

## Development of Scoring Rubrics to Assess Physics Practical Skills

Sang Sang Liew<sup>1\*</sup>, Hooi Lian Lim<sup>1</sup>, Salmiza Saleh<sup>1</sup>, Saw Lan Ong<sup>1</sup>

<sup>1</sup> Universiti Sains Malaysia, Penang, MALAYSIA

Received 5 August 2018 • Revised 16 November 2018 • Accepted 7 December 2018

### ABSTRACT

The Malaysian curriculum for physics aims to produce active learners who are good with physics concepts, practical skills and scientific attitudes. Practical skills are acquired through active involvement in investigative work in the laboratory. This paper discusses the development of a scoring rubric that can be used to assess students' physics practical skills through observation, where scores are assigned to students while they are carrying out practical work. The important physics practical skills are identified through literature review and criteria of scoring each skill are depicted. A total of twenty skills are outlined in the scoring rubric. These skills were subjected to review by a panel of expert teachers. Evidences for validity of test content were gathered. The instrument was tested on 153 Form 4 students and the data obtained was subjected to PLS-SEM analysis. Interpretation of results indicates satisfactory validity and reliability. Results on the scoring rubrics is found to have weak but significant correlation with a written practical test subjected to the same group of students.

**Keywords:** rubrics, physics practical skills, validity, reliability

### INTRODUCTION

Physics is known to many as a difficult subject as it involves understanding of abstract concepts and complex formulae. Practical work plays an important role in the learning of physics. Practical work in the laboratory enables phenomena to be reproduced and thus enhance students' understanding of abstract concepts (Deacon & Hajek, 2011). Practical work also enables students to check on theories and principles and help to correct misconceptions and confusion between what is learnt in schools and what is experienced in daily life, and thus create links between theory and practice (Prades & Espinar, 2010). From the practical work performed in the school laboratory, students can determine the relationship between different variables and physical quantities, from which formulae and laws can be deduced. However, this can only be achieved if the data and the results of the practical work is obtained using reliable and valid methods.

Apart from enhancing the learning of physics concepts and laws, practical work in the school laboratory is also the platform for students to acquire skills of manipulating laboratory equipment and apparatus (Abrahams, 2011). By conducting practical work regularly, students develop their practical skills and learn to collect more reliable and valid data. As such, students' practical skills have to be assessed continuously to determine if they have mastered the required skills to perform practical work and handle data or information. Thus, practical skills assessment has always been a component in the assessment of physics as well as other science related subjects. Science practical assessment can be carried out through centralised practical examination, school based practical assessment or written test. In Malaysia, students' ability to conduct practical work in physics is assessed using written practical test and school based science practical work assessment (PEKA). Scores for the written practical test is included in the centralised examination to obtain the Malaysian Certificate of Education (SPM) while performance on PEKA is reported in a separate certificate (Ong, 2010).

In order to know the ability of students to conduct practical work, students should be assessed while they are performing practical work in the laboratory. According to Reiss, Abrahams and Sharpe (2012), countries with high performing students in science education, such as Singapore and Hong Kong, assessed students' science practical

### Contribution of this paper to the literature

- Quantify student practical skills in the physics laboratory into four domains.
- Provides an alternative to assessment of practical skills in the physics laboratory by using a direct observation approach where student ability of conducting practical work is assessed while practical work is being carried out and quantified through the use of a scoring rubric.
- A comparison between results from direct observation and written test on the same domains of practical skills shows that the two modes of assessment are not comparable and thus should both be applied to assess the actual ability of a student.

work by observing them during the performance of scientific investigation. Yet, assessment through systematic observation is a method that is seldom practised in Malaysian schools. This aspect is also lacking in the assessment of physics practical skills. Scores from observation only contribute a minor part in PEKA (3 out of 26 performance indicators) (MOE, 2009) and is often neglected. There are no clear guidelines on what to observe and how to observe in order to determine students' ability to conduct practical work in physics. Thus, it is necessary to introduce this mode of assessment in Malaysian secondary school physics laboratory.

This paper discusses the development of a scoring rubric that can be used to assess students' practical skills through the method of direct observation in the Malaysian secondary school physics laboratory. The physics practical skills that secondary school students need to master are first identified as items in the scoring rubric. Then, the main objectives of this study are to develop a valid and reliable scoring rubric which is easily applied by all physics teachers to assess these physics practical skills.

The main research questions are as follows:

1. What are the important physics practical skills at secondary school level?
2. Is the developed scoring rubric reliable?
3. Is the developed scoring rubric valid?
4. Is the score obtained from direct observation comparable to score from written test?

This study suggests a more comprehensive method of assessing physics practical work at secondary school level. Currently, there is no assessment framework which is specific to physics practical skills. Scientific skills outlined in the curriculum specifications for physics (MOE, 2005) are science process skills and manipulative skills. These skills are generic for all science and science related subjects, and are listed in all curriculum specifications for these subjects be it primary school or secondary school. In the physics laboratory, techniques of measuring physical quantities are important to obtain reliable data to deduce relationship between variables. However, there is no guidelines as to how these skills should be assessed. As such, this paper reports the development of a scoring rubric which is believed to be able to assess physics practical skills more comprehensively compared to the existing ones especially in the Malaysian context. The scoring rubric in this study is used to score students during observation, which is not being emphasized currently.

