TÜRK FEN EĞİTİMİ DERGİSİ
Yıl 14, Sayı 1, Mart 2017

Journal of
TURKISH SCIENCE EDUCATION
Volume 14, Issue 1, March 2017

http://www.tused.org

Effectiveness of the Proposed Training Formative Assessment Programme and its Impact on Teaching Style Improvements of Saudi Science Teachers in Saudi Arabia

Fayadh Hamed ALANAZI[1]

[1]Assistant Professor, Aljouf University, Saudi Arabia

Received: 09.06.2016 **Revised:** 09.11.2016 **Accepted:** 01.12.2016

The original language of article is English (v.14, n.1, March 2017, pp.35-56, doi: 10.12973/tused.10189a)

ABSTRACT

This work investigates the effects associated with training initiatives consistent with formative assessment to enhance teaching among science teachers in Saudi primary schools. The programme comprised training sessions that were ongoing for a period of two months, which emphasised the strategic skills of adopting the formative assessment. The researcher adopted an experimental approach centred on the design of the control and experimental groups to establish the overall efficiency of the suggested training programme in the development of knowledge among teachers in the field of formative assessment strategy skills. The sample comprised 33 primary school science teachers who were based in Sakaka in the Al-Jouf region; the sample was randomly divided into an experimental group (16 teachers) and a control group (17 teachers). The findings indicate the success of the proposed program, in which the experimental group was more a function of the control group in the telemetric measurement. Moreover, classroom observations were applied by the researcher, in which the science teachers were assessed with the aim to determine whether a form of formative assessment was used following the completion of the programme. The findings suggest that the science teachers adopted various strategies of formative assessment, including sharing learning goals, open questions, assessment-based evidence, and teacher–child classroom dialogue.

Keywords: Science teachers, summative assessment, formative assessment, training program, teaching.

INTRODUCTION

Without question, assessments play a fundamental role in the enhancement of student learning. As stated by the Task Group on Assessment and Testing Report (TGAT, 1988), “assessment is at the heart of the process of promoting children’s learning” (para. 3). In the specific field of education, there are two recognised forms of evaluations: assessment for learning and assessment of learning.

It is essential to differentiate between assessment for learning, which is recognised as a formative assessment, and the assessment of learning, which is recognised as a summative assessment, to facilitate teachers in their application in practice. As stated by Sadler (1989), the formative assessment is centred on how quality-related judgements that pertain to student

Corresponding author e-mail: fayad@ju.edu.sa

responses may be used to enhance and generally shape the competence of students through short-circuiting the inefficiency and randomness associated with trial-and-error learning. Moreover, the view has also been adopted that a summative assessment is focused on summing or otherwise providing a summary concerning the achievement status of a student, and it is focused on providing a report upon course conclusion, particularly in cases of certification (ibid). Nevertheless, summative and formative assessments do not always lack in compatibility, with the information garnered from summative assessments formatively used when learners apply it to guide their activities and efforts in other courses. Moreover, a formative assessment may be applicable for use in informing a summative assessment, with these assessments indicating whether the learner is ready to progress to further material. After the instructor has established that the learner has adequate knowledge and skills, a summative assessment may be implemented, which enables them to showcase their skills and knowledge.

Furthermore, through the Qualification and Curriculum Authority (2009) in the UK, it has been established that assessment for learning goes against learning assessment, which is recognised as comprising judgements concerning student performance consistent with national standards. It is common for teachers to make these statements following the completion of a work unit, year or key stage. Furthermore, student performances in regards to levels are also described through test results.

From the researcher's perspective, the key difference between the two approaches to assessment is related to the target or objective rather than the time. When the key emphasis is on encouraging the learning of the student, which subsequently forms an efficient aspect of the learning process through the adoption of many different instruments, such as sharing learning goals, self and peer assessments, and becoming involved in discussion, learning evaluation would occur. When the focus of an evaluation is centred on reporting the mid-term progress of a student, for example, this approach is considered an assessment of learning.

