Issues and Problems Concerning the Use of Diagnostic Assessment Questions (DAQ) and Formative Assessment In Brunei Schools

Chong Kui Kian
University of Southampton


Abstract

This paper reports on and discusses issues and problems associated with the use of DAQ and formative assessment based on a study of lower secondary classes in 5 government schools in Brunei Darussalam. The study is part of a larger research carried out in a quasi-experimental set-up involving 13 experimental classes and 11 control classes to test the efficacy of DAQ and formative assessment in Brunei schools. A test instrument comprising 26 DAQ is used to provide a baseline of prior knowledge in a pre-test and assess for learning gains after 2-4 weeks of teaching in a post-test. The experimental group was given a teaching intervention package that uses DAQ in a formative assessment framework, while the control group was taught in the ‘usual’ way. Test score analysis, interviews with teachers and students, and classroom observations provided information on problems associated with the use of DAQ and formative assessment in Brunei. They include DAQ designed for mental capacity and language proficiency higher than the level of lower secondary students; inability of some teachers to use them effectively; some teachers lacking knowledge on the subject matter; and classroom culture that does not work well with interactive styles requiring honest feedback. The paper concludes by suggesting that, for DAQ and formative assessment to be effective in Brunei lower secondary schools, (a) DAQ needs to be made simpler and used in small steps, (b) language used in DAQ should be more appropriate for lower secondary students, (c) teachers must be given training, and (d) DAQ and formative assessment must be incorporated into the curriculum.

Introduction

Millar and Hames (2001) explained diagnostic assessment questions (DAQ) as questions that can provide insights into how students understand specific ideas by analysing the responses. Most DAQ comprise of (i) a propositional knowledge statement and (ii) the explanation. A number of studies on the use of Diagnostic Assessment Questions (DAQ) in the classroom (e.g. Millar & Hames, 2001; Black et al, 2003; Millar et al, 2006) have reported significant improvement in both teaching and learning. The majority of the studies carried out so far were on British schools. This study is part of a larger research that is investigating the efficacy of DAQ and formative assessment in a different context: lower secondary government schools in Brunei Darussalam. The underlying purpose of the research was to find a way to address the declining trend in examination performance in lower secondary science.
In the Brunei education system, students spend 3 years (Forms 1 to 3) at the lower secondary level at the end of which they will sit for a public examination to determine whether they are promoted to the upper secondary level and the subjects that they are allowed to take. After 2 years in Forms 4 and 5, they will sit for another public examination: the Brunei GCE Ordinary Level Examination, which is set by the Cambridge Board of Examination in the U.K. The education system is however presently being revised so that beginning in 2009, students will spend 2 years in lower secondary and 3 years in the upper secondary under the new SPN21 (Sistem Pelajaran Negara Abad 21, or National Education System for the 21st Century).

While the country’s education system is based on the British system, including the use of English as the medium of instruction, its classroom culture is quite different. In Brunei schools, there is greater emphasis on conformance to cultural values, such as “proper” dress code and behaviour, respect for the teacher and older persons, and belief that the teacher is knowledgeable and always correct. Hence, classroom interaction is generally authoritative (Scott & Mortimer, 2006) and students tend not to question what was taught in class. In such an environment, incorrect explanations would be readily accepted as correct. In general, students in academically weaker classes tend to have poor command of the English language.

The summative form of assessment still dominates the Brunei education system. Although some semblance of formative assessment strategies are increasingly being used by more committed teachers, they are not widely practised. Several studies (e.g. Kellough & Kellough, 1999; Black & Harrison, 2000; Carl, 2002; Millar et al, 2006), however, have shown formative assessment strategies to enhance teaching and learning primarily through the feedback processes inherent in the practice.

In the Brunei study on which this paper is based, however, the research encountered several problems that may have affected the efficacy of DAQ and/or formative assessment when experimented in 24 lower secondary (Form 2) classes from 5 government schools. Issues and problems associated with the use of DAQ and formative assessment strategy were identified through classroom lesson observations and interviews with the teachers and students who participated in the study. The paper will begin by describing the study group and methodology used, followed by a discussion on the key issues and problems identified. It will conclude with a set of suggestions on how to resolve, where possible, issues and problems associated with the use of DAQ and formative assessment in Brunei schools.