## LITERATURE REVIEW

This paper suggests students' practical skills to be assessed using performance assessment, with a direct observation approach. Performance assessment is the form of assessment which allows the assessor to evaluate both the process and product from conducting a task (Lane & Stone, 2006). Hence, it is the suitable method to be used in assessing practical skills as students are required to conduct practical tasks assigned by the teachers. Performance of practical tasks allow students to show to an observer both the process that is carried out and the product that is created (Musial, Nieminen, Thomas, & Burke, 2009). For the case of practical work in the physics laboratory, the product refers to experimental reports or models produced by the students. In the case of this study, the process is emphasized more than the product. Scoring rubric is used to rate and score the students' skills while they are performing the practical tasks.

Solano-Flores and Shavelson (1997) suggested the development of performance assessment to be cyclical, where the tasks and scoring rubrics should be tried out, pilot tested and revised continuously before it is actually used. Doran, Chan, Tamir and Lenhardt (2002) suggested an assessment development model to develop laboratory assessment. This model suggests that the purpose of the assessment must be clarified before selecting the task format. Scoring rubrics is then developed based on the requirement and administrative procedures of the task. Concluding the suggestion of Solano-Flores and Shavelson (1997) and Doran et al. (2002), development of practical work assessment is an iterative process and should be constantly revised, as had been practised in the development of scoring rubrics in this study, so that the interpretation from the assessment is valid.

As practical assessment had been an essential in assessment of science education, there are existing framework of science practical skills assessment. The assessment framework for PEKA for the physics subject (MOE, 2009) has five constructs, which are designing an experiment, conducting an experiment, tabulating data, analysing data and results, and scientific attitude. There are a total of 26 performance indicators in this assessment framework, seven in the first construct, three in the second construct, four in the third construct, seven in the fourth construct and four in the fifth construct. The constructs of designing experiment and analysing data have the most number of performance indicators, showing that these two constructs are being given more emphasis. All the indicators should be shown in the students' written report for an experiment, except for the three performance indicators for conducting experiment, which should be assessed while students perform the experiment. The assessment framework provided by the Malaysian Examination Syndicate (MES) illustrates the important practical skills that students have to master in a secondary school physics laboratory. However, the execution of the practical work is not being emphasized, considering the fraction of scores given to the actual performance of practical work, which is assessed by teacher's judgement. Details on how the skills should be assessed, especially those that are specific for physics practical work, are not being stated.

Akinbobola and Afolabi (2010) from Nigeria claimed that physics practical skills are science process skills, as those are the skills that are included the senior school certificate physics examination in Nigeria. They identified five prominent science process skills as manipulating, calculating, recording, observing, and communicating, which are skills frequently assessed in examinations. Other skills include measuring, comparing and contrasting, drawing, experimenting, graphing, deducing, investigating, formulating models and interpreting. However, details of essential skills in experimenting and investigating were not discussed. On the other hand, they argued that more integrated (higher order) science process skills should be included to prone creativity, problem solving, and reflective thinking among students. Students should be given chances to test their ideas experimentally through manipulation of materials in the laboratories. As such, it can be deduced from Akinbobola and Afolabi (2010) that skills of manipulation of materials for the purpose of experimenting are important for a physics student, and hence had been included in the scoring rubrics developed for this paper.

On the other hand, Edy and Lilia (2010) had developed a test to assess integrated process skills. They identified integrated science processes of formulating hypotheses, operationally defining variable, identifying and controlling variable and interpreting data as skills that are associated with planning investigation. The assessment framework used by Edy and Lilia (2010) is similar to that used by PEKA (MOE, 2009). For Edy and Lilia (2010), assessment of skills is done through a paper-and-pencil test which consists of 30 multiple choice items. The test was specifically developed for primary school students but has provide information in the practical skills required for designing experiment. There are similarities between physics practical skills stated by Akinbobola and Afolabi (2010) and integrated process skills identified by Edy and Lilia (2010), which were adopted into the practical skills assessed in the scoring rubrics developed in this paper.

Practical skills in the physics laboratory are also closely related with scientific inquiry skills, which include identifying problem, plan investigation, analyse data, and finally make conclusion and justification (NRC, 1996). Wenning (2007) proposed basic science inquiry skills in a secondary school physics laboratory to be: identify a problem to be investigated; using induction, formulate a hypothesis incorporating logic and evidence; using deduction, generate prediction from the hypothesis; design experimental procedures to test the prediction; conduct a scientific experiment to test the hypothesis; collect meaningful data, organise and analyse data accurately and precisely; apply numerical or statistical method to reach conclusion; explain unexpected results; and defend results of investigation. This paper proposes a scoring rubric that can be used to observe the planning and designing of experiment procedures in process so as to analyse students' thinking process in exhibiting science inquiry skills, as well as their ability to carry out their planning successfully.

Science process skills, whether the ones outlined in the curriculum specifications for physics (MOE, 2005) or those discussed by Akinbobola and Afolabi (2010), and the assessment framework of PEKA had not mention the skill of 'making justification', which is also an important aspect in practical work. Making justification to explain unexpected results in order to defend results of investigation is part of the inquiry cycle. This is comparable to the 'concepts of evidence' proposed by Gott and Duggan (1995). Khaparde and Pradhan (2002) believed that exercises of practical skills in the physics laboratory include the aspect of procedural understanding, which is the understanding of 'concepts of evidence' associated with design, measurement, data handling and evaluation of a complete investigative task. Procedural understanding is the understanding which enable one to use systematic experimental skills to verify theory and generate knowledge (Khaparde and Pradhan, 2002) this it is the thinking behind the doing which influence the decision that one have to make while conducting practical work.

