AN INVESTIGATION OF PERFORMANCE-BASED ASSESSMENT IN SCIENCE IN SAUDI PRIMARY SCHOOLS

Paper presented at the 34th IAEA Annual Conference
CAMBRIDGE UK, SEPTEMBER 2008

Abstract

The purpose of this study was to investigate the effects of performance-based assessment on six grade students' achievement and attitudes toward science. The sample of the study consisted of 289 students. The study included twelve science classes comprising 289 primary school students were taught by six teachers in the city of Riyadh formed the population for the study. Six classes were randomly selected and were instructed using a performance-based assessment approach. A second cohort of six classes was instructed traditionally as control groups. The result showed that students in the experimental group had significantly higher scores in the science post-test than the students in the control groups. There was also a significant attitudinal difference towards science between the experimental and control groups in favour of the experimental group. The performance-based assessment procedures were found capable of predicting approximately 23 per cent of variation in the students' final science test scores.

Introduction

Contemporary development in cognitive and constructivist theory changes the researcher's perspective regarding acquisition of knowledge and competencies. In this perspective, meaningful learning is reflective, constructive, and self-regulated; learners are seen not as mere receivers of information but as creators of their own unique knowledge structures (2002; Liu, 2000). In this way, learners can achieve a more meaningful goal in which acquisition of knowledge, skills, and attitudes enable them to act effectively, expertly, and professionally under a teacher's guiding role (Freedman & Lee, 1998). Assessment theory and practice are evolving to reflect these complexities, moving away from a narrow focus on simple tests and scoring that previously dominated teaching (Cumming & Maxwell, 1999; Howell, Brocato, Patterson, & Bridges, 1999). In the information age, Liu (2000) asserts, students do not need to acquire a vast amount of information, typically the focus of traditional tests, but rather the ability to think, and organise that information for specific purposes. Thus, achievement needs to be considered as a qualitative change in a person's conceptions, not simply the amount of knowledge that a person possesses. It is no longer enough to count the number of correct answers on a test.

For two decades researchers have criticised traditional tests such as multiple-choice, fill-in-the-blank, and true or false questions (Akerson, Morrison, & McDuffie, 2002; Joan; Herman, Aschbacher, & Winters, 1992; Higuchi, 1993; Pate, Homestead, & McGinnis, 1993). Traditional techniques do not prove effective for the expanded concept of learning that requires students to demonstrate higher-level thinking skills (Sweeny, 1996). Linn, Baker, and Dunbar (1991) emphasise that traditional tests evaluate a limited number of cognitive functions and skills related only to memory, and students' ability to recall material learned out of context. Moreover, for purposes of accountability, teachers tend to tailor their instructions to students in imitation of multiple-choice questions (Baker, 1996; Kane, Khattri, Reeve, & Adamson, 1997), thus encouraging students to focus only on the options before them (Joan Herman, 1997). Therefore, "teaching to the test" has become a common practice in schools (Bowers, 1989; Izard, 2004; Resnick, 1996), narrowing students' potential to low-level skills, and distorting the curricula: for example, neglecting science, and social studies in favour of reading, and mathematics (Baker, 1996; L. A. Shepard et al., 1996). In addition, science educators claim that traditional tests cannot sufficiently evaluate students' ability to design and undertake experiments or assess their understanding of scientific concepts (Whitman, Klagholz, Schechter, Doolan, & Marganoff, 1998). In regard to this issue Hein and Price (1994) state:

> Assessing science through paper-and-pencil tests is akin to assessing a basketball player's skills by giving a written test. We may find out what someone knows about basketball, but we won't know how well that person plays the game (quoted in Whitman et al., 1998, p. 51).

To assess students on scientific reasoning and understanding rather than simply measuring discrete knowledge, critical assessment methods were developed, with a strong preference emerging for performance-based assessment (Morrison, McDuffie, & Akerson, 2003; Scott, 2002). Researchers such as Hakel (1998) and Howell, Brocato, Patterson, and Bridges (1999) identify performance-based assessment as the focus for education reforms in assessment, curriculum and instruction. The proponents for this type of assessment (Akerson et al., 2002; Guy & Wilcox, 2000; Perlman, 2003; Shavelson, Ruiz-Primo, & Wiley, 1999; L. A. Shepard et al., 1996; Solano-Flores, 1997; Wiggins, 1998) argue that a performance-based assessment methodology provides students with meaningful paths to demonstrate their knowledge. The technique also improves student skills by bringing into play complex functions of cognitive processing that require a higher level of thinking for problem-solving, or the development of options when an individual confronts a new situation.