The Study

The findings from this study are based on a pilot study, which was carried out in late March-early April 2006, and the main research conducted in May- August 2007. In the pilot study, 113 boys and 95 girls (age 13-14 years old) from 3 Government Secondary schools participated in taking DAQ tests and interviews to gain better insights into their experience and views regarding DAQ. The sample included “good”,

(MOE, 1997-2005) by conducting a quasi experiment using a well-established instrument, in this case, DAQ.
“average” and “weak” classes, classified on the basis of teachers’ judgment of the students’ general level of academic ability.

The main experiment originally involved 24 classes from 5 schools with a total sample size of 531 students and 8 teachers divided into an experimental and a control group. However, due to the actions of 2 teachers that comprised the integrity of the experimental conditions, the data from one school is excluded from the analysis. The remaining experimental group is made up of 296 students (137 males, 140 females) in 5 “good” classes, 4 “weak” classes and 1 “average” class. The control group is made up of 235 students (138 males, 97 females) in 2 “good” class, 3 “average” classes and 3 “weak” classes. There are more female students than male students in “good” and “average” classes. In contrast, male students outnumber female students in “weak” classes. The participating teachers aged between 21 and 49, with teaching experiences ranging from 1 to 20 years respectively. They either hold Bachelor Degrees in Education (General Science or Science) from Universiti Brunei Darussalam (UBD) or Diplomas or teaching certificates from the Teacher Training College. Allocation of classes into experimental and control groups was done after consultation with participating teachers.

The study involved setting up a quasi-experiment where all classes were first given a pre-test using a test instrument that consisted of 26 DAQ before students were taught the topic being tested in order gauge their prior knowledge on the topic, in this case, basic electricity concepts. This is followed by a teaching period, where the experimental group was taught the topic (basic electricity) with the help of an intervention packaged provided by the researcher. The intervention packaged was designed based on the national science syllabus and incorporated the use of DAQ to support formative assessment. The control group was taught the topic in the ‘usual way’ of each participating teacher. The teaching intervention period lasted between 2 to 4 weeks depending on individual teachers’ teaching commitments, after which the students were given a post-test using the same 26 DAQ to find out if there is any significant improvement in test performance, and whether the experimental group had performed significantly better than the control group.

Students were not given feedback on how they had performed after the pre-test to avoid students giving correct answers in the post-test based on what they have remembered rather than actually understanding the questions and concepts being tested. The correct answers were also not given to the participating teachers and the researcher did not discuss the questions with the teachers until the end of the experiment. The DAQ used in the test instrument were taken from the DAQ questions bank developed by the Evidence-based Practice in Science Education (EPSE) project network team from the UK (Millar and Hames, 2001).

Classroom observations were also carried out to study how teachers delivered the intervention lessons as well as to capture the classroom culture and interaction patterns in the different classes. Feedback from participating teachers and students regarding their experience and feeling towards DAQ and formative assessment were obtained through interviews. Problems and issues associated with the use of DAQ and formative assessment were identified from interview and classroom observation as well as by comparing the pre-test post-test performance of the participating classes.
Findings

The mean test scores for both the experimental and control groups were found to be very similar in the pre-test (experimental mean 3.74, std. dev. 2.306; control mean 3.46, std. dev. 1.763) but quite different in the post-test (experiment mean 5.51, std. dev. 3.239; control mean 4.99, std. dev. 2.631). The Mann-Whitney (MW) test for two unrelated samples was used to test for significant difference between the groups in the pre-test as well as in the post-test. The Wilcoxon Signed Ranks (WSR) Test was used to test for significant improvement between pre-test and post-test performance for each group. The same tests were also used to compare different class-pairs, i.e. experimental and control classes taught by the same teacher in the pre-test and post-test, and each class’ difference in scores between pre-test and post-test.

The mean pre-test scores for the experimental group and the control group were both very low, i.e. 3.74 (14.4%) and 3.46 (13.3%) respectively. The MW test indicated no significant difference (P > 0.05) in prior knowledge between the two groups, despite the fact that the experimental group consisted of more “good” students. The mean post-test scores for the experimental and control groups were 5.71 (22.0%) and 4.99 (19.2%) respectively. The MW test indicated a significant difference (P < 0.05), suggesting that the intervention program may have had a real impact on teaching and learning. This is supported by the WSR test on the mean pre-test and post-test scores, which indicated a significant improvement (P < 0.05) for the experimental group after the intervention period.