Understanding of issue of reliability of empirical evidence that involves procedures of how reliable evidence can be collected are important features in procedural understanding (Miller, Lubben, Gott, & Duggan, 1994), whereas validity of science depends on objective experimental evidence. Gott and Duggan (1995) introduce

**Table 1.** Concepts of Evidence Associated with Different Stages of Scientific Investigation

Stages of scientific investigation	Concepts of evidence
Design	Variable identification
	Fair test
	Sample size
	Variable types
Measurement	Relative scale
	Choice of instrument
	Repeatability
Data handling	Accuracy
	Tables
	Graph types
	Patterns
Evaluation	Multivariate data
	Reliability
	Validity

Notes. Adapted from *Investigative work in the science curriculum: Developing science and technology education* (p.31) by R. Gott and S. Duggan, 1995, Open University Press.

concepts of evidence to assist and support procedural understanding. Examples of concepts of evidence associated with different stages of scientific investigation is shown in [Table 1](#).

Although the concepts of reliability and validity are associated with the stage of evaluation which is the final stage in scientific investigation, upon the completion of the investigative task, they are actually incorporated into each stage of the investigation. For example, although sample size is associated with the stage of design, one has to consider the size of the sample that is most likely to represent the population, that is the number of measurements that had to be taken, in order for the data collected to be valid when reported. As such, procedural understanding is an important aspect that should be included in the assessment of practical skills. Skills outlined in the scoring rubrics in this paper incorporate procedural understanding into science process skills and inquiry skills discussed in the previous paragraphs so that students know the importance of the thinking behind the doing.

The actual process of performing the practical work involved manipulative skills, which are the skills related to experimenting, observing and measuring. The Malaysian curriculum specifications for secondary school physics had listed manipulative skills as psychomotor skills which enable students to use, handle, draw, clean and store apparatus correctly and safely, which indicates that safety is an important feature while performing practical work. Hidayah and Rohaida (2013) identified manipulative skills as skills related to following instructions, checking apparatus for functionality, efficiency in using time, tidiness of working area, practice of cleaning and storing of apparatus, cautious handling of apparatus, correct and proper technique of handling apparatus, making assumptions, error in measuring and skills in drawing specimens. Moni et al. (2007) included graphing skill as manipulative skill. Although graphing skill is essential while carrying out practical work especially in the physics laboratory, this skill can be taught before students enter the laboratory. The hypothetical-deductive model by Lawson, Abraham, and Renner (1989) which focuses learning process from the formation of hypotheses from observations and deduction of ideas from results of investigations is used in this paper to uncover the practical skills required by students in a secondary school physics laboratory. As practical work in the laboratory helps to enhance student learning, the processes of practical work follow a certain sequence. In order to solve a problem through scientific investigation, the four processes as designated by the hypothetical-deductive model, from making induction then deduction, to comparing and making inference. Hence, four domains of practical skills are extracted from the learning processes of the hypothetical-deductive model as the important constructs for physics practical skills, namely Design, Execution, Analysis and Evaluation. These four domains are also consistent with the stages of scientific investigation indicated by Gott and Duggan (1995) when illustrating the concepts of evidence for procedural understanding.

From the review of existing framework for practical skills required for carrying out practical work in the school laboratory, a total of 20 practical skills had been outlined in the scoring rubrics in this paper. These practical skills were generated from existing framework of science process skills and inquiry skills through a series of review and SMEs' judgement. The skills outlined were checked to be analogous to the skills suggested in the Malaysian curricular specifications for secondary school physics. Skills that are focus on the process of experimenting and data gathering are given more emphasis as they are believed to be important skills especially in the secondary school physics laboratory and can be assessed through observation during the execution of practical work. Procedural understanding is integrated in the definition of the skills and underlines the criterion in the scoring rubrics. Details of development of the rubrics is discussed in the next session. Compared to existing rubrics for practical skills, the rubrics discussed in this paper provides details on the criteria for different mastery levels of

physics practical skills at secondary school level, which enables teachers in schools to help their students to improve in the execution of practical work.

## METHODS

This study uses an exploratory design to look into the possibility of assessing physics practical skills on secondary school students using direct observation approach. For this purpose, the scoring rubric to assess students while they perform practical work is developed. Physics practical skills required for secondary school students were identified from the pool of process skills, inquiry skills, manipulative skills, and procedural understanding. These skills are grouped into the domains of Design, Execution, Analysis and Evaluation which are the domains of practical skills that had been identified from literature review, with hypothetical-deductive model as the foundation. The practical skills are clustered into the respective domains so that they are consistent with the performance indicators in the assessment of physics practical work in the Malaysian secondary school context.

The skills necessary for designing an experiment in a secondary school physics laboratory include identifying variables, choosing suitable apparatus, designing experimental set-up, determining suitable range and intervals for the variables. These skills are grouped under the domain of 'Design'. Skills that are grouped under the domain of 'Execution' are the ones that are important for collecting valid and reliable data, which include setting-up of apparatus, checking and using instruments with correct techniques, manipulating variables, recording of data and taking suitable precautions. Skills that are identified to be under the domain of 'Analysis' involve data handling which include performing correct calculation, analysing data to obtain relationship and making suitable deduction. The domain of 'Evaluation' encompasses skills to evaluate results and make judgement, which includes making conclusion, judging the accuracy of results, identifying sources of errors and weaknesses and suggesting methods to improve results. The skills were judged and approved by a panel of expert teachers in physics education and physics assessment.