Literature Review

The History of Formative Assessment

The term 'formative' was first introduced in the arena of mainstream teaching in the UK when the following statement was made by the TGAT (1988) concerning what should be afforded priority in a National Curriculum: "the results [of assessment] should provide a basis for decisions about pupils' further learning needs: they should be formative" (Paragraph 5, emphasis in original). Importantly, formative assessments were used by the TGAT with respect to planning the subsequent stages, with the theory adopted in support of the view that formative assessments are recognised as pertaining to the application of assessment information by teachers to adapt their planning and teaching; this has been widely recognised in literature that followed in the field of the UK National Curriculum's application. Nevertheless, to make assessment-based decisions, there is a need for the teacher to be well positioned to interpret the assessment outcomes in the correct manner (Frowe, 2005).

In the TGAT model, there is an intention for teachers to utilise the feedback in an effort to make programmatic decisions. Nevertheless, as stated by Sadler (1989), students should be provided with feedback in an effort to identify their strengths and weaknesses and thus recognise the involvement of the student through the establishment of the conditions that may achieve improvement in student learning. The view was also held by Sadler that students should adopt an active role in the participation of self-assessment to ensure that they may achieve the greatest possible formative value from the assessment; moreover, the individual must implement the most suitable actions that may result in improvements, with learners encouraged to select the most suitable strategy or approach to improve their performance.

The perspective adopted by Sadler is comparable to the views of other researchers: for example, the work of Harlen & James (1996) emphasises that formative assessments should enable students to self-assess their progress, as well as enable teaching staff to examine and investigate the areas of strengths and weaknesses in students and subsequently make changes to their learning activities to meet students' learning requirements. Moreover, the work of Black & Wiliam (1998a) further emphasises that "feedback is a concept that is central to formative assessment" (pp. 9–10).

Importantly, Black & Wiliam (1998a) have also made reference to various points concerned with emphasising formative assessment. Student feedback should be centred on the specific qualities of the work, with suggestions made on how the work may be suggested, while ensuring comparisons are avoided. However, from the researcher's own viewpoint, comparison may be positive between students, particularly with good students who require this comparison to feel satisfied with their own efforts. It has also been established by Black and his colleague that pupils would benefit from being trained in self-assessment to facilitate their overall insight into, and understanding of, the main purpose of their work, with additional opportunities regarding how their understanding may be expressed. It is considered that this could be one aspect of teaching.

The Qualifications and Curriculum Authority (2003) has indicated that assessment for learning requires that learning be improved through the use of classroom assessment. Furthermore, an assessment that prioritises serving the purpose of promoting student learning represents assessment for learning (Black at al., 2004). It therefore demonstrates key differences between assessments that are primarily designed with the aim to satisfy accountability or certify competence or ranking.

In conclusion, throughout the earlier stages, formative assessment was applied (1988) with a focus on informing the steps to be taken by the teacher; however, priority is subsequently centred on not only planning teaching but assessment for learning as a principle of teaching and learning, with substantial consideration of participation centred on both teachers and students, such as through sharing learning goals, performing self-assessments, questioning, achieving feedback, completing peer assessments and applying the formative use of summative tests. Notably, however, one of the most in-depth and wide-ranging descriptions of formative assessment was introduced by Shepard (2006), who focused on the works completed by Black & Wiliam (1998b) and Sadler (1989) and devised a framework of formative assessment as an approach to guide student learning, as well as provide information to be used by teachers to enhance their overall instructional practice

It is valuable to recognise that in the history of formative assessment, in 2008, the Council of Chief State School Officers (CCSSO) made formative assessment a critical element of the definition implemented; thus, formative assessment was recognised as a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve student achievements of intended instructional outcomes (Melmer et al., 2008).

In more recent times, Wiliam (2011) has emphasised the value of formative assessment as a critical element centred on the enhancement of overall teacher quality, while also having the most significant effect on student outcomes. He considers several of the more commonly implemented initiatives that seek to enhance student achievement, including learning styles, and presents research that details formative assessment practices as exerting a far more profound effect on educational attainment compared with other reforms.

Effectiveness of Assessment for Learning (AfL) in Enhancement of Student Learning

The achievements of students have significantly improved since assessment for learning practices have become positioned as a fundamental aspect of classroom routines. Evidence

obtained on a global scale indicates that an effect size of half to one, as well as a half standard deviation and more may be directly attributable to the implementation of classroom assessment for learning. Thus, throughout the course of this paper, the effect size is a standardised method of drawing a contrast between the results of multiple pedagogical experiments.