Since performance-based assessment occurs over a period of time, it provides an opportunity for students to individually achieve the highest level of learning (Baker, 1996). Unlike the memory-based traditional testing procedures, performance-based assessment is *authentic* assessment, because it involves the performance of tasks that are valued in their own right, it is situated in a real world context, and it can mirror actual tasks implemented by professionals (Jorgensen, 1994; Linn et al., 1991; Mabry, 1999). These characteristics of performance-based assessment allow students to engage with meaningful problems that foster significant educational experiences (Garbus, 2000; Kulieke et al., 1990; Linn et al., 1991).

In the classroom, performance-based assessment has value for students and teachers. For students, performance assessment provides a realistic approach to science, reinforces the inquiry skills of science that facilitate the art of language and the use of mathematics, and assesses self-progress (Pico II, 1999). For teachers, the methodology provides timely information on the learning needs of their students, and thus the teaching methods they employ (Corcoran, Dershimer, & Tichenor, 2004).

Performance assessment is therefore an appropriate strategy for assessing students' concepts and skills in science, and it prepares students for a productive future within a technologically complex world. The methodology also fits the nature of science, that is, the study of active structures, and frequently changing natural phenomena (Ainley, Hidi, & Berndorff, 2002; Atkin, Black, & Coffey, 2001; Collins, 1997; Guy & Wilcox, 2000; Shavelson, 1994). In addition, Atkin, Black, and Coffey (2001) claim that the current goals for science in educational standards reform present a significant shift to performance assessment. This is due to the fact that standards reform presents science as a subject where students are actively involved in science rather than reactive reading or listening.

Empirical studies of the impact of performance-based assessment show positive effects in the quality of students' learning and attitudes. Baxter and Glaser (1996) found that performance-based assessment not only supports the development of thinking and reasoning in the classroom, but also provides teachers with feedback that can be used to improve the classroom environment. Similarly, Biondi (2001) found that performance-based assessment is a valid, equitable measurement of student progress. Through performance assessment strategies, students become more focused in their work, are able to reflect on their learning activities and abilities, and develop a higher level of vocabulary through group conferences and self-assessments. After studying the effective performance-based assessment for evaluating fifth- and sixth- grade student science achievements, Parker and Gerber (2002) conclude that performance-based assessment is effective in measuring the knowledge and skills of science students.

Many educationalists however propose that performance-based assessment should be considered not merely as a process for assessing students' understanding, but also as a learning process; one that teaches students concepts and requires them to explain and communicate their interpretations of the information, and their methodology for solving problems (Liu, 2000 ; Morrison, McDuffie, & Akerson, 2002). Rudner and Boston (1994) suggest that with performance assessment methodologies, instructional objectives in science should be redefined to include more practical applications and more emphasis on synthesis and integration of content and skills. Therefore, a considerable change in instructional procedures as well as in science curricula must take place to align with theoretical conceptions that underline the new assessment method. With performance assessment, teacher-centered instruction practices must give way to more student-centered techniques that allow students to become engaged actively in the skills they need for their roles (Pfeifer, 2002). In this situation, performance-based assessment can change classroom learning structures in which students merely listen and absorb information to those in which students actively participate (Gopinath, 1999), working together or separately (Haury, 1993). Further, students in this learning experience can assess their own progress and therefore be more responsible for their own learning (Andrade, 2000).

The study was guided by the following research questions

- Q1. What are the differences between the type of science learning outcomes that can be achieved by the implementation of performance-based assessment and traditional testing methods?
- Q2. Are student attitudes toward science affected by performance based assessment?
- Q3. Are students' final science examination outcomes predictable through performance based assessment?

Methodology of research

For this study, six primary boys' schools were randomly selected from schools throughout Riyadh City. The Grade 6 classes were randomly assigned to either a control group (N=143) or an experimental group (N=146), a total of 289 students. Due to absences of students from one or other of the pre-test or post test, data for statistical comparison of only 274, 249 students were compiled on the pre-tests, and 265, 225 on the post-tests of science and SATSS respectively. The minimum and maximum ages of participating students were 10 years and 15 years respectively, with an average age of 12.18 years for the experimental group and 12.10 years for the control group. The study program was administered in the first term of the second semester over a period of nine weeks, and delivered to Grade 6 primary students at their thrice-weekly science classes. Each lesson comprises a forty-five minute block of time during the school day which is from 7am-12:30 pm.

Program Content Methodology

The study program was developed to embed performance-based assessment into learning and instruction whilst using strategies such as reasoning, communication, problem solving skills, and the conceptual understanding of science that reflect the new focus on higher order thinking. This approach was aimed at a change in teaching from rote-based learning to a greater dependence on students seeking out knowledge, learning through practising that knowledge, and acquiring skills in which they receive feedback from the teacher and for which they are assessed. The initial information package for teachers was supported by a number of empirical studies on the use of performance-based assessment in science classes to illustrate issues that arise and how they may be addressed. Further, the package, described teaching styles that encourage students to be reflective and self- regulated, with a focus on problem-solving and inquiry. Exploration was made of the program strategy which involves four elements including technique, assessment, performance, and time. A comprehensive explanation of each element (see Figure 1) and usage guide developed for primary school science classes.