The first draft of the scoring rubric consists of five skills in the domain of 'Design', ten skills in the domain of 'Execution', three skills in the domain of 'Analysis' and five items in the domain of 'Evaluation'. Rating criteria are designed for each of the skill. Each skill is scored at three performance levels with score of either '1', '2' or '3'. Criteria for each performance level is outlined in the scoring rubrics. The scoring rubrics went through three cycles of expert (SMEs) reviews by expert teachers who have more than ten years of experience in teaching physics and assessing physics practical work. In the first cycle, the appropriateness of the language used in describing the criteria was checked and improved. In the second cycle, the clarity of criteria was reviewed. The panel also judge whether the criteria stated are observable and assessable.

The scoring rubric is modified and improved after each cycle of expert review. Number of skills in the domain of 'Design' and 'Execution' are reduced to four and eight skills respectively, based on the suggestions given by the expert teachers. In the final cycle of review, a panel of seven experts were invited to rate the scoring rubric. These seven experts were expert teachers appointed by the Ministry of Education. They came from different parts of Malaysia, where three are from the Northern states, three are from East Malaysia and one is from the middle region of Malaysia. Two of the panellists have master degree in education. The panel rated the importance of the skills on a scale of three where: 1 = not important; 2 = important; and 3 = very important. Content validity index of the scoring rubric is computed from the experts' rating.

The scoring rubrics were then tested on 153 Form Four students from 10 secondary schools in the Kuching district of Sarawak state. Students were required to perform three practical tasks designed by the researcher. Students were observed and scored while they were performing these practical task in the school physics laboratory. Analysis of data and evaluation of experimental result by the students were collected as evidences. Performance of each student was rated individually by two raters. The raters attended a one-day workshop on how to use the scoring rubrics before they participate in the data collection procedures. The data collected were subjected to analysis using PLS-SEM to confirm the domains of physics practical skills that were determined from the hypothetical-deductive model.

At the same time, students also responded to a written practical test. The written test was designed to be consistent with the domains of practical skills as that in the scoring rubrics. The written test is a comparative measure for the purpose of validation for the scoring rubrics, as Malaysia uses written practical test to assess students' physics practical skills. However, due to the nature of the test, not all the skills outlined in the scoring rubrics are able to be tested in the written test, especially that in the domain of 'Execution'. The written test used had also been subject to experts review before it is administered to the students.

**Table 2.** Content Validity Ratio for Items in Scoring Rubrics

Domain	Items	CVR
Design	A1	1.00
	A2	0.78
	A3	0.56
	A4	0.56
Execution	B1	0.56
	B2	0.56
	B3	0.78
	B4	0.78
	B5	1.00
	B6	0.78
	B7	0.56
	B8	0.56
Analysis	C1	1.00
	C2	1.00
	C3	0.78
Evaluation	D1	1.00
	D2	0.78
	D3	0.78
	D4	0.56
	D5	0.78
	<b>CVI</b>	<b>0.76</b>

## RESULTS

### Validity Evidence based on Test Content

A quantitative approach was used as a mean to gather evidence for validity of the test content, which is through the computation of content validity ratio (CVR) and content validity index (CVR). The following formula is used to determine the value of CVR:

$$\text{Content validity ratio, CVR} = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

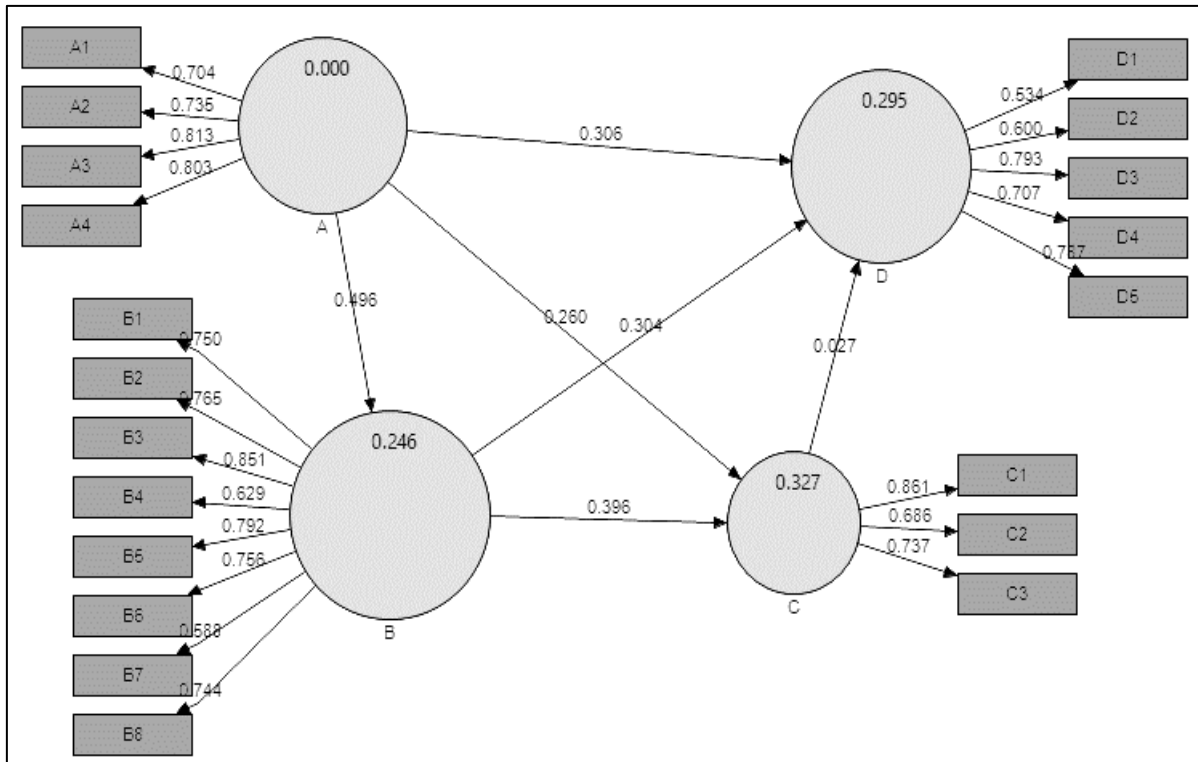
where  $n_e$  is the number of experts who rated the item as essential and  $N$  is the total number of experts in the panel. The value of CVI is calculated from the mean value of CVRs of all the items (Shultz, Whitney, & Zickar, 2014). If the panel consists nine panellists, the minimum value acceptable for CVR is 0.78 for content validity. **Table 2** shows the CVRs for all items and CVI for the scoring rubric. The value 0.76 is slightly lower than the threshold value suggested by Lawshe in Shultz et al. (2014).

