Conceptions and Applications of Scientific Evidence in Physical Science Experiments Among Teacher Trainees in Kuching

by

Tan Ming Tang
Institut Perguruan Batu Lintang

ABSTRACT

The main aims of this study were to identify nine aspects of scientific evidence applied by science teacher trainees in conducting three physical science experiments, and also to describe the related conceptions of scientific evidence. The sample consisted of 87 trainees from a teacher training institute in Kuching. A quantitative methodology involving observation, interview and practical report was used. The results reveal that the three most applied aspects of scientific evidence were repeating measurements, using a suitable scale and a suitable table to organize data collection while the least applied was in giving suitable accuracy to their measurements. As to the sample's conceptions of scientific evidence, the results vary from about 87% possessing the right conception on repeats to only about 5% having the correct understanding on the external validity aspect. Overall, the findings of this study indicate that more trainees have an understanding of the routinized scientific evidence aspects and also in applying them more readily.

INTRODUCTION

Roberts and Gott (2004) refer to "concepts of evidence" or "scientific evidence" as the understanding of a set of ideas that underpin the collection, verification, analysis and interpretation of data in order to handle scientific data effectively. These concepts of evidence involve cognitive abilities such as deciding on how many measurements to take, over what interval and range, how to interpret the pattern in the resulting data etc. and are in turn underpinned by scientific skills. Hence, collecting and using evidence in an investigative task is viewed as a tool kit to help in judging an experimental study for its design, the reliability of the measurements, the validity of the sample and the quality of the resulting data and its interpretation.

In this study, the term "applications of scientific evidence" refers to the 'toolkit' of strategies and approaches applied by students to collect evidence in a scientific investigation. This study aims to investigate whether the respondents are choosing the right 'tools' and at the same time, using them to collect the relevant and correct data in a practical situation. As for the term "conceptions of scientific evidence", this study aims to probe whether the respondents have the right understanding behind the applications of the above toolkit in order to draw valid and reliable conclusions.

THE STATEMENT OF THE PROBLEM

One of the main aims of science education in Malaysia is to develop the potentials of individuals in an overall and integrated manner so as to produce Malaysian citizens who are not only scientifically and technologically literate but also competent in scientific skills. However, owing to a keen emphasis on examination-oriented teaching, 'the teaching and learning of science in some context, has becomes largely teacher-centered, thereby ignoring the development and mastery of scientific and thinking skills among students as required by the curriculum' (Sharifah, 2001; p. 42).

From here, there is thus a need to look into the training of the Malaysian science teachers on whether enough emphasis is placed on teaching and facilitating the use of scientific skills in science laboratory investigations. As regard to their training in the field of laboratory work, there is a lack of knowledge about their conceptions and applications of scientific evidence in science investigations. Thus, the question which arises is: "Do they really understand and at the same time apply the 'thinking behind the doing' (Gott and Duggan, 1995; p. 26) of these scientific procedures in physical science investigative tasks?". The answer to this question is of utmost importance because for science teacher trainees to be effective users and future facilitators of physical science investigative tasks in school, they need to possess appropriate conception of scientific evidence in order to be able to apply its vast repertoire of tools in the teaching and learning of science.

OBJECTIVES OF THE STUDY

Specifically, this study aimed to:

(i) Identify aspects of scientific evidence in the design, measurement and data handling stages that are applied by the science teacher trainees in conducting each novel physical science experiment.

(ii) Describe the science teacher trainees' conceptions of five scientific evidence aspects associated with the measurement reliability and design validity categories that are employed in carrying out each novel physical science experiment.

(iii) Investigate the relationship between the science teacher trainees' conceptions and applications of five scientific evidence aspects associated with the measurement reliability and design validity categories in each science practical and overall.

UNDERSTANDING THE KEY IDEAS OF SCIENTIFIC EVIDENCE

Being scientifically literate nowadays means more than just knowing factual science information. The emphasis now is more on the need of students to understand the procedures of scientific inquiry since it is the single most important ingredient necessary in students' investigative work to enable them to understand how scientific knowledge came to be established and be accepted generally as valid. Thus, secondary school and college level education should provide science students with an understanding, at an appropriate level, of the scientific account of the natural world and of the processes of scientific inquiry (Black, 1993). As a result, practical laboratory work is widely used as a teaching strategy and is also seen as crucial in developing an understanding of the procedures of scientific inquiry.