Figure 1 Performance-Based Assessment Program Strategy

The intervention study was restricted to the Electricity and Magnets units in the science classes. The science content and objectives of the Units of Work in the study program for the experimental groups were aligned to the science curriculum of Grade 6 and were conducted as part of the regular timetable. Thus, variables relating to the school environment and the planned outcome for these science units under primary school conditions in Saudi Arabia were resolved. Further, the teachers reliably followed the constant and aligned teaching procedures for the six experimental science classes.

The teaching procedures used cooperative learning strategies for the experimental classes. At the beginning of each lesson, the teachers prepared materials for the activity and organised students into groups. The students read the instructions for the activity and then worked in pairs or in groups. At the end of each activity, students completed the self-assessment form. As well as class activities, students were given a choice to select the topic of projects for the unit, and were issued with information, a rubric and instructions for that project. The instructions involved discussing the outline of the project with the teacher, following the objectives and timelines set out, then forwarding a draft for feedback before submitting the final project; finally delivering a presentation to a group or the class.

Instrumentation

Science Tests

A science course pre-test was conducted at the end of semester 1 to examine similarity between groups. The test included 20 items related to a science student subject in Grade 6, taught prior to this study's program in semester 1. These items were prepared by the National Achievement Tests Committee to study the achievement progress of primary science in Grade 6 and were designed for students to complete within 35 minutes. The science course post-test was a different measure from the pre-test measure because it had to include items that reflected the content taught during the intervention period. The content was limited to the Electricity and Magnets units that were covered in this research. The test consisted of 24 items adapted and developed by this researcher. Both the pre-science test and post-science test were conducted on different occasions on primary school classes in Grade 6. The data was analysed to determine reliability by evaluating the Cronbach alpha coefficient. For the pre-science test, the Cronbach alpha coefficient was .72, (N=18), and for the post-science test was .71, (N=16). Nunnaly (1978) has indicated 0.7 to be an acceptable reliability coefficient, so that Cronbach's reliability coefficients for the science tests are acceptable.

Student Attitude toward Science Survey (SATSS)

To determine the attitudes of students toward science, a survey of 18 items was administered as a pre-test and post-test, to 12 classes of Grade 6 students, representing both the experimental and control groups. Each of these items based on the Likert scale, had four possible responses from which the student chose by circling. The responses were Strongly Agree, Agree, Disagree and Strongly Disagree. There were a number of negative items for which the scoring was reversed, randomly distributed throughout the survey.

The survey was developed from both TIMSS (1999), and Century (2002), and translated into Arabic and then trialled on two classes of Grade 6 students (N = 43). The internal consistency of the survey was established using the Cronbach alpha coefficient and alpha, if item deleted (Cronbach, 1951). The alpha coefficient, if item deleted, showed that removing any single item will not yield a higher alpha; therefore the items are sufficiently related to combine into an index. The overall alpha for the survey was .84.

Results of Research

Statistical analyses were conducted to ensure that the required assumptions for the independent sample t test and analysis of variance procedures have been met. The assumptions are (1) independency, (2) normality, (3) homogeneity. Independency is generally determined by the structure of the experiment from which they arise. Since the experimental and control groups in this study were assigned at random, all observations obtained from different participants could be assumed to be independent. The second assumption is that the variance of observations is equal across each group. To test the homogeneity, Levene's test was conducted for post-tests. For the science post-test, the result of Levene's test of difference was not significant, $F(1, 263) = .036$, $p = .849 > .05$ (see Table 6-5). The result of Levene's test was also not significant for SATSS, $F(1,223) = 1.96$, $p = .163 > .05$. Thus, the science post-test and SATSS support the assumption of homogeneity of variance; which is, the variance is equal across groups, and therefore, the required assumptions of homogeneity of variance for the ANOVA procedures were fully met for the dependent variables. The third assumption is that the distribution of observations for the dependent variable is normal within each group. The normality assumption was tested for each group by dividing skewness and kurtosis by their standard errors. A result between +2 and -2 is considered to constitute a normal distribution. The normality of the post tests was evaluated for the science test and the SATSS. For the science post-test, the ratio of skewness to standard error for the experimental and control groups was – 1.35, and - 2.08 respectively, and the ratio of kurtosis to standard error for the same groups was -1.34 and .70. For SATSS post-test**,** the results of ratios of skewness and kurtosis for the experimental group were -1.38 and 1.31 respectively, and for the control group were -1.35 and .43. Therefore, the required assumptions of independency, homogeneity and normality for the ANOVA procedures met the dependent variables (science achievement, and student attitudes) in post-tests occasions.

Research question 1

The first research question concerned the effects of the study's performance-based assessment program on students' science achievement levels compared to those reached by traditional assessment methods. The question is: *What are the differences between the type of science learning outcomes that can be achieved by the implementation of performance-based assessment and traditional testing methods?*

H01: There is no significant difference between the experimental and control groups in the science post-test.