However, the control group, which was taught in the ‘usual way’, also showed significant (P < 0.05) improvement in the post-test. This perhaps simply confirmed that students generally would perform better after being taught the lesson than when they had not been taught the lesson. The significant difference in post-test performance between the experimental and control groups however suggested that the experimental group have learned the lessons better. It must be noted again, however, that the experimental group comprised more “good” students than the control group and is therefore expected to learn better. It is therefore unclear if the intervention made any difference.

Table 1: Experimental Group Pre-test Post-test Results

<table>
<thead>
<tr>
<th>No.</th>
<th>Class</th>
<th>Pre-Test (%)</th>
<th>Post-Test (%)</th>
<th>Post-Pre (%)</th>
<th>WSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2B (Good)</td>
<td>21.2</td>
<td>30.0</td>
<td>8.8</td>
<td>S</td>
</tr>
<tr>
<td>2</td>
<td>2E (Good)</td>
<td>12.3</td>
<td>18.1</td>
<td>5.8</td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>2A (Good)</td>
<td>19.6</td>
<td>27.7</td>
<td>8.1</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>2A (Good)</td>
<td>17.7</td>
<td>31.5</td>
<td>13.8</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>2A (Good)</td>
<td>15.4</td>
<td>30.8</td>
<td>15.4</td>
<td>S</td>
</tr>
<tr>
<td>6</td>
<td>2D (Average)</td>
<td>14.6</td>
<td>20.0</td>
<td>5.4</td>
<td>S</td>
</tr>
<tr>
<td>7</td>
<td>2H (Weak)</td>
<td>5.8</td>
<td>12.3</td>
<td>6.5</td>
<td>S</td>
</tr>
<tr>
<td>8</td>
<td>2E (Weak)</td>
<td>10.4</td>
<td>15.4</td>
<td>5.0</td>
<td>S</td>
</tr>
<tr>
<td>9</td>
<td>2K (Weak)</td>
<td>11.9</td>
<td>13.1</td>
<td>1.2</td>
<td>NS</td>
</tr>
<tr>
<td>10</td>
<td>2H (Weak)</td>
<td>8.5</td>
<td>11.9</td>
<td>3.5</td>
<td>NS</td>
</tr>
</tbody>
</table>

WSR = Wilcoxon Signed Rank Test
Tables 1 and 2 provide some details of the poor performances in both pre-test and post-test for individual classes. In the experimental group, “good” and “average” classes scored between 12-21% in the pre-test (mean 16.8%) and between 18-32% in the post-test (mean 26.4%). Weak classes, in comparison, scored between 6-12% and 12-15% in the pre-test (mean 9.2%) and post-test (13.2%) respectively. In the control group, “good” classes fared only slightly better than “average” classes, scoring between 15-18% (mean 16.5%) in the pre-test and 21-23% (21.5%) in the post-test, while average classes scored between 13-16% (mean 14.2%) in pre-test and between 19-24% in post-test (21.7%). “Weak” classes scored between 7-13% in pre-test (mean 10.8%) and between 12-19% in post-test (mean 15.6%). It can be seen that there is little difference between the “Good”, “average” and “weak” classes and their counterparts in the control group in their pre-test performance, suggesting no difference in prior knowledge. As with the combined class post-test data above, “good” and “average” classes in the experimental group performed slightly better than their counterparts in the control group. There is little difference between the “weak” classes in the experimental and control group in post-test performance.

Table 2: Control Group Pre-test Post-test Results

<table>
<thead>
<tr>
<th>No.</th>
<th>Class (Type)</th>
<th>Pre-Test (%)</th>
<th>Post-Test (%)</th>
<th>Post-Pre (%)</th>
<th>WSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2D (Good)</td>
<td>17.7</td>
<td>21.2</td>
<td>3.5</td>
<td>NS</td>
</tr>
<tr>
<td>2</td>
<td>2B (Good)</td>
<td>15.4</td>
<td>21.9</td>
<td>6.5</td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>2H (Average)</td>
<td>15.8</td>
<td>22.7</td>
<td>6.9</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>2D (Average)</td>
<td>13.8</td>
<td>18.5</td>
<td>4.6</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>2C (Average)</td>
<td>13.1</td>
<td>23.8</td>
<td>10.8</td>
<td>S</td>
</tr>
<tr>
<td>6</td>
<td>2F (Weak)</td>
<td>11.9</td>
<td>18.5</td>
<td>6.6</td>
<td>S</td>
</tr>
<tr>
<td>7</td>
<td>2F (Weak)</td>
<td>7.3</td>
<td>11.5</td>
<td>4.2</td>
<td>S</td>
</tr>
<tr>
<td>8</td>
<td>2G (Weak)</td>
<td>13.1</td>
<td>16.9</td>
<td>3.8</td>
<td>NS</td>
</tr>
</tbody>
</table>

