. After watching each film, each participant received a copy of the original lab report submitted by the student, in his handwriting. Each teacher was asked to assess the student's overall performance (the performance of the actual experiment and the lab report) in each film as he would usually do in his everyday teaching. In what follows we shall refer to these scores as the scores in the "subjective" (S) assessment.

In the second part of the first meeting, the teachers received the checklist we had developed. In order to train them in the use of this instrument, the teachers were then shown a third film (film C), and given the student's lab report. They were asked to assess the student's performance, this time using the list. Each of the teachers was then asked to start using the instrument in his regular laboratory lessons.

During the four months following the first meeting, teachers used the list to assess their students' performance in the physics laboratory in their schools.

At the end of this period, a second meeting took place, and 21 of the 25 teachers who started the project took part. In this meeting, the teachers were shown the same two films (A and B) which they had seen in the first meeting, and the same preliminary procedure was followed (i.e., they were given the instructions for the student and could check the equipment he had used). After watching each film, they again received the lab report, and were asked to assess the student's performance. This time, however, they were requested to use the instrument, to which they were by now accustomed. Each teacher was asked to assign five grades to the student, one grade for each of the five components of the list (Appendix II). The teachers were asked to state what weights would be adequate in their opinion for each component (1-5) in the list. The average weight (over all teachers) for each of the five components was computed. The results are given in Table I. These were the weights used in computing the students' scores. The weighted average of the five component scores constituted the overall "objective" (O) score for each student.

Thus, the data collected and used in our analysis can be summarized as follows:

a) "Subjective" scores assigned by each of 21 teacher participants to students in films A and B, collected during the first teacher meeting. Each student received a single "subjective" (S) score from each teacher.
b) "Objective" scores assigned by each of 21 teacher participants to students in films A and B, collected during the second teacher meeting. Each student receive five "objective" scores, one for each component of the instrument, from each teacher. A weighted overall average was computed from the five "objective" scores.

At the end of the project, the teachers who participated were asked to express their opinions about the use of the proposed assessment scheme in their physics teaching.

Results and Discussion

In Table II, the means and standard deviations over all 21 teachers are presented. These data will be discussed with reference to the objectives of our study.

Increase in Objectivity

As already stated, we aimed for an improvement in the consensus among teachers, when assessing the same student. From the data in Table II it is seen that even though the change in the mean grade (from S to O) is moderate, there is a substantial decrease in the standard deviations for both student A and student B. "Morgan" tests (Morgan, 1939) showed this decrease to be significant in both cases. If we regard the mean as our best approximation for the "correct" grade, then more teachers came closer to grading the student correctly in the objective mode of assessment. We believe this improvement in objectivity to be due to the use of well defined criteria by all the teachers, rather than grading the student subjectively according to their own preferences.

TABLE II

Means and Standard Deviations of Teachers' Grades (N = 21)

	Student A		Student B	
	Mean	Std. Dev.	Mean	Std. Dev.
S	7.60	1.23 *	7.05	1.04 **
0_1	8.79	0.90	7.60	1.24
0_2	8.40	0.89	6.95	0.92
0_3	8.62	0.91	7.31	1.05
0_4	6.90	1.30	6.38	1.07
0_5	7.33	1.29	6.14	1.01
0	7.84	0.81 *	6.74	0.64 **

* Morgan test (Morgan, 1939): $t = 2.15$ $p = 0.02$

** Morgan test: $t = 3.04$ $p = 0.003$

Increase in Precision

A second aspect of student assessment concerns precision. If a teacher considers in his assessment some aspects of student performance and disregards others, he will obviously score highly a student who performs well in the aspects he considers. Another student, who may perform well in some aspects this teacher does not consider, will receive a low score, if his performance is weaker on the skills the teacher does consider. This situation, or similar ones with different combinations of student skill versus teacher preferred aspects of performance clearly occur in classroom situations, in which teachers have to evaluate different students on different occasions.

