# DIFFERENCES IN EIGHTH GRADERS' STRATEGIES IN PERFORMANCE ASSESSMENT TASKS

**Toh Kok Aun & Ho Boon Tiong**
*Nanyang Technological University, Singapore*

Abstract: Using an eclectic of student interviews, observational techniques as well as written reports, the problem-solving strategies of 9 eighth graders, who were carrying out TIMSS-type performance assessment tasks (scientific investigations), were studied. The participants individually worked to find a solution to a task and provided details of his/her work in a written form. Simultaneously a trained researcher, on a one-to-one basis, recorded on a checklist every action taken by the participant. The participant was also interviewed at the end of the task to enable the researcher to clarify any missing links/doubts concerning a particular action. Twelve activities, constituting procedural and cognitive actions, performed by the participants are identified and described. The performance in these activities by the better students is compared with those of the weaker ones.

## The Solution Process In Performance Assessment Tasks

The research reported in this study was conducted to understand how students were thinking and the strategies they were applying when carrying out TIMSS-type performance assessment tasks (scientific investigations). It was also intended to understand the solution process in terms of the fundamental components that constitute any strategy or set of strategies used in these tasks. In the context of performance assessment these fundamental components are just as important as the macroscopic strategies required in the solution process. A set of generic fundamental components will assist teachers in their guidance of students working through a performance assessment task.

A literature search has provided useful insights (see, for example, Padilla, Okey and Gerrard 1984, Lock 1986, OCEA 1987, Kind 1999) as to what other researchers consider as fundamental components normally invoked during performance assessment. Twelve activities are identified by the researcher and they are classified into three levels – conceptual, procedural and reasoning. Conceptual level includes conceptualising the problem, what data to collect, and what variable to control. Procedural level takes into account the use of appropriate equipment, how many data points are needed, collection and organisation of data into tables/lists/graphs, application of algorithms, reading data from the table, comparing and computing differences. Reasoning embraces the area of identification of trends/patterns, drawing conclusions, predicting, extrapolating new points and relating findings to original question. Under these rubric the 12 activities identified are:

*At the Conceptual Level*

*Activity 1:* Sizing up the situation: Formulating testable ideas

*Activity 2*: Quick trial testing with equipment: Trying to get a better feel of the problem

*Activity 3:* Conceptualising the problem: Translating the problem to something familiar

*Activity 4*: Making decisions concerning variables: Deciding what to measure/ vary/ maintain constant to ensure a fair test

*At the Procedural Level*

*Activity 5*: Preparing for data entry: Getting tables ready for tabulating results

*Activity 6*: Setting up equipment for investigation: Carrying out task with the set-up, showing awareness of errors in measurement and the need for repeating measurements

*Activity 7*: Making measurements and tabulating readings: Consideration for how many data points are needed and concern for accuracy

*Activity 8*: Graphing data: Translating numerical data to graph, including accuracy ofgraphing work

*At the Reasoning Level*

*Activity 9*: Making sense of data collected and/or graph plotted: Making inferences from data

*Activity 10*: Re-checking of "suspicious" data and re-visiting the set-up for confirmation

*Activity 11*: Making hypothesis from data: Formulating tentative answers to the problem

*Activity 12*: Formulating conclusion(s): Refining/polishing the tentative answers and translating them into conclusions

## Methodology

An investigative task presents a problem situation whereby the solver explores the solution path to reach a goal based on some given information. Ausubel (1968) refers to such problem-solving situations as involving the cognitive representation of prior experience interacting with the components of the current problem situation to achieve a designated objective, popularly termed as "a solution". He considers problem solving as learning that bridges the gap between the learner's existing knowledge and the solution to the problem through transformation of information by analysis, synthesis, hypothesis formulation and testing, translation, and integration. The activities reported in this paper may therefore be more representative than the general strategies for performing these transformations.