Throughout the mastery learning study presented by Bloom (1984), a summary of the effects into the mastery learning models was provided, which drew a contrast between standard whole-class instruction and two experimental interventions: primarily within mastery learning conditions and secondarily within individual tutorial conditions. The in-depth, wide-ranging application of classroom assessment for learning has become a key element of instruction in both types of experimental conditions. It was highlighted throughout the course of the analysis that there were differences that ranged from one to two standard deviations in the achievements of students who favoured the assessment for learning experimental conditions. Nevertheless, other evaluations that utilised the same experiments achieved more minor effects, such as in the case of Cohen, Kulik & Kulik (1982), where the effective size was only 0.4 standard deviations.

In an assessment of 250 studies performed worldwide, as presented for the period that spanned 1987–1998, Black & Wiliam (1998a) suggested that teachers placed emphasis on assessment for learning rather than assessment of learning, and this focus created a significant increase in the achievements of learners. Regarding formative assessment, the scholars concluded that the ongoing adoption of classroom formative assessments would result in improvements in student achievements by 0.4 to 0.7 standard deviations, which represents a figure sufficiently adequate to raise the United States to the top five in international rankings. Furthermore, it was established that an improved assessment was useful in assisting low achievers more than other students and would therefore decrease the achievement range while also increasing achievements overall. Nevertheless, it was established that teacher-centred teaching and summative-focused assessments achieve positive outcomes in relation to the performance of learners.

The Iowa Tests of Basic Skills were examined by Meisels et al. (2003) via the assessment of the scores of third- and fourth-grade learners who had been involved in the Work Sampling System (WSS) programme; the aim of this initiative was centred on encouraging a curriculum-embedded performance assessment compared with a demographically comparable contrast group. The findings suggest that “Children in WSS classrooms made greater gains in math than children in the other two groups, although the results were only marginally significant when compared with gains by the matched contrast group. Effect sizes ranged from .75 to 1.5 standard deviations” (p. 2). Furthermore, Rodriguez (2004) has discussed smaller size effects with respect to an assessment of the relationship between the achievements of US math students involved in the TIMSS (Third International Math and Science Study) and assessment practices; it was subsequently concluded that students could be helped by teachers, particularly in the case of low-achievers in math, when they were taught that their own success in learning could be controlled.

In the work conducted by Ruiz-Primo & Furtak (2006), four teachers were involved in a five-day training in the application of formal embedded assessments in science instruction. The way in which these assessments should be used to garner immediate feedback for learners was taught. Student performance was measured with the use of pre-tests, embedded assessments and post-tests. It was established that the intervention had a notable effect on student performance. The findings indicated that fidelity and the overall quality of the application varied among teachers, and, in general, teachers with a greater degree of quality implementation reaped better scores in the post-test. Notably, however, it was indicated that the four-teacher sample size is substantially limited. Achievement gains cannot be attributed

to intervention when considering that other teacher characteristics may have contributed to the various differences identified in the performance of the teachers. Importantly, tasks were performed correctly by students; however, the students were often not able to provide explanations as to why they answered a specific way.

Overall, the OECD (Organisation for Economic Cooperation and Development, 2007) has made the suggestion that the studies conducted by the OECD's Centre for Educational Research and Innovation (CERI) support the previously described effects. The equity of student findings would also be improved through formative assessment, which has been applied by the schools' present data not only in mind of attaining general gains in academic achievements but also particularly high gains for students previously recognised as underachieving. The retention of learning and overall attendance have exhibited improvements, in addition to an overall improvement in the work of learners. Various aspects of the work conducted by Black & Wiliam (1998a) have been called into question in the work of Dunn & Mulvenon (2009), with the latter scholars contending that the limited scientific evidence concerning the effects of formative assessment in education is, at least in some regard, a result of the issue that there is no sound consensus regarding what constitutes formative assessment. Without agreement in this regard, the authors state that the establishment of a well-formed body of research is problematic. The authors further challenge the 1998 work as a result of the belief that it exhibits significant reliance on a study that utilised a sample composed of disabled students, which equated to 83%. The authors further emphasise that higher quality students with a more wide-ranging representation of the student population may have resulted in smaller effect sizes. The authors cited numerous concerns in relation to methodological issues in the article Black & Wiliam (1998a); however, they agree that there is support concerning the application of formative assessment and thus indicate the need for higher quality students in an effort to strengthen the research base.