To describe this distinct set of conceptions relating to the procedures of scientific inquiry, Gott, Duggan and Roberts (2002) have come up with a list to define these concepts of evidence (Appendix A), which they believe, can be taught and which is a necessary but not sufficient condition for creative problem-solving. As an example, in order to ensure that the test or experiment conducted is fair, an understanding of the importance of isolating only the relevant variables while controlling others is necessary so that the resulting data collected is valid.

Some studies had been carried out to probe different aspects of students' conceptions on measurement reliability such as the need for repeats (Schauble, 1996; Varelas, 1997), the treatment of anomalous data (Chinn and Brewer, 1993), and the reliability of data sets (Allie, Buffler, Kaunda, Campbell & Lubben, 1998; Lubben and Millar, 1996). Yet other studies focused on different aspects of students' conceptions on experimental validity such as fair test (Schauble, Klopfer & Raghavan, 1991) and data collection strategies (Strang, 1990).

METHODOLOGY

In this study, three physical science experiments were designed by the researcher based on the topic "Force and Motion" to gather data on the science teacher trainees' conceptions and applications of scientific evidence. The aims of these experiments were to find the relationship between (i) the time of the bounce of a ping-pong ball with the height from which it was dropped, (ii) the time an object travelled through a tall column of water with the height from which it was dropped, and (iii) the depth of immersion of a plastic container with its carried load. To identify the science teacher trainees' applications of scientific evidence in the practical task, a quantitative methodology involving three instruments, a Laboratory Observation Instrument (LOI), the Students' Evidence Applications Rating Scale (SEARS), and the Students' Evidence Application Interview Scale (SEAIS), was used. All these instruments were developed by the researcher by referring mainly to Gott and Duggan's (1995) 'Principal Sources of Evidence Table'.

Two categories of conceptions of scientific evidence were probed in the practical task. In the measurement reliability category, the three scientific evidence aspects investigated were that of repeated trials, evaluating the trustworthiness of data and treatment of anomalous data. Whereas in the design validity category, the scientific evidence aspects of internal and external validity were probed. To elicit the necessary information regarding the trainees' conceptions of scientific evidence immediately after they had completed their practical task, a focused 'Level of Conceptions Interview Protocol' (LoCIP) instrument, designed by the researcher based on a revised Lubben and Millar's (1996) "Levels of Students' Understanding of the Collection and Evaluation of Empirical Data" Model and the researcher's own "Levels of Students' Understanding of the Design Validity" Model, were used.

RESPONDENTS

The sample studied consisted of 87 science teacher trainees from the January 2002 intake in a teacher training college in the Kuching Division. These prospective science teachers had been routinely taught to use certain experimental procedures such as identifying key variables, controlling variables for fair test, doing repeats, devising data table, and drawing graph. They were also taught to write their practical reports by following a standard format: Aim of experiment, Equipment/Materials used, Methodology, Data table, Graph, Conclusion and Precautions taken (Lembaga Peperiksaan, 2002). There were 45 male and 42 female trainees in this selected sample and their mean age was 22.6 years (sd = 2.1). The majority (94.3%) of the respondents in this selected college were art-streamed students, having only taken the General Science paper in their upper secondary school years.

FINDINGS AND DISCUSSION

Identification of Scientific Evidence Applied

To answer research question (i), percentages and frequency counts were used to aggregate the science teacher trainees' applications of nine aspects of scientific evidence. The quantitative approaches involved in the ratings of these scientific evidence aspects were observations, interviews and practical reports. The results of the ratings to identify the scientific evidence applied in each novel physical science experiment as a whole are presented in Table 1.

The results show that more respondents (~78 to 100%) applied the following scientific evidence correctly in conducting all three practical task: identifying key variables, controlling appropriate variables, repeating measurements, recognizing and identifying pattern in graph correctly, using a suitable scale and a suitable table to organize data collection. The majority of the respondents (at least 90% in each experiment) also possessed an excellent understanding of the substantive nature of science as were shown in their abilities to describe correctly the influence of at least one factor on the dependent variable in each practical before the commencement of their investigations.