To test the null hypothesis that emanates from Question 1, the independent sample t test was applied. The result revealed significant statistical differences in the science post-test scores between the experimental (N= 136, *M* = 16.69, *SD* = 3.49), and control (N= 129, *M* = 15.37, *SD* = 3.55) groups, $t(263) = 3.05$, $p = .003 <.01$ (see Table 1). Therefore, the null hypothesis is rejected.

Table 1 Independent Samples t Test for Science Post-test

Group	***Mean***	***SD***	***df***	***t value***	***P***
Experimental group	*16.69*	*3.49*	*263*	*3.05*	*.003*
Control group	*15.37*	*3.55*			

Research question 2

The second research question investigated the effect of the study project on the students' attitudes: *Does performance-based assessment have an effect on students' attitudes' toward science?*

H02: There is no significant difference in the means of scores between the experimental and control groups for the students' attitudes toward science.

Data for this variable was collected through the use of the Student Attitude toward Science Survey (SATSS) which was administrated at the beginning and the end of the study. The data collected from both the pre and the post administration of the SATSS for both the experimental and control groups (see Table 2). The data were analysed by one-way analysis of variance. The result as can be seen in Table 3 showed significant statistical differences in the post-test of the SATSS between the experimental (*M* = 51.17, *SD* = 6.12) and control (*M* = 49, *SD* = 7.10) groups, $F(1,223) = 6.08$, $p = .014 < .05$. Therefore, the null hypothesis that there are no statistically significant differences between the attitudes of the experimental and control groups toward science is rejected.

Table 2 Means and Standard Deviation for groups on SATSS

SATSS	**Group**	**N**	**Mean**	**Std. Deviation**
Pre-test	Experimental	118	50.39	6.17
	Control	131	50.11	6.97
Post-test	Experimental	111	51.17	6.12
	Control	114	49	7.10

Table 3 Analysis of Variance (ANOVA) for Science pre and post-tests

Measure	**Source**	**df**	**F**	**p**
Pre-test				
	Between	1	.108	.743
	Within	247		
Post-test				
	Between	1	6.08	.014
	Within	223		

Research question 3

The third question investigated whether performance-based assessment results are predictors of student performances in the final science test. Specifically, the question asked: *Can performance-based assessments predict students' scores in the final science exam?*

H03: Performance based assessment cannot predict final science examination scores of the experimental group.

Data were collected through applying the study program's performance assessment in science classes, comprising six science activities, and two tasks for each unit. The students' achievements were scored for three activities, as well as both tasks for each unit, and were randomly collected from 10 students in each class. The final performance-based assessment scores were generated by averaging the individual scores for the ten activities and tasks. The capacity of performance assessment to predict students' scores in the final science exam was investigated by using linear regression analysis on the data from students' results on the performance assessment, and then from their final examination results. The result in Table 4 demonstrated that approximately 23 per cent of the variation in the final science test scores can be predicted by the performance-based assessment scores at < .05. The results of the linear regression analysis are summarised in Table 5. The relationship between performance–based assessment scores, and final science test scores shown in Figure 2 is a moderately positive linear relationship, r = .484.

Table 4 *Linear Regression Coefficients for Performance-Based Assessment predicting final science test*

Variable	Coefficient	Std. Coeff.	t	p
(Constant)	7.38	.484	2.723	.009
Performance assessment	3.11		3.984	.000
N = 54	R = .484 R^2 = .234		Std. Error.= 2.96	

Table 5 *Analysis of Variance*

Source	SS	df	MS	F	p
Regression	138.94	1	138.94	15.876	.000
Residual	455.06	52	8.75		

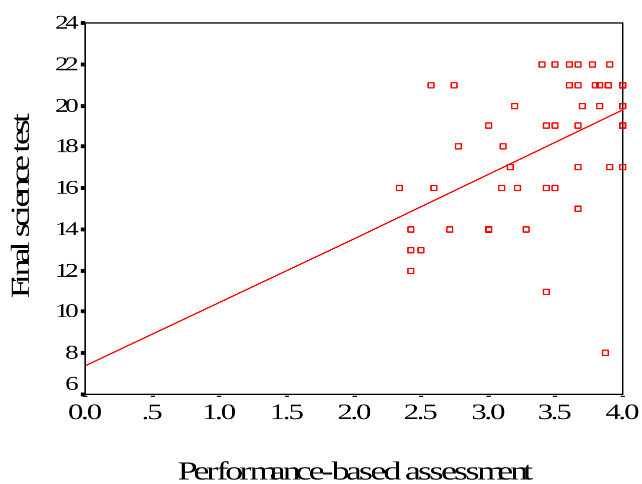

*Figure 2*Linear Relationship between Performance Assessment and Final Test Scores

Discussion

A priority for this study was to determine whether the use of a performance-based assessment approach showed a marked result for grade 6 science students' achievement scores. For this purpose, independent sample t test was applied to examine the difference between groups. The result indicates that the mean differences between the two groups were significant at < .05 level in favour of the experimental group. So, it suggests that a performance-based assessment approach promotes and supports learning science at the classroom level. This result is consistent with a considerable body of research (e.g., Biondi, 2001; Enger, 1997; Parker & Gerber, 2002) that has found that performance-based assessment supports learning science.