Moreover, Kingston & Nash (2011) completed a meta-analysis, which established more modest mean effect sizes that notably equated to approximately 20; these findings suggest that the work of Black & Wiliam (1998a) was overly generous when estimating the effects on achievements. Numerous different elements were queried in the work of Briggs & Ruiz-Primo (2012) in relation to the study by Kingston & Nash (2011); however, there is a consensus that a notable degree of uncertainty remains in terms of the effect of formative assessment on achievement. There is increasing momentum underpinning the application of formative assessment practices in classrooms; however, the analyses completed by various scholars, such as Bennett (2011), Briggs & Ruiz-Primo (2012), Dunn & Mulvenon (2009), and Kingston & Nash (2011), provide the collective suggestion that the evidence base concerning the effects of formative assessment on the academic attainments of students has remained somewhat questionable. More specifically, they suggest the adoption of two solutions: more suitably designed studies and studies that analyse various elements that affect the overall efficacy of formative assessment.

Professional Development Related to Formative Assessment

Alkharusi (2011) indicates that previous and recent studies of classroom assessment have consistently expressed a concern regarding the sufficiency of teachers' assessment skills. However, a limited number of studies have been recognised as associated with the link between formative assessment and professional development (Cole, 2010). Meisels et al. (2003) have analysed student performance (for further details, refer to previous Section) as an outcome measure associated with the effects of professional development focused on formative assessment. Standardised test scores were analysed in terms of assessing the effects of a three-year implementation of an embedded performance assessment system in schools

self-selected for the programme's application. This study further established significantly increased standardised reading scores for the treatment group.

Furthermore, the work of Wiliam et al. (2004) has detailed numerous different elements inherent in an 18-month professional development programme aimed at providing teachers with support in further improving their application of formative assessment. Twenty-four teachers at the secondary school level (12 mathematics and 12 science) spent 6 months learning about four strategies of formative assessment, namely, descriptive feedback, questioning, self and peer assessment, and sharing learning criteria with students. Action plans were subsequently created by the teachers, which outlined the various aspects of formative assessment to be adopted in their classrooms. Throughout the duration of the school year, teacher observations were performed, with the opportunity to consider and discuss with other professionals how the formative assessment practices detailed in their action plans could be implemented. The findings suggest that individuals involved in the development demonstrate notable changes in their perceptions of themselves as professionals.

A study that utilised a sample of 25 secondary teachers in math and science, which was conducted in New South Wales Australia by Panizzon & Pegg (2007), aimed to investigate assessment and its changing nature, as well as the effects on in-classroom practices. Across a two-year study, teachers attended numerous professional development sessions and received continuous consultative support. All sessions were recorded and subsequently transcribed; interviews were completed with all teachers at the end of the school year. The key aspect inherent in the study was the questioning, whereas the six contributing aspects included assessment linked to pedagogy, attention to cognition, classroom advantages for students, classroom advantages for teachers, pedagogical practices and teaching strategies. These findings signify a significant change in the perceptions of teachers regarding assessment, which redirects the focus from the accumulation of student marks to diagnosis with the objective to direct teaching towards improving the mathematical and scientific understanding of students.

Teachers were provided with a short-term professional development session by Yin et al. (2008) with the aim to teach them how a particular embedded assessment could be applied in their science classroom. The scholars designed and implemented formative assessments in an inquiry science-based unit. The sample utilised comprised 12 middle-school science teachers and their students, all of whom were randomly selected and subsequently assigned to control or experimental groups, with 6 students per group. The embedded assessments were provided to the control teachers, without the inclusion of training. The findings indicated that teachers significantly varied in terms of their effects on achievement, conceptual change and student motivation. Nevertheless, the effects of the formative assessment treatment on these outcomes were not significant.