The eighth graders involved in this study were between 12 and 13 years of age, and had approximately 5 years of prior exposure to science teaching (since science was only taught from third grade). Most of them were not likely to have been exposed to performance assessment. They had, however, been exposed to the basic skills of observation, controls, hypothesis testing, trends and patterns as part of science skills acquired in the laboratory. While these skills have been taught, there is still the need to know the extent to which students can put together these various component skills during performance assessment tasks. Many researchers have written about the whole being not the same as the sum of its parts (see, for example, Woolnough 1989, Solomon 1988).

In the spring of 1999 a sample of nine students, who could be followed through by the researcher in one-to-one observations, were interviewed, and their written reports examined. These nine students were randomly picked from more than 400 eighth graders from two schools. The students all carried out the same performance assessment task, viz. "*finding out how adding a single strip of adhesive tape to a 15 cm straw affects the mass that it can support*." Students would essentially need to develop a measurement strategy to determine the minimum force required to bend the straw which was supported at both ends, like in the case of a bridge.

Students were given an hour to carry out the task and to write up about it. They were encouraged to spend about half of that time carrying out the task to obtain the necessary data, and devote the remaining half of the time to communicate what they had done on paper. They were allowed to abandon the task at any time and start afresh if they so wished. The researcher used a detailed checklist to track each action taken by the student in a one-to-one observation. The checklist took into account the different pathways taken by the student to solve the task. This enabled the researcher to reconstruct what the student did, by merely studying the checklist. Follow-up interviews were pursued immediately at the end of the hour to enable the researcher to clarify with each student why a particular action was observed. This was important for establishing what took place for *Activity 1* through to *Activity 4* – these being at the conceptual level and are not always reflected as actions or apparent on paper in the written report. The evidence for *Activities 5 - 12*, which are at the procedural and reasoning level, were apparent from the checklists and the written reports. A decision was taken to refrain from interrupting the students while they were carrying out the task to avoid any giveaways. The researcher was also very careful not to provide leading questions during the post-task interview to mitigate the effects of leading the student.

## Findings And Discussion

Using the various data sources identified in the preceding section the researcher was able to triangulate the research evidence from one source with those obtained from other sources, thereby piecing together each student's performance in *Activities 1 - 12*. The nine students were ranked according to their total scores with those achieving more than 75% of the total possible scores being considered as high scorers (represented as H1 → H3), while those scoring below 50% (represented by L1 → L3) were reckoned as low scorers. The middle group (represented by M1 → M3) were those scoring between 50% and 75%. These are shown in Table 1. The performance of the 3 ability groups (high, medium and low) in each of the 3 levels (conceptual, procedural, and reasoning) is shown in Table 2. The findings in Table 2 are intended to bring out the differences, if any, in the level of conceptualisation, reasoning and procedural efforts, arising from their different abilities. Table 3 provides the correlation across the 3 levels, that between conceptual and procedural, conceptual and reasoning, and procedural and reasoning. The intention here is to find out whether good conceptualisation of the problem translates to better procedural efforts. Likewise the other correlation coefficients may reveal interesting insights. The discussion concerning the findings is taken up as assertions.

**Table 1: Students' Performance in Activities 1 - 12 (Numbers within [ ] show maximum possible score for each activity)**

| | H1 | H2 | H3 | M1 | M2 | M3 | L1 | L2 | L3 |
|---|---|---|---|---|---|---|---|---|---|
| Activity 1 [1] | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| Activity 2 [2] | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 |
| Activity 3 [1] | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Activity 4 [3] | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 1 |
| Activity 5 [1] | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Activity 6 [3] | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 1 | 1 |
| Activity 7 [3] | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 3 |
| Activity 8 [3] | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 |
| Activity 9 [2] | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| Activity 10 [2] | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| Activity 11 [2] | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Activity 12 [2] | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 |
| ***Total* [25]** | 22 (88) | 21 (84) | 19 (76) | 18 (72) | 16 (64) | 14 (56) | 11 (44) | 10 (40) | 9 (36) |

Numbers within () are percentages.