### Interrater Agreement

Scores were assigned to the students by trained raters. Overall percentage of interrater agreement is 86.6%. Kappa values computed shows substantial strength of agreement with the Cohen kappa's correlation as 0.71. This indicates that the rubric developed in this study can be used to assess students' physics practical skills.

### Construct Validity

PLS-SEM analysis was conducted on the scores obtained by the students to check for construct validity through the determination of convergent validity and discriminant validity of the scores. Convergent validity is the extent to which an item in a construct is positively correlated with other items in the same construct, while discriminant validity is the extent to which a construct is distinct from other constructs. Path diagram used for analysis in PLS-SEM is shown in **Figure 1**.



**Figure 1.** Path diagram for measurement model of practical skills in physics laboratory. The alphabets 'A', 'B', 'C', and 'D' represent the domains of 'Design', 'Execution', 'Analysis' and 'Evaluation' respectively.

Path coefficients shown on arrows between the circles in **Figure 1** shows that the domains identified are interrelated although the effect of one domain on the other is small. The main loading for each item is shown on the arrows from the circles of A, B, C and D which represent the domains of Design, Execution, Analysis and Evaluation respectively. Item can be accepted as an indicator of the domain if the main loading of the item is more than 0.70, meaning that it could explain at least 50% of the variance on the domain (Hair et al., 2014). There are five items which do not fulfil this criterion, namely items B4 and B7 from the domain of 'Execution', item C2 from the domain of 'Analysis' and items D1 and D2 from the domain of 'Evaluation'.

**Table 3** shows the main loading of each item and the AVE (average variance extracted) value is the grand mean value of the squared loadings of the items associated with a construct. AVE value greater than 0.50 is considered acceptable as it explains more than half of the variance of the items (Hair et al., 2014). **Table 3** shows that only the domain of 'Analysis' has AVE lower than 0.50. Composite reliability to assess internal consistency reliability is also shown in **Table 3**. The values range from 0.81 to 0.91 which are considered as satisfactory. However, high value of 0.91 in the domain of 'Evaluation' is not desirable as value greater than 0.90 implies existence of redundant items.

**Table 3.** Summary from PLS-SEM for Domains of Practical Skills (Construct Validity)

Domains	ITEMS	Main loadings	AVE	Composite Reliability
Design	A1	0.71	0.59	0.81
	A2	0.74		
	A3	0.81		
	A4	0.80		
Execution	B1	0.75	0.59	0.85
	B2	0.77		
	B3	0.85		
	B4	0.63		
	B5	0.79		
	B6	0.76		
	B7	0.59		
	B8	0.75		
Analysis	C1	0.86	0.48	0.82
	C2	0.69		
	C3	0.74		
Evaluation	D1	0.53	0.55	0.91
	D2	0.60		
	D3	0.79		
	D4	0.71		
	D5	0.79		

**Table 4.** Inter-Correlation Matrix for Domains (Discriminant Validity)

	Design	Execution	Analysis	Evaluation
Design	<b>0.765</b>			
Execution	0.496	<b>0.765</b>		
Analysis	0.457	0.525	<b>0.692</b>	
Evaluation	0.469	0.470	0.327	<b>0.739</b>

**Table 5.** Paired Sample T-test on Performance in each Domain using Direct Observation and Written Practical Test

Domains	t-value	Degree of Freedom	Sig. (2-tailed)
Design	- 2.914	152	0.004
Execution	2.973	152	0.003
Analysis	7.626	152	0.000
Evaluation	8.828	152	0.000
Overall	- 6.712	152	0.000

For discriminant validity, the square root of the AVE values of each domain is calculated. The values are shown on the diagonal of **Table 4** in bold. According to the Fornell-Larcker criterion, the square root of AVE values for each domains should be greater than its highest correlation with any other construct and should exceed the squared correlation with any other construct (Hair et al., 2014). Thus, the values in bold should be the highest value in each column and each row in **Table 4**. The Fornell-Larcker criterion is satisfied as the value of the square root of AVE for each domain on the diagonal of **Table 3** is higher than the off-diagonal values.

Paired sampled t-test was conducted to determine whether there is significant difference between scores collected through direct observation and scores from written test. The same domains of practical skills were assessed in both modes of assessment although they are different in nature. Results shown in **Table 5** indicate that students performed differently in performance for each domains and in overall performance of practical skills. This showed that although clustered under the same domains, skills measured by written test and direct observation are different. Written test is more suitable for assessing cognitive knowledge while direct observation assesses performance of psychomotor skills apart from cognitive knowledge.