The abilities of the trainees in applying all these scientific evidence aspects correctly in the practical task seem to be the result of the routinization of these strategies in the college practical program coupled with an excellent understanding of the substantive nature of science. As to the non-routinized aspects of using an appropriate range and interval and on evaluating the trustworthiness of data, only about 54% (47) to 72% (63) and 43% (37) to 51% (44) of the sample was implementing them correctly in each of the three experiments respectively. The least correctly applied evidence aspect by the sample in the practical task was in giving suitable accuracy to their measurements.

Identification of Conceptions of Five Scientific Evidence Aspects

The results (Table 2) reveal that the percentages of college students having the correct conceptions of five scientific evidence aspects in the practical task are as follows (in ascending order): rationale of repeats (~87% to 89%), fair test (~64% to 68%), how to evaluate the trustworthiness of data (~36% to 38%), treatment of anomalous data (~25% to 26%) and the external validity aspect (~5% to 9%).

For each science experiment, it was found that from about 87% (76) to 89% (77) of the respondents who possessed the right conception on the rationale of repeats, only about 25%

Table 1

Percentages and Frequency Counts of Applications of Scientific Evidence for Sample in Each Experiment as Obtained Through Observations and Interviews

	Experiment		
Scientific Evidence	First % (f)	Second % (f)	Third % (f)
Design Stage			
Identified key variables as independent and dependent correctly	82.8 (72)	87.4 (76)	78.2 (68)
Appropriate variables controlled for fair test	88.5 (77)	95.4 (83)	88.5 (77)
Measurement Stage			
Range and Interval appropriate	69.0 (60)	54.0 (47)	72.4 (63)
Scale – quantities chosen sensibly	100.0 (87)	100.0 (87)	100.0 (87)
To give suitable accuracy	8.1 (7)	6.9 (6)	6.9 (6)
Measurements repeated	100.0 (87)	100.0 (87)	100.0 (87)
Data Handling Stage			
Suitable table used to organize data collection	100.0 (87)	100.0 (87)	100.0 (87)
Pattern in data recognized and identified correctly	50.6 (44)	42.5 (37)	43.7 (38)
Pattern in graph recognized and identified correctly	87.4 (76)	78.2 (68)	83.9 (73)

Table 2

Percentages and Frequency Counts of Sample's Conceptions of Five Scientific Evidence Aspects in Each Physical Science Experiment.

Conceptions of Scientific Evidence	Physical Science Experiment (N = 87)		
	First	Second	Third
Repeats			
Wrong conceptions	11.5 (10)	11.5 (10)	12.6 (11)
Correct conceptions on rationale of repeats only	63.2 (55)	62.1 (54)	63.2 (55)
Correct conceptions on both rationale of repeats & way to handle repeats	25.3 (22)	26.4 (23)	24.1 (21)
Evaluating trustworthiness of data			
Wrong conceptions	62.1 (54)	64.4 (56)	63.2 (55)
Correct conceptions	37.9 (33)	35.6 (31)	36.8 (32)
Handling anomalous data			
Wrong conceptions	74.7 (65)	73.6 (64)	74.7 (65)
Correct conceptions	25.3 (22)	26.4 (23)	25.3 (22)
Fair test			
Wrong conceptions	33.3 (29)	35.6 (31)	32.2 (28)
Correct conceptions	66.7 (58)	64.4 (56)	67.8 (59)
External validity aspect			
Wrong conceptions	91.0 (79)	93.1 (81)	95.4 (83)
Correct conceptions	9.0 (8)	6.9 (6)	4.6 (4)

(22) to 26% (23) of them understood the correct way to handle the repeats measured in each of the three experiments conducted. When asked on why they repeated their measurements and how they subsequently handled the repeats, their typical responses are as follows: "By taking one reading only, it may be less accurate" (*Dengan mengambil satu bacaan sahaja, ia mungkin kurang tepat*) and "By finding its average" (*Dengan mencari puratanya*).