By using a well defined list of criteria, every teacher is forced to consider all the aspects of student performance which are included in the list. Clearly, some teachers are more lenient than others, but this leniency (or lack of it) should manifest itself in all the aspects which they assess. Hence, precision should manifest itself in that the scores teachers give to different students will be highly correlated. We computed the correlation (Pearson r) between the sets of scores obtained by student A and by student B, both for the subjective (S) and the objective (O) modes of assessment. For the S scores we obtained $r = 0.36$ (p = 0.11) while for the O scores we find $r = 0.70$ ($p = 0.0004$). These results are also clear from Figures 1 and 2, where we plot each score received by student A versus the score received by student B from the same teacher.

The results confirm the hypothesis that by using assessment schedules which were based on ready made criteria, teachers assessed students much more precisely: they assigned grades to each of the abilities which should be considered, and consequently became much more consistent in their overall assessment of student achievement in the physics laboratory.

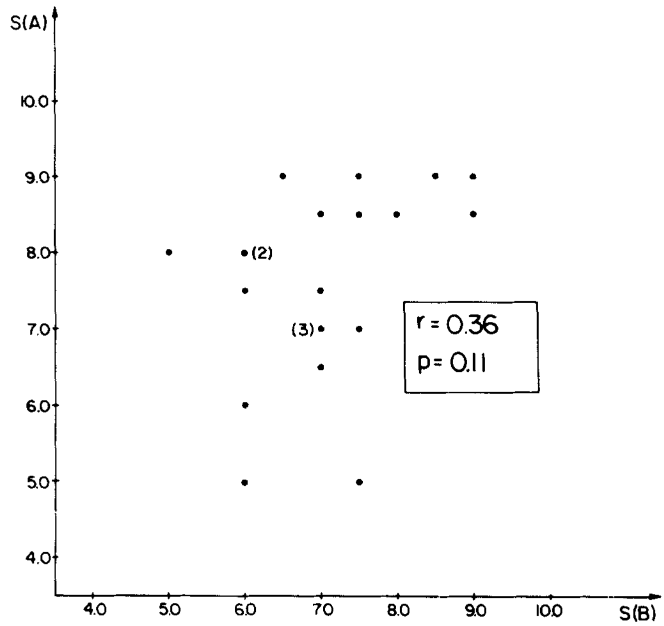

Figure 1. Grades received by student A versus grades received by student B—subjective (S) assessment.

Teachers' Attitudes

When trying to introduce a new assessment scheme into schools, teachers' attitudes and their willingness to adopt it should not be overlooked. In order to evaluate this aspect, a 16 item Likert type questionnaire was administered to the 25 teachers who started this project. In this questionnaire, the teachers were asked to express their opinions about the usability of the list, whether they thought it helped them to arrive at a better assessment of their students, and whether they would continue using it in the future. It was found that teachers agreed that the use of the list greatly improved the quality of their assessment. However, the four teachers who did not take part in the second part of the experiment (the objective assessment) did not use it in their classrooms. These teachers agreed that the list should be used in the final practical examinations, to ensure precision and objectivity, but they found it difficult to use it continuously in their regular teaching in school.

We also interviewed all the teachers at the end of the project, and they expressed their opinions about the assessment scheme. Some of the remarks made by teachers during these interviews are quoted below:

". . . presently, the criteria (for assessment) are totally vague, and such a list is an absolute must . . ."

". . . one single experiment and the impression of it (in the matriculation examination) cannot come instead of continuous observations by the teacher during the year"

". . . if in the first meeting (i.e., where the assessment was done subjectively, without the list) there was such a wide distribution of grades, based on all kinds of arguments, then the use of such a list should be made compulsory. . ."

". . . I had some difficulties using the list, so I posted it at the entrance to the lab. My students could