Pearson product-moment correlation coefficient was also computed to determine whether there is any relationship between scores collected from direct observation and scores obtained from written practical test. The correlation coefficient was found to be 0.338 which is significant at  $p < 0.01$ . This indicated that there is a weak but significant correlation between the two modes of assessment. Correlation matrix for the relationship between direct observation and written test for each domains of practical skills is shown in **Table 6**.



**Table 6.** Correlation Matrix of Relationship between Direct Observation and Written Test

	<b>Design</b>	<b>Execution</b>	<b>Analysis</b>	<b>Evaluation</b>
Design	.206*			
Execution	.121	.108		
Analysis	.261**	.214*	.201*	
Evaluation	.196*	.142	.137	.233**

\* $p < 0.05$  (2-tailed)\*\* $p < 0.01$  (2-tailed)

## DISCUSSION AND CONCLUSION

The process of developing instrument for assessment using a direct observation approach is iterative. This paper discusses the earlier process of developing the scoring rubric to assess students' physics practical skills when they are performing practical work in the secondary school laboratory. As recommended by DeVellis (2012), expert (SMEs) review is used to analyse the validity of the scoring rubric. Physics expert teachers assisted the development of the scoring rubrics by suggesting the suitable criteria for different performance level of a certain physics practical skill.

The rubric was rated by a panel of expert teachers. Evidence for validity of test content was obtained from the ratings of the SMEs which is then subjected to the calculation of content validity ratio (CVR) and content validity index (CVI). Seven items were found to have low CVR. Although the CVI obtained is within the acceptable range, where more than half of the panellists considered the skills listed in the scoring rubric as important, it also indicates that the scoring rubric can be further improved, especially in the items with lower CVR.

Finding also indicates that skills which are not included in the Malaysian assessment scheme for practical skills have lower CVR. As the content validity is rated by the Malaysia expert teachers, this findings implied that expert teachers are bounded with the current Malaysian assessment scheme. Thus, more experts have to be referred. Experts other than secondary school teachers, for example professors or lecturers in the higher education institution and professionals in industrial laboratories, can be included in the expert panel so that a more comprehensive view of important physics practical skills can be obtained.

Construct validity is evaluated from the PLS-SEM analysis. Results show that there are five items which have main loading lower than the suggested threshold value. These five items assess skills of testing experimental set-up, recording of data, analysing relationship, concluding findings and making judgement. Low main loadings indicate that students have low scores in these skills. These five skills are important skills that students should master when conducting practical work, as agreed through expert reviews. Students performed poorly in these skills causing the scores for these items to be low, which shows the need and importance of these skills to be reinforced. Thus, these five items need to be improved in order to improve the validity of the scoring rubric.

Interpretation of results from PLS-SEM show that the scoring rubric have satisfactory convergent validity and discriminant validity when tested on a sample of 153 Form 4 students. Internal consistency reliability was found to be satisfactory as well. This shows that the scoring rubric is valid and reliable to be used to assess students' physics practical skills in the secondary school physics laboratory. However, improvement is still needed for the scoring rubrics to be useful, especially if it is to be applied on large scale assessment.

This study shows that physics practical skills are multidimensional with at least four domain as confirmed from the analysis in the study. When assessed through written test, scores were found to be different as that collected from direct observation. This suggested that practical skills should be assessed with different methods, as certain domains are better assessed through observation while others are better tested through written test. As such, the scoring rubric developed can be used as a complimentary to assessment of practical skills which is currently based on written products of students only.

Consequently, the scoring rubrics discussed in this paper benefits the expectation of the test administrator, and verify the fact that practical skills are assessable through the method of observation. This rubrics was developed with the anticipation that it can be used in the Malaysian physics practical skills assessment framework. With the emphasis given on STEM education by the Malaysian government nowadays, there is a high potential for the scoring rubrics to be used in school based assessment as it provides teachers and educators with information necessary to help their students to improve skills in conducting practical work. It is more practical to use the scoring rubrics in formative assessment, so that students are notified of their weaknesses and know where to improve. However, further analysis and improvement is needed before the rubrics is presented to the relevant stake holders.