Students in the experimental group were engaged in science education with consideration for the diversity of their needs, attitudes, and abilities, or as Gray and Sharp (2001) suggest that "the more children can identify and are engaged with a task, the more effort they will put into it and, therefore, the greater their success rate is likely to be" (p. 87). Another possibility for student outcomes is that teachers in this study used performance assessment formatively. They provided students with information in different ways such as making comments on student's work or within conducting an experiment, and students used this information as feedback to improve their learning.

Finding that performance-based assessment supports low achievers obtained in this study project parallel and confirm the empirical findings of Gray and Sharp (2001). Gray and Sharp also studied grade 6 science students in an assessment context, comparing the results of traditional and performance-based assessment and found students, particularly lower achievers, perform better on interactive performance assessment than on comparable tasks presented in a purely pencil and paper format.

However, the positive results of this study reflected in students' outcomes in science do not focus on performance assessment as a test method, that is, students merely practising science for later examination. Instead these results encompass a holistic approach to learning, based on the theoretical framework of performance assessment, which is different from the traditional approach of assessing and teaching science. The approach contains interactive factors in learning, teaching and assessing, based on constructivist learning theories, as has been described in the theoretical framework and formulated in the study project. In this approach, performance-based assessment required the learners to adopt higher order thinking, engage in inquiry and problem-solving for a period of time, and the vitalisation of teaching methods that encourage active participation. These activities occur within a social constructivist learning environment where students are able to work cooperatively, and reflect on their work (Roth, 1995; Lorrie Shepard, 2000).

However, using performance assessment solely as a separate process from instruction, or within a learning environment based on behaviourist learning principles, focusing on memorisation or recall cannot be an effective assessment method, whether for classroom or accountability use. Accordingly, researchers who used performance-based assessment separately reported no positive effects for performance assessment on students learning science. Shymansky and Chidsey (1997) gave students a one-hour block of time to complete the individual performance tasks to parallel the on-demand nature of the Iowa Test of Educational Development (ITED). They found that students performed poorly on all performance tasks. A similar result, determined using similar performance assessment processes, was found by Huff (1998). He studied the effects of the use of multiple-choice item formats and performance formats for the assessment of learning science at the second grade level (n =16) over two weeks. Both forms of assessment were used after science classes. The results showed no positive effects for performance assessment. However, as he suggested, some disadvantages of performance assessment may be related to the fact that students were not familiar with performance tests. Another possibility is that performance assessments have essential elements which require drastic changes in teaching styles, curriculum materials, classroom environment, and learning methods.

Another example of research that separated performance-based assessment from instructional procedures was conducted by Century (2002) who also compared the impact of alternative and traditional tests with sixth grade students. Whilst I used in this study different teaching methodologies under traditional assessment and performance-based assessment, Century utilised the same teaching methods for both the control and the applied groups, but they were assessed differently by either performance-based assessment, or a traditional test form. Century's study showed that there was no significant difference between students' performance on the two types of assessment. The lack of clear procedures with Century's work is therefore considered in light of the teaching and learning techniques employed in his study, which were similar in both groups, thus hindering the provision of the basic conditions necessary for applying performance assessment.

Nevertheless, research studies that combined performance-based assessment to instructional procedures found encouraging results. For example, Biondi (2001) who combined performance assessment with instructional procedures found that performance-based assessment is a valid, equitable measurement of student progress. Students became more focused in their work, were able to reflect on their learning activities and abilities, and developed a higher level of vocabulary through group conferences and self-assessments. In addition, he found that performance-based assessment provided students with tangible evidence of their work as they analysed their strengths and weaknesses, became more focused on their assignments and were able to apply their knowledge of the material in a creative manner.

The procedures for implementation of performance assessment as shown in the previous studies have essential impact on the expectation outcomes, so addressing these procedures and considering their recommendations assisted the researcher to achieve positive results for the performance assessment approach. For instance, after an expected result, Shymansky et al.(1997) highlight two important issues for producing a valid performance assessment in science classroom. These are the development of teaching practices and alignment of goals, teaching, learning and assessment processes.