In a study by Phelan et al. (2009), 91 teachers were randomly assigned within schools to treatment or control groups. For the treatment, teachers were provided with training on three lessons in the 6th grade, which were to be taught throughout the year, with an emphasis placed on math and formative assessments. The POWERSOURCE© intervention specifically directs attention to middle-school-level mathematics, notably beginning at the 6th grade, as well as aiding in assuring that learners have the understanding required for them to be successful in the algebraic field. This particular intervention focuses on big ideas and associated skills in four specific fields that underlie algebraic success: (a) rational number equivalence (RNE), (b) properties of arithmetic (PA—the distributive property), (c) principles for solving linear equations (SE), and (d) the application of core principles in these areas to other critical areas of mathematics, namely, geometry and probability (RA). These areas were selected because of their perceived value in subsequently mastering algebra, as well as with respect to their valuable position in state mathematics standards across the 6th grade.

In contrast, the individuals in the control group were not provided with training; instead, they were instructed to teach the students the same math topics with their own lesson plans. Furthermore, following each of the three lessons, the teachers in the experimental group met with the researchers to review the work of the learners and subsequently analyse common errors. In a comparison of the conditions, when considering student performance of standardised math outcome measures, as well as teacher math outcomes, the findings indicated numerous key differences in favour of the treatment group. For example, overall, the POWERSOURCE© students did not outperform the students in the control groups, when considering that the researchers did not identify statistically significant main effects associated with the treatment. Notably, however, they identified a notable interaction between 39 treatment and pre-test scores, which implies that the learners with higher pre-test scores are more likely to draw benefit from the intervention formative assessment when contrasted with the learners with lower pre-test scores. The student effect size is as much as a 0.5 pooled standard deviation; thus, the intervention was more significant in relation to the higher performing students compared with the lower performing students. The PA items were recognised as being more problematic for students compared with the items concerned with other domains. Moreover, qualitative data obtained through the completion of the teacher observations and interviews highlighted that teachers recognised their students as experiencing greater problems with the PA content compared with the other POWERSOURCE© areas. Thus, there was a greater impact of POWERSOURCE© on the more difficult items.

The Teacher's Role in Formative Assessment

The adoption of formative assessments is common when teachers seek to change their teaching methods in line with student learning or when they seek to enable students to make changes to their learning in alignment with their performance feedback (Gibson & Shaw, 2010). Formative assessment is diagnostic in nature because it provides insight concerning student learning that may subsequently be applied to change and enhance both learning and teachers to improve the fulfilment of student needs (Black & Wiliam 1998b). Making changes to assessment-centred practices in schools necessitates more in-depth, precise conceptual changes among educators, as well as critical changes in thinking with respect to teaching and assessment and the overall link between the two factors (Earl, 2006).

The implementation of formative assessment in a practical situation further necessitates that teachers make changes regarding how they work with their students. It is not only necessary to enhance existing practices but also requires practice to be reconsidered to improve the essential feedback loops involved, which assist students in interacting and collaborating more in their overall learning and assessment. These changes in assessment are centred on the views that all students can learn and the role of the teacher is to establish numerous ways to ensure that this particular learning occurs (Earl, 2006). Assessment essentially becomes a fundamental element of the learning–teaching process. Thus, formative assessments that are conducted throughout the learning process itself enables changes in teaching and learning (Gibson & Shaw, 2011). This is a novel approach to assessment for the majority of teachers, and this method is in contrast to more long-term experiences and practices of teachers (Earl, 2006). Studies have emphasised that the way in which a teacher adopts formative assessment within the classroom environment ultimately depends on the teacher and their individual experiences (Stephens, 2011).

In providing effective formative assessments, dialogue was highlighted as paramount (Fox-Turnbull, 2006). More specifically, in the current work, it has been established that providing students with the opportunity to discuss their work not only aids them in their knowledge and growth but also provides teachers with insights into the way in which their

learners think. Thus, teachers may utilise this as an opportunity to respond to students in relation to the practicality, usefulness and quality of their thinking, such as through posing questions in an effort to expand understanding (Fox-Turnbull, 2006). Nevertheless, there is also a need for teachers to enable students to have the opportunity to discuss their thoughts in rationalising and explaining their choices (Fox-Turnbull, 2006). One important finding in the work of Fox-Turnbull (2006) is that a significant factor is teacher knowledge, which has a significant impact on the quality of the feedback provided to students by their teachers.