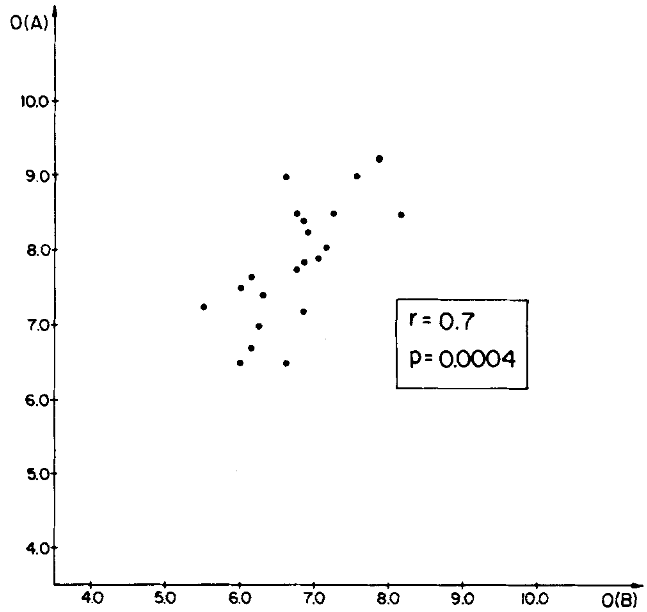

Figure 2. Grades received by student A versus grades received by student B—objective (O) assessment.

read it and check which of the skills (in the list) is necessary in the experiment they were performing. I think this was even more important than the grades I gave them at the end."

Conclusions

For many years, science educators have recognized the importance of the laboratory in science teaching and learning. If one agrees that the laboratory is a unique mode of instruction, then the assessment of this mode should not be overlooked both by teachers and by those involved in teacher training. Assessment of laboratory skills, which is based exclusively on written evidence, will necessarily neglect those skills which manifest themselves during the actual performance of the laboratory exercise. Even practical examinations have severe drawbacks, if they are used only on special occasions, such as in final examinations. Furthermore, practical examinations in which the assessment is not based on clearly defined criteria are not very useful, since their outcome will be greatly influenced by personal preferences and biases of the particular examiner.

We have tried out an assessment scheme which is based on a clearly defined list of criteria. It was found that by using this procedure, the assessment of laboratory performance was greatly improved, both in precision and in objectivity. Furthermore, this scheme is useful for continuous assessment by the teacher in his classroom during the school year, and therefore offers a remedy to most of the disadvantages discussed above.

Finally, we feel that those involved in teacher training should realize that there is a need to incorporate such methods of evaluation of student performance in the laboratory into their training programs, so that in the future it will be used by more teachers as part of their physics instruction.

Appendix I

Description of the Experiments Which Were Filmed

Experiment A: Collisions in Two Dimensions

The apparatus includes a "launching run-way" on which a steel ball can slide and collide with another ball (steel or glass) which is held at the end of the slope. The experiment can be arranged so that the collision angle is different each time. By marking the positions at which the two balls hit the floor after a collision, the student can construct a vector parallelogram for each collision, and investigate momentum conservation and also energy conservation in elastic collisions.

Experiment B: The Magnetic Field of a Solenoid

The apparatus includes an air-core solenoid (500 turns, in six layers, 15 cm long, 4 cm internal diameter) and a "current balance." The current balance is used to measure the force which the magnetic field at the center of the solenoid exerts on a section of copper wire, which carries electric current. The student investigates the dependence of this force on the current through the wire and on the current through the solenoid (which creates the magnetic field).

He is also able to calculate μ_0 (the permeability of the vacuum) from the results of his measurements.

Experiment C: Measuring the Energy Stored in a Capacitor Using a Solid Minicalorimeter

A capacitor is repeatedly charged and discharged through a resistor embedded in a small insulated aluminium cube The student measures the temperature rise of the cube, and calculates the heat energy needed to cause such a rise. He investigates the dependence of the energy on the charging voltage (V), and finds both the proportionality to V^2 and the capacitance.

Appendix II

Assessment List

1. *Constructing the experimental setup and other manipulative skills* (15%)*
   —identifies the components of an experimental setup from a schematic diagram.
   —constructs the experimental setup according to a scheme or to instructions.
   —uses the correct equipment for each measurement or observation.
   —overcomes simple malfunction of equipment.
2. *Observing and measuring* (25%)*
   —reads the instruments correctly (scale, parallax, etc.).
   —performs observations and measurements correctly.
   —records the results (of observations and measurements) clearly.
3. *Ordering and organizing work* (10%)*
   —arranges the equipment on the bench in an orderly manner.

—keeps tidy and orderly records in the notebook.
—observes safety regulations.
—completes the prescribed task in the time allotted.
—concentrates on his work and does not disturb his fellow students.