WSR = Wilcoxon Signed Rank Test

Based on the WSR test, 8 out of the 10 experimental classes and 6 out of 8 control classes showed significant improvement between pre-test and post-test (Tables 1 and 2). The 2 experimental classes that did not show any real improvement were “weak” classes. However, 2 other “weak” classes did perform better between pre-test and post-test. In the control group, 2 weak classes also performed significantly better in the post-test, but 1 weak class did not achieve any significant improvement. The other control class that did not show improvement, interestingly, is a “good” class. When the classes are compared in experimental-control class pairs, the pattern is still not very clear, with various class-pair combinations of “good”-“good”, “good”-“average”, “weak”-“weak”, etc., classes showing significance and no significance in the different permutations.

Problems and Issues

This poor performance and lack of pattern in the data suggest that student performance may be influenced by other factors, such as teacher’s attitude and capability, teaching quality, students’ aptitude, classroom environment, etc. In a nation-wide study of conceptual understanding in science using DAQ in Taiwan, Chiu et al. (2007) found that culture has a profound influence on how students responded to
DAQ and hence, affected the quality of the data obtained. The poor performance in the Brunei study could also be explained partly by DAQ design, which was found to be not very suitable in the Brunei context. The study identified 7 main problems when experimenting with the use of DAQ and formative assessment in Brunei schools. They include the design of the DAQ, the teacher’s ability to use it effectively, proficiency in the English language; teacher’s knowledge on the subject matter; and classroom culture.

1. **Too Many Options of Similar Answers**

The multiple-choice sections of the DAQ are deliberately designed to test if the student can differentiate between the various scientific terms and concepts used. Each option provides insight into how the student understands or misunderstands the different terms and concepts. This is the primary objective of a diagnostic question. The words used and the manner in which the sentences are constructed therefore must be very precise in order to differentiate between the closely related or associated terms and concepts as accurately as possible. However, while the DAQ thus constructed may be very clear to a researcher or specialist, high school students (and even teachers), in general, do not have the same level of clarity of concepts and in the case of Brunei, level of proficiency in the English language used. This was found to be the case in the present study, where many teachers and students expressed difficulty in choosing the correct answer because (a) “too many” choices were given for each part of the questions; (b) some of the choices given “sounded similar”; and (c) some of the sentences sounded “confusing”.

Furthermore, many students and some teachers were found to use a number of interrelated terms interchangeably. For example: current, electricity and energy (and even battery) are synonymous to many students; electric charge is hardly used. Consequently, they often find at least 2 of the 5 options given to mean the same thing, as they were unable to differentiate between the terms. It was found that many people in general (perhaps only in Brunei) tend to use words in an approximate sense, i.e. using a word to mean something else that is related, even though it does not mean what was intended. For example, the word “confusing” is sometimes used when the person really meant that he is unfamiliar with some of the words used and therefore decided not to try to understand what is being asked. It is also sometimes used when the person feels that he does not want to answer the question in case he gives an incorrect answer.

The words and sentence structure used in DAQ, while serving to maintain precision of meaning, may be regarded as a design flaw in the Brunei context. This is because it compromises the instrument’s function to diagnose the exact cause of the problem. A student choosing an incorrect answer from 2-3 options that sounded similar to him or her does not necessarily indicate misconception of the scientific concept being tested. It could be that his or her level of clarity of the concept or language proficiency is not up to the level used in the DAQ and so is unable to differentiate between some options.