To test the college students' understanding on how to evaluate the trustworthiness of their measurements in both the measured and given data, about 36% (31) to 38% (33) of the respondents managed to provide a viable explanation on how they evaluated the trustworthiness of their measurements in each of the three experiments. By looking at the spread of the measurements, a typical response was "All the measurements are almost the same" (*Semua ukuran adalah hampir sama*). They also justified correctly the existence of an anomalous data point in the three experiments by pointing out significant differences in the measurements. A typical answer in response to the question, "From the data, is there any reading which is less believable and if there is, identify it and explain why you think so?" is illustrated as follows:

> "Yes, at the height of 80 cm where the time at t_4, that is 13.88 because this reading t_4 differs quite a lot from the rest of the readings at this height" *(Ya, iaitu pada ketinggian 80 cm dimana bacaan masa pada t_4 iaitu 13.88 kerana catatan bacaan t_4 ini agak jauh berbeza dengan bacaan yang lain pada ketinggian ini)*. (Respondent no.: 55, first experiment)

As to the conception on handling the anomalous data in each of the three experiments (an anomalous data was included in the results of each supplementary question), about 25% (22) to 26% (23) of the respondents identified and handled the anomalous data correctly when asked on whether a measurement which differs appreciably from most of the others can be included in calculating an average. This group of students realized that if the anomaly is included, it will affect the value of the mean calculated. When questioned on whether all the data in the supplementary question can be accepted or not, a typical response from this group of respondents is as follows:

> "Cannot. This is because the calculation of unacceptable data will influence the value of the mean at that measured height (*Tidak boleh. Ini kerana pengiraan data yang tidak boleh diterima akan mempengaruhi purata pada ketinggian yang diukur*). (Respondent no.: 36, first experiment)

In the design validity aspect, about 91% (79) to 95% (83) of the sample failed to have a real understanding of the importance of selecting a suitable range and interval in collecting their data in each of the three experiments conducted. When asked to provide a rationale for their preference of choosing an adequate range of independent variable values and at equal interval in their measurements, about 44% (38) to 47% (41) of the respondents described issues related to both graphical construction and interpretation in each of the three experiments conducted. The following excerpts contain portions of the respondents' rationales:

> To make easier the transfer of data or information onto the graph (*Untuk memudahkan pemindahan data atau maklumat ke dalam graf*). (Respondent no.: 55, first experiment)

> To make the construction of graph easier (*Untuk memudahkan membina graf*). (Respondent no.: 53, second experiment)
>
> Because with suitable interval, it will make reading and the drawing of graph easier (*Kerana selang kelas yang sesuai akan memudahkan bacaan dan melukis graf*). (Respondent no.: 57, third experiment)

Hence, nearly half of the sample chose a suitable range and appropriate interval in their measurements in order to make the construction and/or interpretation of the resulting graph drawn easier. Thus, these trainees failed to see the connection between the usage of a suitable range and interval in collecting data and the external validity of experimental design overall.

As to the concept of fair test, about 64% (56) to 68% (59) of the respondents in each experiment understood the necessity of controlling all the relevant variables in order that only the independent variable is allowed to affect the dependent variable. When questioned on whether they can consider their experiments to be a fair test, a typical excerpt of their responses is as illustrated below:

> "Yes, because there is no obstacle for us to do the experiment. Time is only influenced by the height" (*Ya, sebab tiada ada halangan untuk kita melakukan ujikaji. Masa dipengaruhi ketinggian sahaja*). (Respondent no.: 77, second experiment)

Of these respondents, about 59% (51) had the right conception on fair test in all three experiments and another 8% (7) in any two experiments. This shows that these respondents were quite consistent in their conceptions of the fair test.

Correlations between Trainees' Conceptions and Applications of Five Scientific Evidence Aspects

One of the main aims of this study, as shown in research question (iii), was to investigate the relationship between science teacher trainees' conceptions and applications of the scientific evidence aspects of repeats, evaluating the trustworthiness of data, anomalous data, fair test and external validity in each and all three physical science experiments. The total overall score for the conception and application of each scientific evidence aspect in the three experiments was computed by adding up the related individual evidence score of all three experiments. Pearson-Product-Moment correlation coefficients were computed for each experiment and overall. The results are presented in Table 3.

An examination of the correlations between the scores reveals that science teacher trainees' conceptions of the purpose of repeats, treatment of anomalous data and fair test were moderately correlated with their applications for the whole sample in each practical, and overall. As to the scientific evidence aspect of how to evaluate the trustworthiness of data, the Pearson correlation values obtained range from .66 to .82. The moderately high positive significant correlation values obtained for this scientific evidence aspect could be due to the fact that the reasoning required for this data interpretation task is only at a low cognitive level that is of recognition. However, weak correlation values (r = .14 to .31) were obtained for the external validity aspect across each and all three practical overall.