**Table 2: Performance of Different Ability Groups at Conceptual/Procedural/Reasoning Levels (Numbers within parenthesis refer to percentages)**

| | High Group | Middle Group | Low Group | *Total* |
|---|---|---|---|---|
| Conceptual level | 20 (95) | 15 (71) | 10 (48) | 45 (71) |
| Procedural level | 27 (90) | 23 (77) | 14 (47) | 64 (71) |
| Reasoning level | 15 (63) | 10 (42) | 6 (25) | 31 (43) |
| ***Total*** | 62 (83) | 48 (64) | 30 (40) | 140 (62) |

**Table 3: Correlation Coefficients Across the 3 Levels (All values are significant at the 0.05 level)**

| | Conceptual | Procedural | Reasoning |
|---|---|---|---|
| Procedural | 0.86 | | |
| Reasoning | 0.70 | 0.80 | |

*Assertion 1: Conceptual level activities are strong determinants of success in procedural level activities*

Performance assessment tasks are ideal situations to observe how students can integrate thought with action. When students are involved in the actual carrying out of performance assessment tasks, they are engaged in some purposeful activity. The ability to carry out such tasks is more than simply knowing about things; it is also more than simply being able to do things. Rather, it is being able to effectively integrate rational and creative thought with the ability to take action through manipulating materials to tackle the task at hand. The inter-dependence of thinking and doing, and not just one or the other, is crucial when students are involved in performance assessment tasks. This means that the mere drill of component skills by themselves can become quite mechanistic. It involves "doing" without much "thinking", especially after repeated trials. Both doing without thinking, and thinking without the opportunity of doing, are equally undesirable situations.

The high correlation coefficient of 0.86, significant at the 0.05 level, reflects the inter-dependence of thinking and doing (see Table 3). Those who are able to conceptualise the problem well are seen to be better in their procedural activities. This is very clearly reflected in the outcomes shown in Table 2. The deterioration in the quality of conceptual level (thinking) activities from the high group (with 95%) to the middle group (with 71%) and the low group (with 48%) is likewise reflected in corresponding drops in the quality of performance in the procedural level (doing) activities. Even though *Activities 1 - 4* are conceptual level activities, there are strong links between them and the procedural level activities. It is not possible to perform *Activities 1 – 4* in the absence of a task framework. A degree of appreciation of the context of the task is required in the performance of the conceptual level activities. By the same token, detailed planning for the task at hand, which involves in-depth thinking through of the conceptual level activities, will act as strong determinants of success in procedural level activities. The high correlation is therefore something to be intuitively expected. One cannot expect to perform well at the procedural level without sufficient thinking through of the problem posed by the task, and conceptualising well concerning the action to take. The procedural activities enable students to put into action the plans drawn up under the conceptual level activities.

*Assertion 2: The standard of achievement at reasoning level activities are much lower than those at the conceptual and procedural level activities.*

The activities at the reasoning level require the student to “make sense” of the data collected and/or graph plotted to explain, translate or re-work the information into another form. Students may require to re-visit the experimental set-up to verify something missed earlier or to follow-up something of a suspicious nature. It may also be necessary to explore the limitations of the data collected, the inaccuracies embodied in them, and the extent to which extrapolation could be made. They may also use the data collected to predict what might happen in other similar situations.

These reasoning level activities are normally considered to be fairly high up the hierarchical scale of intellectual reasoning and a certain degree of anticipation and foresight are necessary ingredients for performing well. A helicopter perspective of the requirements of the task may not be a bad thing here. For the typical eighth grader, these reasoning level activities would be ranked a number of notches up the difficulty ladder in relation to the conceptual and procedural activities. While the conceptual and procedural level activities are related to the task at hand, the reasoning level activities require maturity in the use of scientific reasoning to perform well. Those who have not had prior exposure to the challenge of scientific reasoning are likely to give a cursory answer as the conclusion to the task, without due diligence paid to requirements of *Activities 9 - 12*. The significant all round drop in the level of performance in the reasoning level activities (43%) *vis-à-vis* the conceptual and procedural level activities (71%) is a testimony to this.