The results of data analyses also showed that student attitudes toward science have increased significantly for the experimental group, compared to the control group, as measured by the Student Attitude toward Science Survey (SATSS). Thus, the findings of this study support the use of performance-based assessment to enhance students' attitudes toward science. This result is inconsistent with the findings of Century (2002), who investigated the impact of alternative and traditional assessment on students' attitudes, and science learning outcomes. Century found no significant difference between alternative and traditional groups in their attitudes toward science, and in their science outcomes. This inconsistency may be related to the issue of separating performance assessment from teaching and learning procedures. This separation prevents students from being actively engaged in science class, using both science processes and critical thinking skills as they search for answers (Gibson & Chase, 2002). On the other hand the current result is consistent with the findings of studies that had an influence on students' engagement in science. Using a sample of 699 students from 27 high schools science classes, Myers and Fouts (1992) found that the most positive attitudes were related to a high level of involvement, very high level of personal support, strong positive relationships with classmates, and the use of a diversity of teaching strategies and innovation learning activities. In a similar study, Siegel and Ranney (2003) used activity-based science curriculum focused on connecting science to students' lives. The curriculum concerned scientific evidence to make decisions involving social consequences and their findings were that students' attitudes toward science were enhanced. Similarly, Gibson and Chase

(2002) conducted an inquiry-based science program, as a form of constructivism similar to performance-based assessment, to stimulate greater interest in science, and scientific careers within middle-school students. They found that students maintained a more positive attitude towards science and a high interest in science careers. Bilgin (2006) conducted intervention on grade 8 students that included hands-on science activities using a cooperative learning approach. The researcher found that the experimental group had better performance on the attitude scale toward science and on the science-post test.

This consistency between the results of these previous studies and the result of the current study can be attributed to common factors related to classroom variables such as using enquiry, cooperative learning, and linking activity-based science to real life situations.

The results for the third research question which is *Are students' final science examination outcomes predictable through performance-based assessment?* Showed that approximately 23 per cent of the variation in the final science test scores can be predicted by the performance-based assessment scores. This means that the forms of performance assessment such as projects, portfolios and tasks can be used as an accurate indicator of students' progress, particularly in the new Saudi *Continuous Assessment* educational system bylaw, which depends on continuous assessment for promoting students.

However, the method of demonstrating tasks that were designed to assess a student's progress at the end of each unit undermined to some extent the direct correlation between the forms of performance assessment and the final science examination. This is because these tasks should be demonstrated individually to give a valid indicator of students' progress, there was a shortage of experiment instruments, and students performed tasks in groups comparable to other forms of performance assessment. This finding is consistent with the results of Gallant (2005), who concluded that "a curriculum-embedded performance assessment can be used to predict students' performance on a state's criterion-referenced assessment in a later grade" (p. 106).

In this study, although performance-based assessment methodologies were used formatively, and within a cooperative environment, the regression analysis results demonstrate a positive association with the final science test outcomes. A finding of this research is therefore that utilising tools such as tasks, projects, and portfolios for summative purposes has a greater possibility of positive outcomes than the use of traditional tests such as multiple-choice and true or false, particularly with primary school pupils. As performance assessment improves students' learning and attitudes toward science, it could also be considered as valuable for summative assessment.

Traditional testing procedures however can be useful. As proponents of performance assessment, for example, Herman, Aschbacher, and Winters (1992) suggest, using a variety of modes of assessment. A balanced assessment system involving different types of assessment is needed to give a detailed, multi-perspective picture of student accomplishments, that may best serve all functions, knowledge domains, and learners (Haury, 1993; Tillema, 2003). Nevertheless, Haertel (1999) believes that performance assessment should not be used as external assessment, but just for daily classroom instruction. There is however, no external assessment for accountability in the Saudi education system. The scope of this study is to examine the viability of using performance assessment for promoting students to the next class within the new assessment. Thus research finding 3 is that the Saudi *Continuous Assessment* educational bylaw can use forms of performance assessment such as projects, portfolios, and tasks as trusted indicators of students' achievements.

References

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*(3), 545-561.

Akerson, V., Morrison, J., & McDuffie, A. (2002). Performance assessment in science methods: a program description. Retrieved December 20, 2004, from http://www.ed.psu.edu/CI/Journals/2002aets/f7_akerson_morrison_m.rtf

Andrade, H. G. (2000). Using rubric to promote thinking and learning. *Educational Leadership, 57*(5).

Atkin, J. M., Black, P., & Coffey, J. (2001). *Classroom assessment and the national science education standards*. Washington, DC: National Academies Press.

Baker, E. (1996). What the research says about student assessment, *Improving America's Schools: A Newsletter on Issues in School Reform*: the U.S. Department of Education.

Baxter, G., & Glaser, R. (1996). *Assessment and instruction in the science classroom* (No.). Los Angeles: Centre for Research on Evaluation Standards and Student Testing (CRESST).

Bilgin, I. (2006). The effects of hands-on activities incorporating a cooperative learning approach on eight grade students' science process skills and attitudes towards science *Journal of Baltic Science Education*(9), 27-37.