Buck & Trauth-Nare (2009) have suggested that few teachers have a real understanding regarding the pedagogical implications of scaffolding or their role in the application of formative assessments. Thus, Buck & Trauth-Nare (2009) conducted a study centred on the development of understanding the experience of being a teacher, which aims to improve their levels of knowledge through formative assessment and accordingly utilise their understanding with respect to enhancing teacher education practices. The further observation was made that questioning the class was an approach widely applied by teachers; nevertheless, most questions analysed only what students could remember. The findings of the work draw the conclusion that following the implementation of formative assessment methods within the classroom, teachers soon began to recognise an increase in student participation in the development of formative assessment. Moreover, in contrast to high-achieving students, other learners who had functioned unsuccessfully accepted the formative process and began to thrive in the classroom.

The ways in which science teachers have analysed their roles in the assessment of students and teaching have been investigated by Gioka (2009), in which the emphasis of the study was centred on teacher assessment and the feedback provided to learners. It was considered that teachers face numerous challenges; teachers are required to instruct and further support the learning of students. In contrast, teachers must also assess and analyse and subsequently grade students (Gioka, 2009). Overall, teachers provided limited feedback and students were left to make decisions regarding how they could progress their learning.

How teachers apply formative assessments was examined by Stephens (2011) in the context of middle-school classrooms. Data were obtained through the completion of field observation, teacher interviews, and written responses to student work; the results demonstrated that the adoption of formative assessment creates structure and subsequently achieves student engagement. Positive communication among students is pivotal not only in improving the levels of student motivation but also increasing their willingness to participate and cooperate. Providing learners with assignments or activities means there is less discussion time. Moreover, providing learners with the opportunity to work in small groups may also provide a platform for students to learn more through questions. The findings further emphasise that the lack of formative assessment may induce resistance to what the teacher aims to achieve.

Formative assessment in Saudi Arabia:

From a historical perspective, summative assessments, which emphasise the generation of results that highlight student performance, have been demonstrated to be the key assessment practice in schools in the Kingdom of Saudi Arabia (KSA): the examination system is recognised as being the only educational assessment tool in adoption (Al-Sadaawi, 2007). However, in recent times, Saudi Arabia has implemented a shift towards a more constructivist approach in the educational domain, which highlights problem-solving (Alaudan, 2015), analysis and research as opposed to mere repetition and memorisation (ibid). Effective assessment approaches based on constructivist views promote the integration of assessment and teaching (Bekiroglu, 2007). However, regardless of these changes, there has been the neglect to consider FA, which is regarded as best utilised in a constructivist

environment (ibid). This study poses the view that the reason for the failure to adopt formative assessment is that education in the KSA is associated with a lack of teaching training with respect to implementing an efficient assessment.

Professional development is rarely provided to science teachers in the KSA (Al-Sadaawi, 2007), with the suggestion made by Doran et al. (1994) that teachers have limited training in teacher education programmes regarding how classroom formative assessments should be conducted. Moreover, they also have limited technical assistance across their day-to-day professional practices. In addition, it is stated by Berry (2008) that numerous teachers are not even provided with adequate professional development on learning assessments. In this area, it was also established that Saudi teachers lack the necessary skills to evaluate students in the most suitable approach.

Consistent with these findings, the present work poses the argument that there is a pressing need for teaching staff to acknowledge the aim and overall method of formative assessment, as well as the effective tools of formative assessment, including collaboration among students and teachers, explicit learning goals, descriptive feedback to students with regards to their performance, and self and peer assessments. In this regard, the view is adopted by Yin (2005) that teachers engaged in training programmes should be provided with formative assessment tools and helped in their application.

The present work considers that there is a need for relevant professional training to be delivered to teachers to ensure the assessment results for improving education in schools may be used. It is held that student learning may be enhanced through formative assessment, as established in other works previously discussed, which indicates professional development on formative assessment should be focused on improving student attainment. Thus, there is a need for formative assessment to be one element of professional development because this is recognised as a fundamental aspect of teaching, which encourages the higher-order thinking of students and further assistants students in satisfying the standards to a high level of proficiency (Trumbull & Gerzon, 2013).