4. *Organizing and processing data (including graphs)* (25%)*
—processes the results of observations correctly.
—organizes tables of measured data efficiently and clearly.
—records results of measurements with correct units and significant digits.
—construct graphs correctly (scales, mark data points in coordinate system, best line through data points, . . .).
—obtains numerical values from graphs.
—chooses correct formulae for calculating unknown variables.
—substitutes data correctly in formulae.
—performs computations correctly.

5. *Drawing conclusions and critical discussion* (25%)*
—understands the objectives of an experiment, and the connection between them and the experimental procedure of its execution.
—decides how many measurements are necessary in order to find relations between variables.
—identifies the shape of a graph and uses it to find a functional dependence between variables.
—discovers and identifies reasons for illogical results.
—identifies reasons for improper performance of an experimental system.
—understands the limitations of an experiment, and tries to find ways for improvement.
—formulates new assumptions in view of experimental results.

* Percentage of the component in the total score.

The authors wish to thank Batia Levi, Bat-Sheva Eylon, and Yetty Salomon for their help in conducting the experiment and analyzing the data. The authors also thank the referee for some helpful remarks concerning both contents and style.

References

Ben-Zvi, R., Hofstein, A., Samuel, D., & Kempa, R. F. Modes of instruction in high school chemistry. *J. Res. Sci. Teach.*, 1977, *14*, 433–439.

Buckley, J. G. An investigation into the assessment of practical abilities in 6th form chemistry courses, unpublished M.Sc. dissertation. The University of East Anglia, 1970.

Doran, R. L. Assessing the outcomes of science laboratory activities. *Science Education*, 1978, *62*, 401–409.

Eglen, J. R., & Kempa, R. F. Assessing manipulative skills in practical chemistry, *School Sci. Rev.*, 1974, *56*, 261–273.

Giddings, G., & Hofstein, A. Trends in the assessment of laboratory performance in high school science instruction. *Aust. Sci. Teach. J.*, 1980, *26*, 57–64.

Grobman, H. *Developmental curriculum projects: decision points and processes* Itaca, Illinois, Peacock Publishers, Inc., 1970.

J. M. B. (Joint Matriculation Board) The internal assessment of practical skills in chemistry, Universities of Manchester, Leeds, Liverpool, Sheffield, and Birmingham, January, 1979.

Kelly, P. J., & Lister, R. E. Assessing practical abilities in Nuffield A level biology in J. F. E. Eggelston and J. F. Kerr (Eds.). *Studies in Assessment*, London, English Universities Press, 1969.

Kruglak, H. Evaluating laboratory instruction by the use of objectives type tests. *Amer. J. Phys.*, *1958*, *26*, 31–32.

Lunetta, V. N., Hofstein, A., & Giddings, G. Evaluating science laboratory. *Sci. Teach.*, 1981, *48*, 22–25.

Mathews, J. C. The Nuffield A-level chemistry examinations, *Educ. Chem.*, 1969, *6*, 205–207.

Morgan, W. A. A test for the significance of the differences between the two variances in a sample from a normal bivariate population. *Biometrika*, *1939*, *31*, 13–19.

"*Physics Experiments for High School*", Rehovot (Israel), The Weizmann Institute of Science, Department of Science Teaching, 1975 (in Hebrew).

Robinson, J. Evaluating laboratory work in high school biology. *Amer. Biol. Teach.*, 1969, *31*, 236–240.

Shulman, L. S., & Tamir, P. Research on teaching in the natural sciences in R. M. W. Travers (Ed.). *Second Handbook of Research on Teaching.* Chicago: Rand McNally, 1973.

Tamir, P. The practical mode—A distinct mode of performance in biology. *J. Biol. Educ.*, 1972, *6*, 175–182.

Tamir, P. An inquiry-oriented laboratory examination. *J. Educ. Measure.*, 1974, *11*, 25–33.

Tamir, P., & Glassman, F. A practical test for BSCS students: a progress report. *J. Res. Sci. Teach.*, 1971, *8*, 307–315.

University of London, Notes and guidance and report form for the optional internal assessment of special studies, University of London, 1977.

Accepted for publication 10 November 1981