## REFERENCES

- Abrahams, I. (2011). *Practical Work in Secondary Science: A Minds-on Approach*. London: Continuum International Publishing Group.
- Akinbobola, A. O., & Afolabi, F. (2010). Analysis of science process skills in West African senior secondary school certificate physics practical examinations in Nigeria *Bulgarian Journal of Science & Education Policy*, 4(1), 32-47.
- Deacon, C., & Hajek, A. (2011). Student Perceptions of the Value of Physics Laboratories. *International Journal of Science Education*, 33(7), 943-977. <https://doi.org/10.1080/09500693.2010.481682>
- DeVellis, R. F. (2012). *Scale development: theory and applications* (3rd ed.). California: SAGE Publications, Inc.
- Doran, R., Chan, F., Tamir, P., & Lenhardt, C. (2002). *Science Educator's Guide to Laboratory Assessment*. Virginia: NSTA Press.
- Edy, H. M. S., & Lilia, H. (2010). Development and validation of a test of integrated science process skills. *Procedia - Social and Behavioral Sciences*, 9(0), 142-146. <https://doi.org/10.1016/j.sbspro.2010.12.127>
- Gott, R., & Duggan, S. (1995). *Investigative work in the science curriculum: Developing science and technology education*. Buckingham, England: Open University Press.
- Hair, J. J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least square structural equation modeling (PLS-SEM)*. USA: SAGE Publication.
- Hidayah, M. F., & Rohaida, M. S. (2013). Phenomenographic study of students' manipulative skills during transition from primary to secondary school. *Jurnal Teknologi*, 63(2), 71-75.
- Khaparde, R. B., & Pradhan, H. C. (2002). Procedural understanding: A neglected aspect of physics laboratory training. *Physics Education*, 19(2), 147-154.
- Lane, S., & Stone, C. A. (2006). Performance Assessment. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 387 - 432). Connecticut, USA: American Council on Education and Praeger Publishers.
- Lawson, A. E., Abraham, M. R., & Renner, J. W. (1989). *A theory of instruction: Using the learning cycle to teach science concepts and thinking skills*.
- Millar, R., Lubben, F., Gott, R., & Duggan, S. (1994). Investigating in the school science laboratory: conceptual and procedural knowledge and their influence on performance. *Research Papers in Education*, 9(2), 207-248. <https://doi.org/10.1080/0267152940090205>
- Ministry of Education. (2005). *Curriculum Specifications PHYSICS Form 4*. Putrajaya: Curriculum Development Centre.
- Ministry of Education. (2009). *Physics Practical Work Assessment (PEKA) Assessment Guide: SPM 2010*. Kuala Lumpur: Malaysian Examination Syndicate.
- Moni, R. W., Hryciw, D. H., Poronnik, P., Lluca, L. J., & Moni, K. B. (2007). Assessing core manipulative skills in a large, first-year laboratory. *Advances in Physiology Education*, 31, 266 - 269. <https://doi.org/10.1152/advan.00020.2007>
- Musial, D., Nieminen, G., Thomas, J., & Burke, K. (2009). *Foundations of Meaningful Educational Measurement*. Boston: McGraw-Hill Higher Education.
- National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academic Press.
- Ong, S. L. (2010). Assessment profile of Malaysia: high-stakes external examinations dominate. *Assessment in Education: Principles, Policy & Practice*, 17(1), 91-103. <https://doi.org/10.1080/09695940903319752>
- Prades, A., & Espinar, S. R. (2010). Laboratory assessment in chemistry: an analysis of the adequacy of the assessment process. *Assessment & Evaluation in Higher Education*, 35(4), 449-461. <https://doi.org/10.1080/02602930902862867>
- Reiss, M., Abrahams, I., & Sharpe, R. (2012). *Improving the Assessment of Practical Work in School Science*. Retrieved from <http://www.gatsby.org.uk/~media/Files/Education/Improving%20the%20assessment%20of%20practical%20work%20in%20school%20science.ashx>
- Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2014). *Measurement theory in action: Case studies and exercises* (2nd ed.). New York: Routledge.
- Solano-Flores, G., & Shavelson, R. J. (1997). Development of performance assessment in science: conceptual, practical and logistic issues. *Educational Measurement: Issues and Practice*, 16(3), 16-25. <https://doi.org/10.1111/j.1745-3992.1997.tb00596.x>
- Wenning, C. J. (2007). Assessing inquiry skills as a component of scientific literacy. *Journal of Physics Teacher Education Online*, 4(2), 21-24.

APPENDIX  
PRACTICAL TASK 1  
**Linear Motion at Constant Acceleration**  
(Teacher's Guide)

**Students' Prior Knowledge**

Before conducting the task, students should know the definition of acceleration and the formulae that can be used to calculate acceleration in linear motion.

- Acceleration = rate of change of velocity
- $a = \frac{v-u}{t}$ ,  $2as = v^2 - u^2$ ,  $s = ut + \frac{1}{2}at^2$

**Expected learning outcomes:**

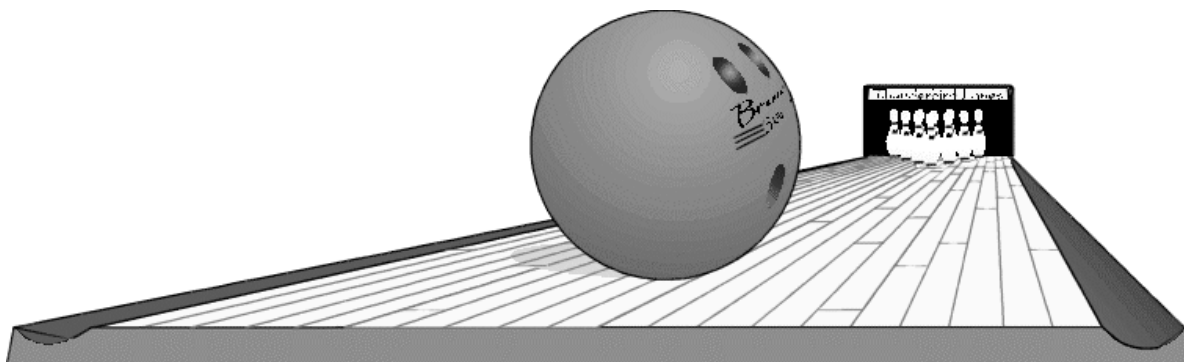
Students are expected to

- use apparatus and instruments available to determine the type of motion of an object
- determine whether an object is moving at constant velocity or accelerating from ticker tape
- calculate velocity and acceleration from ticker tape

**Expected performance:**

From the situation given and the apparatus provided, students are expected to

- design a simple experiment to investigate how an object moves down an inclined plane
- carry out the plan by taking precautions on how accurate data can be collected
- demonstrate how the velocity and acceleration of the object can be determined
- identify the problems faced during the experiment and ways to overcome them

**Situation:**

A bowling ball move faster as it rolls down the bowling track.

Similarly, the velocity of an object will increase as it moves down an inclined plane.