Taking Q16 in the test instrument to illustrate the problem described, this DAQ shows 3 identical bulbs connected in series to a battery. Its first tier asked how bright would each bulb be lighted and offered 7 options to test the student’s understanding of
current flow in a series circuit. Several teachers commented that seven options are too many and believed that students would be confused by what is presented to them. This is because they (the teachers) also found difficulty in choosing the correct answer from the options given. One experienced teacher said that the many different ways that DAQ presents the same concept makes the questions “unclear”. It is unclear however how much his view is affected by a common observation in Brunei classrooms where doubt tends to creep into one’s mind when asked the same question ‘too many’ times, suspecting a ‘trick question’ that nearly everyone – teachers and students alike – is likely to have experienced having to answer at some point in their lives.

The second tier asked for the explanation and provided 6 possible answers, two of which sounded similar to many students. Option 3 said that “the electric current is the same all round the circuit”, while option 4 said that “the electric current is shared equally between the 3 bulbs”. The two options sounded similar because to many students, electric current and voltage across the bulbs or the battery means the same thing. In the words of the students interviewed, one said that “voltage is the measure of current”, while another said it is the “measure of the battery”. They see the battery as the source of current, and do not differentiate between voltage and current, and so current is thought to be divided among the 3 bulbs (when they are actually thinking of voltage, which would be divided among the bulbs) and also is the same all round the circuit (the correct understanding of current in a series circuit). Millar et al (2006) noted similar observations among UK students, where students do not differentiate electrical energy and current, and often see electric circuits in term of a “source” (battery) and “consumer” (bulbs) and hence, the bulbs would consume the current/energy equally.

2. Unfamiliar/Unfriendly Format

Several teachers were concerned that their students will find the format and layout used in DAQ confusing. They said that the questions were too wordy and layouts where multiple choice questions were arranged in 2 columns in one section and in 1 column in the next section with diagrams in between are not easy to comprehend and students would need time to orientate themselves in order to understand them. They said that each DAQ contain too much information, which require time for the students to read and understand, and many would lack the mental stamina to take in so much information. They were also concerned that it would consume time and resources (e.g. photocopying many DAQ test sheets for over 100 students at a time), which many schools do not have. The majority of the students interviewed however said that they did not have serious problems comprehending the DAQ. However, the teachers are not completely mistaken as many students in general lack the mental stamina to process large amount of information.

3. Students Find Difficulty in Expressing their Thoughts

Many students could not answer the open-ended part of the DAQ correctly. Of the 6 questions that required an explanation, students consistently performed considerably better when the correct answers minus the explanation part were considered than when the explanation was taken into account. This is true for both experimental and control groups. For example, while 57% of students in the experimental group knew
that the bulb in Q1 is “not lit” because the circuit is incomplete, only 17% was able to explain it correctly. In comparison, 48% of the control group got the correct answer but only 9% was able to give the correct explanation. There are differences between “good”, “average” and “weak” classes in expressing ideas in English. In Q1, 48% of “good” students in the experimental group got the correct answer but only 25% could explain it. In contrast, 49% of “weak” student got the right answer but only 2% was able to give the correct explanation. For Q25, where the open-ended part of the question requires that the student give the correct value, the difference is much smaller (49% for multiple choice part, and 40% to include the open ended part).

Students did particularly poorly in Q4 and Q7, which presented the same question differently (Q4 shows the circuit drawn realistically while Q7 uses a circuit diagram), and Q26, which required an explanation for current flow in parallel circuits.

To illustrate the point that students have trouble expressing their thoughts, the following are some of the explanations given for part 2 of Q1, where the correct answer should be that the circuit is not complete or is open.

- “The battery is not connected”. The diagram however shows that the 2 cells are connected to the lone bulb from 1 terminal, although not to one another from the other terminal. This suggests that the student understood the concept but is unable to express his or her understanding completely.
- “There is no other wire”, meaning, another length of wire is needed to complete the circuit. Again, the student is unable to explain fully the correct answer.
- “The cells are separated but not combined”. This could mean that the student intended to say that the cells are not connected to complete circuit (which is correct) or that they are used to seeing cells placed side by side in circuit diagrams and therefore thought that the bulb would not light up because the cells are separated one on either side of the bulb (which is a misconception).