Biondi, L. (2001). Authentic assessment strategies in fourth grade. *(ERIC Document Reproduction Service No. ED460165)*

BouJaoude, S. (2003). Science and technology education in the Arab world in the 21st century. *Connect, XXVIII*(3-4).

Bowers, B. (1989). Alternatives to standardised educational assessment. *(ERIC Document Reproduction Service No. ED312773)*

Century, D. (2002). *Alternative and traditional assessments: their comparative impact on students' attitudes and science learning outcomes: an exploratory study* Unpublished the Degree Doctor of Education, The Temple University.

Chang, S.-N., & Chiu, M.-H. (2005). The development of authentic assessments to investigate ninth graders' scientific literacy: in the case of scientific cognition concerning the concepts of Chemistry and Physics. *International Journal of Science & Math Education, 3*(1), 117-140.

Collins, A. (1997). National science education standards: looking backward and forward. *The Elementary School Journal, 97*(4), pp.299-313.

Corcoran, C. A., Dershimer, E. L., & Tichenor, M. S. (2004). A teacher's guide to alternative assessment taking the first steps. *Clearing House, 77*(5), 213-216.

Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policy & Practice, 6*(2), 177.

Elshout-Mohr, M., Oostdam, R., & Overmaat, M. (2002). Student assessment within the context of constructivist educational settings *Studies in Educational Evaluation, 28*, 369-390.

Enger, S. K. (1997). Linking instruction and assessment in science: science learning opportunities and student performance on a set of open-ended science questions. *(ERIC Document Reproduction Service No. ED415271)*

Freedman, R., & Lee, H. (1998). Constructivist assessment practices. *(ERIC Document Reproduction Service No. ED415118)*.

Gallant, D. J. (2005). *Predictive nature of a curriculum-embedded performance assessment for young children.* Unpublished Doctoral Dissertation University of South Carolina, Columbia.

Garbus, P. (2000). More than MCAS: performance-based assessment in the days of standardised tests, for the Massachusetts Charter School Fellowship Program.

Gibson, H. L., & Chase, C. (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. *Science Education, 86*(5), 693-705.

Gopinath, C. (1999). Alternative to instructor assessment of class participation. *Journal of Education for Business, 75*(1), 10-14.

Gott, R., & Duggan, S. (2002). Problems with the assessment of performance in practical science: which way now? *Cambridge Journal of Education, 32*(2).

Graham, P. (2005). Classroom-based assessment: changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education, 21*(6), 607-621.

Graue, E. M. (1993). Integrating theory and practice through instructional assessment. *Educational Assessment, 1*(4), 283.

Gray, D., & Sharp, B. (2001). Mode of Assessment and its Effect on Children's Performance in Science. *Evaluation and Research in Education, 15*(2), 55-68.

Guy, D., & Wilcox, J. (2000). Assessing elementary science teaching in a performance setting. Retrieved November 14, 2004 from www.ed.psu.edu/CI/Journals/2000AETS/18guy_wilcox.rtf

Haertel, E. (1999). Performance assessment and education reform. *Phi Delta Kappan, 80*, 662-666.

Hakel, M. D. (1998). *Beyond multiple choice: evaluating alternatives to traditional testing for selection* Mahwah: Lawrence Erlbaum Associates.

Haury, D. (1993). Assessing student performance in science. *(ERIC Document Reproduction Service No. ED359068)*.

Herman, J. (1997). Assessing new assessments: How do they measure up? *Theory Into Practice, 36*(4), 196-201.

Herman, J., Aschbacher, P., & Winters, L. (1992). A practical guide to alternative assessment. *(ERIC Document Reproduction Service No.ED352389)*

Higuchi, C. (1993). *Performance-based assessments and what teachers need* (CSE Technical Report No. 362). Los Angeles: Centre for Research on Evaluation, Standards, and Student Testing (CRESST).

Howell, L., Brocato, D., Patterson, K., & Bridges, T. (1999). An evaluation of performance task instruction in the classroom. *(ERIC Document Reproduction Service No. 435761)*

Huff, J. (1998). Assessment in science: an examination of multiple-choice and performance science assessment for a second grade classroom. *(ERIC Document Reproduction Service No. ED423286)*

IEA. (2003). TIMSS 2003 results. from http://nces.ed.gov/timss/TIMSS03Tables.asp?Quest=3&Figure=6

Izard, J. (2004). *Gathering evidence for learning.* Paper presented at the Annule Conference of the Australian Association for Research in Education (AARE) Melbourne.

Jorgensen, M. (1994). *Assessing habits of mind: performance-based assessment in science and mathematics*. Ohio.

Kane, m., Khattri, N., Reeve, A., & Adamson, R. (1997). Assessment of student performance: studies of education reform: Office of Educational Research and Improvement, U.S. Department of Education.