**Apparatus and materials:**

Ticker timer, ticker tape, metre rule, power supply, connecting wires, retort stand and clamps, trolley, runway, wooden block, wooden plank

**Assessment Scheme:**

Students will be observed while carrying out the task. Students should show clearly how data is collected including the number of times measurement is done.

Scores will be given based on the observation and evidence shown in the activity sheet.

## SCORING RUBRICS

### Domain of Design

#### 1. Identify variables correctly

0	1	2	3
Unable to identify any variables	Able to identify the one variable correctly	Able to identify any two variables correctly	Able to identify at least three variables correctly

#### 2. Design functional experimental set-up

0	1	2	3
Unable to draft the experimental set-up while designing the experiment	Able to draft the experimental set-up but the design is not functional	Able to draft functional experimental set-up but there is missing apparatus	Able to design a functional experimental set-up in the form of labelled diagram

#### 3. Determine suitable range for manipulated variable

0	1	2	3
Unable to determine the range while designing the experiment	Determine the range but is beyond the measurable (minimum or maximum) limit	Determine the range based on the apparatus available	Determine suitable range that is within the measurable limit

#### 4. Select suitable interval for values of manipulated variable

0	1	2	3
Unable to select systematic intervals while designing the experiment	Select interval for the values of manipulated variable which is not suitable with the task	Select intervals of manipulated variable which are systematic	Select suitable intervals which are systematic and compatible with the selected range

### Domain of Execution

#### 1. Choose suitable apparatus for measurement

0	1	2	3
Unable to choose suitable apparatus	Choose apparatus with assistance from others	Choose apparatus with unsuitable measurement range	Choose suitable apparatus without assistance from others

#### 2. Check the functionality of apparatus and instruments

0	1	2	3
Do not check the apparatus and instruments before starting the experiment	Check the physical conditions of the apparatus and instrument	Check the sensitivity of the apparatus and instruments	Check the functionality of the apparatus and instrument, and correct the errors, if any

#### 3. Set-up of functional apparatus

0	1	2	3
Set up of experiment is not compatible with the task	Try different set up before deciding the final set up	Set-up is functional but placement of apparatus may cause errors in results	Set up is suitable and functional

#### 4. Test-run the experimental set-up

0	1	2	3
Do not perform any test run	Check the stability of the experimental set up	Check the experimental set-up by trying out one measurement	Test the experimental set up with the minimum and maximum values that can be applied

5. Use measuring instrument with the correct techniques

- to be assessed separately for instruments used to measure each variable

0	1	2	3
Unable to use the instrument for measurement	Need guidance in using the measuring instruments	Use the instrument correctly but the measurements have errors	Use the instruments skilfully and correctly

6. Take precautions to improve accuracy of data collected

0	1	2	3
Show no precaution to improve accuracy of data	Show effort to avoid parallax error while taking measurements (or any one precaution)	Avoid parallax error and repeat experimental steps to improve accuracy (or any other two precautions)	Take all necessary precautions and checking the constant variables

7. Record all measurements

0	1	2	3
No record of measurement	Record measurement but not systematically	Record all measurements systematically in a table	Record all measurement systematically, to the correct significant figure and unit

8. Ensure safety (self, others and environment) in the laboratory

0	1	2	3
Do not show effort to ensure safety while carrying out the experiment	Perform experiment with minimum consideration for safety	Only shown effort in ensuring safety at the beginning of the experiment	Show effort to ensure safety throughout the experiment

**Domain of Analysis**

1. Perform correct calculation for secondary data and analysis of data

0	1	2	3
Shows no calculation for secondary data and analysis of data	Calculation is irrelevant to the task	Show relevant calculations but with errors	Show all relevant calculations correctly

2. Analyse data to obtain results / relationship

0	1	2	3
Do not show any effort in analysing the data	Perform analysis which is irrelevant to the task	Use suitable graph or chart for analysis but with limited graphing skills	Show good graphing skills in using suitable graph / chart to analyse data

3. State the correct relationship / make correct deduction

0	1	2	3
Make no deduction	Deduction made is irrelevant to the task	State a general relationship between variables	State the specific and correct relationship between the variables based on the graph

**Domain of Evaluation**

1. Conclude the findings of the experiment

0	1	2	3
Do not make any conclusion for the experiment.	Able to state a conclusion but is irrelevant with the experiment	Able to state a general conclusion for the experiment	Able to state a conclusion for the experiment based on the relationship

2. Make judgement on the accuracy / reliability of the results

0	1	2	3
Unable to judge the accuracy or reliability of the experimental results	Make judgement on accuracy without any explanation	Judge the accuracy of the results based on prior knowledge and assumptions	Judge the accuracy of the results based on errors occurred during the experiment

3. State the sources of errors

0	1	2	3
State sources of errors that is irrelevant with the task	State relevant source of error but give no explanation	Able to explain logically one sources of error that occur in the experiment	Able to explain logically more than one sources of error that occur in the experiment

4. Identify weakness in the design of the experiment

0	1	2	3
Unable to identify the weakness in the design of the experiment	Able to state weakness in the design but is irrelevant with the experiment	Able to explain logically one weakness in the design of the experiment	Able to explain logically more than one weakness in the design of the experiment

5. Suggestions to overcome weaknesses or sources of errors in the experimental design

0	1	2	3
Unable to give any suggestions to improve the experimental design	Able to suggest precautions but is irrelevant with the task	Able to suggest one suitable method to improve the experimental design	Able to suggest more than one suitable methods to improve experimental design

<http://www.ejmste.com>