METHODS

a) Research Methodology

Both quantitative and qualitative methods were used in this study.

b) Research Question

What are the effects of the proposed training programme based on formative assessment to improve instructional practices among Saudi science teachers in primary schools?

c) Quantitative Approach

The researcher adopted the experimental method based on the design of the control and experimental groups to determine the effectiveness of the proposed training programme in the development of strategies for using formative assessments among Saudi science teachers in primary schools. The researcher designed the programme. The reason for using the current experimental approach was because the researcher controlled the independent variable, namely, the proposed training programme.

c1) Sample

The study population comprised all science teachers in primary schools in one educational administration in Saudi Arabia (Aljouf); the total number of teachers in this administration was 120. The specific administrations were selected because it would be easier to conduct the study where the researcher lived and worked. The administration department of the Saudi Ministry of Education was asked for a list of e-mails for all science teachers in Aljouf City where the research was conducted. An email was subsequently distributed among all teachers (120). After four weeks, 33 teachers had responded. The researcher subsequently sent another e-mail to the 33 teachers to explain the purpose of the current study and the other processes of investigation (knowledge test, the proposed training programme based on formative assessment).

Sixteen teachers agreed to participate in the experimental group. Thus, the study sample consisted of 33 science teachers from primary schools in Sakaka in the Al-Jouf region, who were divided and randomly assigned to an experimental group (16 teachers) and a control group (17). As a result of the influence of the Saudi culture and considerations regarding religion, all schools are single-sex in Saudi Arabia. The gender separation also indicated it was obligatory to conduct the study from boy only schools and among a sample of male teachers.

From a research perspective, the practicalities of the Saudi context indicated that the study could only include a sample of boys. It is expected that differences may exist between boys and girls as a result of the ways in which different genders are socialised in Saudi Arabia and considering that the inclusion of girls as participants may help establish new and important issues in Saudi education that otherwise may not be identified with boys alone. Furthermore, it should be noted that the average experience is very tight between teachers, i.e., an average of ten years. However, the homogeneity of the two samples was ascertained before applying the proposed training programme through the use of the test T (refer to the table below).

This type of test demonstrated that there are no statistically significant differences between the experimental and control groups in terms of the total score on the questionnaire for testing the level of knowledge in relation to the strategic skills of formative assessment; this finding indicated the homogeneity of the two groups prior to the application of the proposed training programme.

Table 1: Differences between the averages (the level of knowledge in relation to the strategic skills of formative assessment) of the experimental and control groups prior to the application of the proposed training programme.

Statistics / group	N	Mean	Std. deviation	df	T	Sig.
Exper	16	168.94	17.510	31	0.772	0.446
Cntrl	17	173.65	17.511			

Table 1 indicates that there are no statistically significant differences between the averages of the experimental and control groups in terms of the degree of knowledge test to measure the level of science teachers on the use of formative assessment strategies; these findings indicate the homogeneity of the two samples.

c2) Tools of the Quantitative Approach

The tools used in this approach were as follows: first, a knowledge test was prepared by the researcher to measure the effectiveness of the proposed training programme in developing the strategies and skills of formative assessment use among Saudi science teachers

at the primary level; second, a training programme was implemented (refer to the following section), which encompassed 15 training sessions for two months, with only the experimental group exposed to this programme. The test comprised 20 questions, which were multiple-choice in nature. The test questions were devised by the researchers and were prepared consistent with a literature review. A pilot that utilised a sample of 5 science teachers was conducted. With the use of the KR-20 (Kuder-Richardson), the test reliability was established. The result was identified as 0.85. The test validity was determined and validated by five experienced science teachers and six professors. As implied by this study, knowledge improvements may result in improvements in instructional practices. Importantly, the knowledge test used the Arabic language.

c3) Programme Session Description

The programme used in the current research was a 15-training session focused on the strategic skills of formative assessment use among Saudi science teachers at the primary school level. The training sessions revolved around the following themes:

- The differences between behaviourist learning theory and constructivist learning theory.
- The differences between formative assessment and summative assessment.
- The differences between learning and teaching, as well as assessment and evaluation.
- Training the teachers in the use of formative assessment strategies: for example, explicit learning goals, collaboration among teachers and students, self and peer assessments, and descriptive feedback to students regarding their performance.
- Provide teachers with practical examples of the importance of formative assessment in improving teaching.