*Assertion 3: The purposeful thinking through of the problem by the better performers is in stark contrast to the poor thinking applied by the weaker performers.*

One-to-one observation of each of the nine subjects, together with the post-task interviews, has provided valuable insights concerning the differences in performance of the better students compared with the weaker ones. While the weaker students are seen to be employing trial-and-error strategies to find a solution to the task, the better ones are using more systematic planning in identifying the variables to measure, to vary and to control, when seeking a solution to the task.

The performance in each of the twelve activities (see Table 1) is a reflection of how well the subjects have grasped/understood the problem. The strategies adopted for *Activities 3 - 7* reflect the varying degrees with which each of them has thought the problem through. The weaker ones, for example, did not appear to understand the need to control variables to ensure that the test was fair. However the concept of a fair test was evidently with these weaker subjects when they were shown, during post-test interviews, the scenario of a young man competing with an elderly woman in a sprint race:

- the trim young man in track shoes, using starting blocks, all ready to go;
- the portly and elderly woman in high heels, with no starting blocks, and obviously not in a state of preparedness.

All of them “saw” that it was not fair, and could identify the reasons why the sprint race was not fair. Yet when there was a shift from the sports arena to a different task situation they did not “see” the need for adequate controls to ensure a fair test. This was a transfer problem, where the weaker students were not transferring what they knew about a particular concept in one situation to another. The fair test concept and other similar situations within the performance assessment reflected the poor thinking through of the problem by the weaker students.

The better performers, on the other hand, were more purposeful in their thinking through of the problem. Their systematic planning at the conceptual level translated to better action at the procedural level. Their purposeful thinking in the use of reasoning skills had resulted in better quality when distilling inferences from the data collected, making hypothesis, and formulating conclusions. What emerged was that the better students were making more judicious decisions concerning their observations, data and other relevant information to support their formulation of conclusions.

The better language command of the more able students had also meant better understanding of the problem posed by the task. The written product of the better performers is evidence of this, as a piece of written work is a reflection of the writer's ability to retrieve, synthesise and organise information.

## Conclusion

The analysis of different students' performance in the fundamental components making up the TIMSS-type performance assessment tasks was used to inform how low-achieving performers were doing *vis-à-vis* the high-achieving ones. The results of this study indicate that there are structural differences in the way in which better students go about a performance assessment task compared to their weaker counterparts. The random and trial-and-error strategies used by weaker students reflect the low level of conceptualisation of the task. The performance of better-performing students is characterised by systematic planning and purposeful thinking in their attempts to seek a complete solution to the problem posed by the performance assessment task.

## References

Ausubel, D.P. (1968). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart and Winston.

Kind, P.M. (1999). Performance assessment in science - what are we measuring? *Studies in Educational Evaluation, 25* (3), 179-194.

Lock, R.J. (1986). *Assessment of Science Practical Skills in 15 Year Olds*. University of Leeds: Unpublished doctoral thesis.

OCEA (1987). *OCEA (Oxford Certificate of Educational Achievement) Science Teachers' Guide*. Oxford: Oxford University Press.

Padilla, M.J., Okey, J.R., and Gerrard, K. (1984). The effects of instruction on integrated science process skills achievement. *Journal of Research in Science Teaching, 21* (3), 277-287.

Solomon, J. (1988). Learning through experiment. *Studies in Science Education, 15*, 103-103.

Woolnough, B.E. (1989). Towards a holistic view of processes in science education (or the whole is greater than the sum of its parts, and different), in Wellington, J. (Ed.) *Skills and Processes in Science Education: A Critical Analysis.* London: Routledge.