On inspecting the results for the applications of this external validity aspect, it was found that out of about 54% (47) to 72% (63) of respondents who applied it appropriately in each of the three experiments, only about 5% (4) to 9% (8) of them had a real understanding of the importance of selecting a suitable range and interval in collecting their data in each experiment. The findings here seem to suggest that a majority of these respondents had applied this evidence aspect without grasping the appropriate understanding of it yet. The interview results reveal that nearly half of the sample had the misconception that the reason for choosing a suitable range and appropriate interval in their measurements was to make the construction and/or interpretation of the resulting graph drawn easier.

Table 3

Correlations between Trainees' Conceptions and Applications of Five Scientific Evidence Aspects in Each Practical Task and Overall

Practical Task	Pearson Correlation Coefficient (r)				
	Rationale of Repeats	Evaluating the Trustworthiness of Data	Anomalous Data	Fair Test	External Validity
	All	All	All	All	All
First	.43**	.68**	.49**	.58**	.21*
Second	.43**	.82**	.62**	.46**	.25*
Third	.40**	.67**	.64**	.35**	.14
Overall	.50**	.78**	.54**	.46**	.31**

Note: * $p < .05$ level, ** $p < .01$ level (2-tailed)

As to the correlations between trainees' conceptions and applications of the five scientific evidence aspects in all three experiments overall, the correlation values obtained ranged from about .31 to .78. Hence, these college students' conceptions of these scientific evidence aspects tallied moderately well with their applications of the corresponding strategies in the practical task.

IMPLICATIONS AND RECOMMENDATIONS

Applications of Scientific Evidence

The findings concerning science teacher trainees' applications of scientific evidence in conducting the practical task suggest that more trainees tended to apply those experimental procedures which are incorporated as part and parcel of the practical work in their college's diploma of education program. Hence, the routinization of certain scientific evidence aspects such as repeating measurements, identifying key variables, controlling appropriate variables, using a suitable table for data collection and utilizing graph seems to help trainees to recall these 'tools' promptly to aid in problem-solving and in the process, enhance their procedural performance in the practical task.

As for those scientific evidence which were less or incorrectly utilized by the trainees, two aspects were found in the measurement stage (giving suitable accuracy and in using an appropriate range and interval) and another in the data handling stage (recognizing and identifying an anomaly in a table correctly). One major factor that could have contributed to the “underutilization” of these scientific evidence aspects is that trainees were not informed of the characteristics and relevance of these scientific evidence aspects in the investigative process. In this study, it was observed that a large number of teacher trainees appeared to be unaware of the importance of these underutilized scientific evidence aspects in the practical task. It can be seen from the lackadaisical attitude given towards these scientific evidence aspects which were not routinized as part and parcel of their college’s laboratory experimental procedures.

Since the findings of this study also indicate that a large majority of these college students possessed an excellent understanding of the substantive nature of science, it is suggested that teacher trainers use an integrated approach to teach subject-specific physical science concepts together with these underutilized aspects of scientific evidence in order that the teaching of this procedural knowledge be carried out in a meaningful context. According to McDermott (1990), this integration approach makes sense since students learn not only science concepts but also the evidence and reasoning that were used in developing these concepts.

Conceptions of Scientific Evidence

The findings seem to suggest that the routinization of scientific evidence aspects did enable more trainees to grasp the appropriate conceptions of the related evidence aspects. In other word, the constant applications of these scientific evidence aspects did increase the strength of the weightings in the memory units, thus aiding more college students in generating appropriate generalizations or inferences about the scientific evidence applied. This finding thus substantiates Rumelhart and McClelland’s (1986) connectionistic information processing model which states that learning can occur with gradual changes in connection (memory) strength by experience and an environment within which the system must operate. However, the results also reveal that there are still some college students who have yet to grasp the necessary conceptions of the routinized scientific evidence aspects from their practical experiences in the science laboratory. Thus, more time is needed for these college students to grasp the necessary understandings of their applications of the related scientific evidence aspects or they might not even grasp it at all.