4. Understanding the Questions

While the medium of instruction in all the participating schools is English, some students, particularly in “weak” classes, do not have a good grasp of the language. This was found during the pilot study (Chong, 2007) and so a section on the degree to which the student understood the question was added to the test instrument in the main experiment. Students were asked to indicate whether they understood the question very well, fairly well, not so well, or that they did not understand the question. It was found that, on average, only 68% of the experimental group and 63% of control group understood each question. In “good” classes, the proportion was over 70%, whereas in “weak” class, just over half of the students indicated they understood the question. This clearly shows that DAQ would not be very effective in “weak” classes in Brunei schools, where proficiency in the English language is low.

5. Teacher’s Ability to Teach the Subject

After the post-test, the researcher discussed the answers to the DAQ with the participating teachers. Some teachers also tried answering the 26 DAQ. It was found that several teachers misunderstood some of the basic electricity concepts being tested. One teacher, for example, was convinced that a resistor behaves like a gate that
restricts the amount of current flowing through it but not affecting the current flowing into it. This has serious implications on the use of DAQ in a formative assessment system, as teachers with misconceptions would have more opportunity to teach the wrong concepts to the student. In such a case, the misconceptions would be reinforced in both teacher and student, while giving both the sense that the teaching and learning process has been enhanced with the use of a diagnostic tool and feedbacks that reinforce the process. This is a serious issue that must be addressed because it is very likely that there are other teachers with inadequate knowledge teaching in the various schools in the country.

The few teachers, who admitted that they were not able to answer some of the DAQ, believed some of the concepts being tested are of a level that is too “high” and “deep” for lower secondary students. They revealed that other teachers who did not participate in the study concurred when consulted. One teacher said that she cannot answer a circuit diagram question “because no numbers were given”, revealing her own ‘low’ and ‘shallow’ level of understanding. Indeed, most of the teachers expressed some difficulty in choosing the correct answers for many of the DAQ and said that DAQ may not be suitable for lower secondary school students. These are however young teachers with only between 1 to 5 years of work experience. The only experienced (> 20 years) teacher who participated in the study was able to use DAQ in his lessons and knew the correct answer the test questions. This highlighted that fact that the ability to teach using DAQ is directly related to the teacher’s depth of knowledge of the subject matter. This in turn determines the level of confidence and proficiency with which he or she is able to use DAQ as a diagnostic tool to correct misconceptions in a process of teaching and feedback. Without a sound understanding of the concepts being tested, misconceptions could be transmitted to students and reinforced in both the minds of the teacher and student.

6. Teacher’s Attitude towards DAQ

During the study, 2 of the 8 participating teachers were not very co-operative even though they had voluntarily agreed to participate in the study. They had expressed interest in the use of DAQ and the method proposed to test if it would have an effect in their classrooms. However, their attitude during the experiment showed that they did not think much of DAQ as a tool to enhance teaching or learning. They demonstrated a lack of responsibility as participants in the study and a lack of confidence and interest in the instrument. Consequently, the classes that they taught did not gain much from the use of DAQ. Their classes were not included in the data analysis on efficacy of DAQ and formative assessment.

7. Classroom Culture

Brunei classroom culture is characterised by an authoritative teacher imparting knowledge to generally well-behaved students that do not question the teacher. Consequently, attempts at interactive, dialogic styles of teaching have not been successful. The main classroom interaction pattern observed may be regarded as authoritative-interactive/non-interactive (Scott & Mortimer, 2006), which when coupled with a teacher with inadequate knowledge, is not suitable for DAQ-formative assessment style of teaching/learning.
In addition, students tend not to give honest answers. Instead, there is a tendency to respond with answers that they believed is expected of them. Culturally, it is also considered arrogant to express strong confidence or to flaunt talent. This was confirmed through interviews, where students admitted to knowing the correct answer but saying that are “not confident” or only “fairly confident”. This cultural trait would compromise the effectiveness of formative assessment strategies because students have the tendency to respond by giving answers that they think the teachers are looking for thereby making it difficult for the teacher to know the true situation.

Adapting DAQ and Formative Assessment For Brunei Schools

The majority of students in Brunei schools tend to rote-learn theories and formulae and lack the ability to relate theory and reality. Misconceptions are common because some teachers lack the ability to teach the subjects assigned properly, largely because of inadequate knowledge on the subject matter and hence, confidence to teach and correct misconceptions. DAQ and a formative assessment framework helps teachers to refine their understanding of the subject matter and build their confidence while they teach and correct misunderstandings in the classroom. In an ideal classroom, this iterative process of teaching and feedback, while using a diagnostic tool, would certainly enhance teaching and learning. However, as the paper has discussed, unless the problems identified are address and mitigated, DAQ and formative assessment would not only lose their functions, but they could lead to increased spreading and reinforcement of misconceptions. The following are four suggestions on how to make DAQ and formative assessment more effective in Brunei schools.