Kulieke, M., Bakker, J., Collins, C., Fennimore, T., Fine, C., Herman, J., et al. (1990). Why should assessment be based on a vision of learning? Retrieved November 8, 2004, from www.ncrel.org/sdrs/areas/rpl_esys/assess.htm.

Linn, R., Baker, E. L., & Dunbar, S. B. (1991). *Complex, performance-based assessment: expectations and validation criteria* (CSE Technical Report No. 331). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Liu, J. (2000). *The effect of performance-based assessment on eighth grade students Mathematics achievement.* Unpublished Doctor of Philosophy, University of Missouri-Columbia.

Mabry, L. (1999). *Portfolios plus: a critical guide to alternative assessment*: Crowin Press, Inc., California.

Morrison, J., McDuffie, A., & Akerson, V. (2002). A Focus for collaboration: developing and implementing Science and Mathematics performance assessment tasks. *(ERIC Document Reproduction Service No. ED465611)*

Morrison, J., McDuffie, A., & Akerson, V. (2003). Preservice teachers' development and implementation of science performance assessment tasks. *(ERIC Document Reproduction Service No. ED 478065)*

Myers, R. E., Fouts, J. T., & (1992). A cluster analysis of high school science classroom environments and attitude toward science. *Journal of Research in Science Teaching, 29*(9), 929-937.

Parker, V. A., & Gerber, B. L. (2002). Performance-based assessment, science festival exhibit presentations, and elementary science achievement. *Journal of Elementary Science Education, 14*(1), 59-67.

Pate, P., Homestead, E., & McGinnis, K. (1993). Designing rubrics for authentic assessment. *Middle School Journal, 25*(2), 25-27.

Perlman, C. C. (2003). Performance Assessment: Designing Appropriate Performance Tasks and Scoring Rubrics. *(ERIC Document Reproduction Service No. ED480070)*

Pfeifer, G. R. (2002). *The influence of authentic assessment tasks and authentic iInstruction on Lutheran elementary school fifth and sixth grade students' attitudes toward social studies and authentic projects.* Unpublished doctoral dissertation, the University of Minnesota

Pico II, R. A. (1999). *A case study of three Puerto Rico-statewide systemic initiative intermediate science teachers' views on and use of performance assessment.* Unpublished doctoral dissertation, The Pennsylvania State University, Malvern

Resnick, L. B. (1996). *Performance puzzles: issues in measuring capabilities and certifying accomplishments*. Los Angeles: CRESST/University of Pittsburgh, LRDC.

Roth, W.-M. (1995). *Authentic school science*. Netherlands: Kluwer Academic Publishers.

Rudner;, L. M., & Boston, C. (1994). Performance Assessment. *The ERIC Review, ED 369 389*.

Scott, S. J. (2002). Using multiple methods of assessment to examine student achievement in Music from http://www.brandonu.ca/Music/People/FacultyPages/scotts/using_multiple_methods_of_assess.htm

Shavelson, R. J. (1994). Performance assessment. *International Journal of Educational Research, 21*(3), 233-350.

Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (1999). Note on sources of sampling variability in science performance assessments. *Journal of Educational Measurement, 36*(1), 61-71.

Shepard, L. (1995). Using assessment to improve learning. *Educational Leadership, 52*(5), 38-43.

Shepard, L. (2000). *The Role of Classroom Assessment in Teaching and Learning* (CSE Technical Report No. 517). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1996). Effects of Introducing Classroom Performance Assessments on Student Learning. *Educational Measurement: Issues and Practice, 15*(3), 7-18.

Shymansky;, J. A., Chidsey, J. L., Henriques, L. E., Yore, L. D. S., Wolfe, E. W., & Jorgensen, M. (1997). Performance assessment in science as a tool to enhance the picture of student learning. *School Science and Mathematics, 97*(4), 172-183.

Siegel, M. A., & Ranney, M. A. (2003). Developing the changes in attitude about the relevance of science (CARS) questionnaire and assessing two high school science classes. *Journal of Research in Science Teaching, 40*(8), 757-775.

Solano-Flores, G. S., Richard J. (1997). Development of performance assessments in science: conceptual, practical, and logistical lssues. *Educational Measurement: Issues and Practice, 16*(3), 16-25.

Sweeny, B. (1996). The transition to a performance-based classroom [Electronic Version]. *Best Practice Resources* from http://www.teachermentors.com/.

Tillema, H. H. (2003). Integrating Developmental Assessment with Student-directed Instruction: a case in vocational education in the Netherlands. *Journal of Vocational Education & Training, 55*(1), 113-128.

TIMSS. (1999). Science teacher questionnaire main survey. TIMSS International Study Center, Boston College: Chestnut Hill, MA 02467, USA.

Whitman, C., Klagholz, L., Schechter, E., Doolan, J., & Marganoff, B. (1998). New Jersy science curriculum framework. from http://www.state.nj.us/njded/frameworks/science/

Wiggins, G. (1998). *Educative assessment.designing assessment to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.