The three critical areas were the sample’s failure to grasp the correct conceptions on how to evaluate the trustworthiness of the data, how to handle anomalous data and on the external validity issue in all three experiments. For the first aspect, teacher trainers might want to instruct them on the use of variance and standard deviation to evaluate the trustworthiness of the results. As to the second aspect, the students’ attention needs to be drawn to see the difference in the mean values calculated if the anomalous data is both included and excluded in their calculations. In this way, it will help the respondents to determine which mean value is the better representative of the measure of central tendency of the collected data. For the last aspect, teacher trainers can help the trainees by encouraging them to reflect on the consequences of failing to use an appropriate range and interval on the overall experimental generalization drawn. To do this, trainees are encouraged to construct graphs from the provided data sets. Besides encouraging the practice of self-regulation of learning among the trainees, teacher trainers might also want to instruct them on the correct way of constructing line graphs so that the full pattern in a relationship can be explored and displayed.

Since the results also indicate that a large majority of respondents possessed an excellent understanding of the substantive nature of science and that they also tended to integrate this conceptual knowledge in their evaluation of certain scientific evidence aspects, it is also recommended that the above aforementioned teaching strategies be incorporated as part and parcel of the integration approach. In this way, trainees can be instructed in acquiring not only the relevant procedural knowledge but also in understanding them in a meaningful context.

Correlations between Trainees' Conceptions and Applications of Five Scientific Evidence Aspects

The findings in this study show that there were positive relationships between the conceptions and their corresponding applications of scientific evidence in the practical task. The Pearson correlation values obtained for the scientific evidence aspects of the purpose of repeats, evaluating the trustworthiness of data, treatment of anomalous data, fair test and external validity ranged from about .31 to .78 for the whole sample in the three experiments overall. Although the correlational data cannot address causality, it appears that college students who had grasped the correct understandings of the scientific evidence aspects also tended to apply their related evidence aspects well and more readily. In view of this relationship, teacher trainers might want to promote appropriate applications of scientific evidence aspects among the science teacher trainees by helping them to acquire the correct understandings of the related evidence aspects.

To aid trainees' understandings of the ideas that underpin scientific evidence, explicit teaching of procedural knowledge through the aforementioned integrated approach utilizing a multi-method strategy is needed so that trainees may then be able to grasp the correct understanding inherent in the applications of these strategies. In fact, Millar et. al. (1994) stressed that 'this domain of scientific evidence contains ideas which must be taught: it cannot simply be assumed that they will be picked up through experience.' (p. 245)

CONCLUSION

The findings of this study show that the routinization of certain scientific evidence aspects enhanced trainees' conceptions and applications of the related scientific evidence aspects. However, other factors such as trainees' level of maturity, interest in science, experience in undertaking science investigative task etc. may also influence their conceptions and applications of scientific evidence in science investigative task. Thus, there is a need for more research to be undertaken in order that more light may be shed on this matter.

Also, since the present study is cross-sectional in nature, in that data was gathered from a group of final year science teacher trainees at a point in time, the findings could not provide information on the continuity of progress of trainees' conceptions and applications of scientific evidence. It is suggested that a multi-year longitudinal study that involves laboratory observations, interviews and other suitable strategies be undertaken to chart the trainees' progress in their conceptions and applications of scientific evidence from their time of entry until their graduation from the college.

Such a study could also be used to investigate the effectiveness of intervention activities undertaken to resolve trainees' concerns and in the process monitors the effectiveness of the whole implementation program. In this way too, a check can be put into place to ensure that there is a balance between introducing new scientific practices to trainees and the development of meaningful understandings on how science works and how it is based on the analysis and interpretation of evidence.

REFERENCES

Allie, B., Buffler, A., Kaunda, L., Campbell, B., & Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measurements. *International Journal of Science Education, 20*, 447-459.

Black, P. (1993). The purposes of science education. In R. Hull (ed.), *ASE Secondary Science Teachers' Handbook* (pp. 6-22). London: Simon and Schuster.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: Atheoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1-49.

Gott, R. and Duggan, S. (1995). *Investigative work in the science curriculum.* Buckingham: Open University Press.