1. **DAQ in Smaller Steps**

To address the problem concerning too much information contained in each DAQ, perhaps DAQ could be broken down to test only one concept in separate questions in a set of 3-4 smaller DAQ, rather than 3-4 concepts at the same time. Although the EPSE question bank has DAQ of varying simplicity or complexity, most DAQ contain 2-4 parts not including the question on confidence in answering the question. It is suggested that DAQ should contain only one question, with the only additional questions to be regarding the student’s level of confidence and understanding. While it is important to have options to identify misconceptions, there should be 2-4 answer options instead of 5-7 to make the option choices clearer. If smaller DAQ presented in a less cluttered layout is used in a step-wise manner, beginning with test series containing 3-5 DAQ on simple basic concepts and terms before leading to sets that test more complex concepts, it is believed that teachers and students would find them more effective as tools to enhance the learning process. DAQ in smaller steps also addresses the problem of low mental stamina exhibited by many lower secondary students in Brunei, and the diagnostic uncertainty associated with complex DAQ that test multiple concepts.

2. **Using Language Appropriate for Lower Secondary School Students**

The study showed clearly that over 30% of the students involved had trouble understanding the DAQ. The proportion was nearly 50% for academically weaker classes. Clearly, while the words used were carefully chosen for their precision in
conveying the exact meaning of the question and answer options, there is a need to rephrase the DAQ with language that is more appropriate to lower secondary school students. For more effective application of DAQ in Brunei schools, the language must be simplified. As many children now learn better through multiple modes, i.e. when textural information is augmented with visual information, especially in animated format, DAQ may be more effectively employed using electronic media, e.g. as an on-line animated computer quiz complete with sounds and instant answers. However, this would require adaptation of existing DAQ for use with ICT (information-communication technologies) or creation of new DAQ made for ICT learning. This in turn would require a section of the Ministry of Education, perhaps the Curriculum Development Department, to take on this mammoth task, as it would certainly be unrealistic to expect teachers to develop the e-DAQ, given that few teachers can spare the time or have the resources and ability to do it. However, for starters, some modification to simplify the language used in existing DAQ and presenting them, for example, using Power Point may improve their effectiveness in Brunei schools.

3. Training Teachers to Use DAQ Effectively

Discussions with the project supervisor, colleagues and participating teachers indicate that teachers need to be given training on how to use DAQ effectively. Training programmes must also take into consideration that some teachers may resist having to use DAQ as part of their teaching because of a variety of reasons including a reluctance to disrupt or replace something familiar with something new and unfamiliar; and seeing the introduction of DAQ as additional burden to a curriculum that is already larger than the time and resources allocated to complete it. In addition, many teachers, particularly the young and inexperienced, need to be given coaching on questioning techniques and teaching abilities, as well as classroom management training, in order to use formative assessment strategies. Finally, the most critical part of teacher training must focus on strengthening teachers’ knowledge on the subject they teach. This could be done through workshops involving the use of DAQ so that they may discover for themselves their misconceptions and correct them with the help of the training facilitator, who must of course be an expert on the subject matter. Using DAQ with teachers would also help young inexperienced teachers that are not very sure about the concepts (and the associated terminologies) that they are assigned to teach gain a clearer understanding as DAQ are designed deliberately to zoom in on the exact meaning of terms and concepts. For practical purpose, DAQ training may be done in a staggered one afternoon per month, teaching-the-teacher system that is implemented in different geographical areas simultaneously to cover all schools.

4. Incorporate DAQ and Formative Assessment into Curriculum

As time and resource are two persistent constraints to any new developments in the classroom, DAQ and formative assessment strategies must be incorporated into the curriculum for it to be practical. The Curriculum Development Department therefore must be engaged to review the current situation and be convinced to incorporate DAQ and formative assessment into the existing curriculum. This would require consultation with experts, funding and commitment. Given the efficacy of DAQ as shown by many studies, including this study, this would be a challenge that is worth the time, energy and investment it demands.


This document was added to the Education-line database on 10 October 2008