Gott, R., Duggan, S. & Roberts, R. (2002). *Research into Understanding Scientific Evidence.* Retrieved June 10, 2003 from http://www.*Understanding%20 Scientific%20 Evidence.htm_*

Lembaga Peperiksaan (2002). *Format Pentaksiran Mata Pelajaran Fizik Mulai SPM 2003: Instrumen Contoh Bagi Kertas 3 – 4531/3.* Kuala Lumpur: Kerajaan Malaysia.

Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education, 18*, 955-968. _

McDermott, L. C. (1990a). A perspective on teacher preparation in physics and other sciences: The need for special science courses for teachers. *American Journal of Physics, 58,* 734-742.

Millar, R., Lubben, E., 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.

Roberts, R., and Gott, R. (2004). A written test for procedural understanding: a way forward for assessment in the UK science curriculum? *Research in Science and Technological Education, 22*(1), 5-21.

Rumelhart, D. E, & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102-119.

Schauble, L., Klopfer, L. E. and Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28*, 859-882.

Sharifah Maimunah Syed Zin. (2001). *Malaysia*. Retrieved August 12, 2003 from h*ttp://www.ibe.unesco.org/National/China/NewChinaPdf/11Malaysia.pdf_*

Strang, J. (1990). *Measurement in School Science.* Assessment Matters No. 2. London: SEAC/ EMU. _

Varelas, M. (1997). Third and fourth graders' conceptions of repeated trials and best representatives in science experiments. *Journal of Research in Science Teaching, 9,* 853-872.

Appendix A

Examples of Concepts of Evidence and their Definitions

(Adapted from Gott, Duggan and Roberts, 2002, pp. 1-12)

Reliability And Validity	Concepts of Evidence	*Definition*
Associated with design	**Variable Identification**	The design of an investigation requires variables to be identified and measured. The independent variable is the variable for which values are changed or selected by the investigator whereas the dependent variable is the variable the value of which is measured for each and every change in the independent variable.
	Fair Test	A fair test is one in which only the independent variable has been allowed to affect the dependent variable. Laboratory-based investigations ………. involve the investigator changing the independent variable and keeping all the controlled variables constant.
Associated with Measurement	**Relative Scale**	… the choice of sensible values for quantities is necessary if measurements of the dependent variable are to be meaningful e.g. in differentiating the dissolving times of different chemicals, a large quantity of chemical in a small quantity of water causing saturation will invalidate the results.
	Range and Interval	…….the range over which the values of the independent variable is chosen is important in ensuring that any pattern is detected. …….the choice of interval between values determines whether or not the pattern in the data can be identified.
	Choice of Instrument	Measurements are never entirely accurate for a variety of reasons…..of prime importance is choosing the (right) instrument to give the accuracy and precision required.

Non-repeatability	….repeated readings of the same quantity with the same instrument never give exactly the same answer. (Due to the inherent variability in any physical measurement, repeats are necessary to give more reliable data).

Table 1 (continued)

Examples of Concepts of Evidence and their Definitions

(Adapted from Gott, Duggan and Roberts, 2002, pp. 1-12)

Reliability And Validity	Concepts of Evidence	*Definition*
Associated with Measurement (con't)	**Accuracy or trueness**	….trueness is a measure of the extent to which repeated readings of the same quantity give a mean that is the same as the 'true' mean. According to Gott and Duggan (1995), an appropriate degree of accuracy is required to provide reliable data which will allow meaningful interpretation.
Associated With Data Handling	**Tables**	….a table is a means of reporting and displaying data. Simple patterns such as directly proportional or inversely proportional relationship can be shown effectively in a table but it has limited information about the design of an investigation e.g. control variables.
	Anomalous data	…patterns in tables or graphs can show up anomalous data points which require further consideration before excluding them from further consideration (the 'bad' measurement due to human error perhaps)
	Patterns	Patterns can be seen in tables or graphs or can be reported by using the results of appropriate statistical analysis and they represent the behavior of variables.
Associated with the evaluation of the complete task	**Reliability**	….the reliability of the design includes a consideration of all the ideas associated with the measurement of each and every datum. It relates to the question 'Will the measurements result in sufficiently reliable data to answer the question?'
	Validity	….the validity of the design includes a consideration of the reliability and the validity of each and every datum. Beside the above question, another overarching question is 'Will the design result in sufficiently valid data to answer the question?'