c4) Statistical Analysis

This study aimed to identify the effects of the proposed training programme on the experimental group compared with the control group; thus, comparisons between the experimental and control groups were conducted, as well as between the tribal and dimensional measurements using T-tests to denote differences between the two groups via SPSS (Ver programme. 20.0).

d) Qualitative Approach

The researcher implemented a classroom observation with science teachers who were subjected to the proposed training programme. The reason for this approach was the need to determine the effectiveness of the proposed training programme in improving Saudi teaching in actual classrooms.

d1) Sample

Five science teachers were randomly selected from the experimental group (N = 16). The researcher used classroom observations with the five teachers prior to the proposed training programme to investigate whether the teachers used formative assessment strategies; after the programme, this method was used to determine what strategic skills were used by the teachers in relation to the formative assessment. It should be noted that interviews are not customary in the Saudi context and with this instrument (EL-Deghaidy et al, 2015). However, Flick (1998) indicated that in addition to the competencies of speaking and listening, which are used in the interviews, observation is another everyday skill that is methodologically systemised and applied in qualitative research. It is because the observer may use more than

one sense; he/she may simultaneously smell, touch, hear, and see what is observed (Flick, 1998). In the current research, I observed teachers and the way they benefited from the proposed training programme in relation to the formative assessment in class.

d2) Tools of the Qualitative Approach

Different tools were used in this approach: first, a digital recorder was used to record the session; and second, card notes were used to facilitate the codification of the items that could not be recorded or documented by the recording device. It should be noted that the researchers were non-participant observers; this role was helpful in understanding what occurred from the inside. The data obtained through the use of classroom observation were digitally recorded. Digital recording helped the researcher to capture all teacher and student conversations. Two digital recording devices were used during the entire class observations: one device was placed close to the board where the Saudi teacher frequently stood, and another device was placed at the other end of the room. Each voice recording was downloaded and converted to specific file formats.

Verbatim transcription was implemented because this approach could provide a full transcript of everything that was said in the recording, including the common instances of ‘erm’ and ‘er’, ‘you know’ and ‘I mean’. Thus, after completing all transcriptions, the decision was made to send, via email, all files of the digital recordings from the classroom observations to two colleagues who work as Arabic language teachers. They were asked to check the transcripts and review them alongside the digital recordings to provide the percentage of accuracy in the transcriptions. Overall, the feedback was helpful, and the percentage of accuracy was very high, i.e., greater than 99%. All data were collected in Arabic, whereas the research report was written in English. Thus, all data included in the research report had to be translated from Arabic to English. The researcher decided to translate all data into English that could help achieve holistic rather than piecemeal data. Thus, the researcher also decided to use machine translation (MT); this approach is currently used by many professional translators to assist their work (Aiken and Balan, 2011). The original transcription and translation drafts were sent to two Saudi friends of the researcher, who are currently earning their PhDs in Biology at Glasgow University and Science Education at York University. The researcher asked these individuals to follow the same process used when the researcher translated the original transcription. The researcher subsequently asked them to provide a percentage of the accuracy in the work. The average results demonstrated greater than 98.5% accuracy.

d3) Analysis of the Qualitative Approach

A content analysis and conversation analysis were used in this approach. The content analysis is a method that aims to investigate the meaning and content of a text (Hyde et al., 2004). A conversation analysis was applied in an effort to examine the teachers and children’s discussion in interaction.

FINDINGS

a) Presentation of Quantitative Findings

This section presents the quantitative findings based on T-tests to determine the effects of the proposed training programme in developing the level of knowledge in relation to the strategies and skills of formative assessment. For example, Table 2 indicates that there is a significant difference between the averages of the tribal measurement before the implementation of the programme and the telemetric after the implementation of the

programme, in which the experimental group was in favour of the dimensional measurement. This finding infers the success of the proposed training programme in the development of the level of knowledge in relation to the strategic skills of formative assessment among Saudi science teachers in primary schools.

Table 2: T-test of the pre and post measurements in the experimental group

Statistics Variable	Sample size	average	Std. deviation	Degrees of freedom	T- value	Significance
Pre	14	169.93	18.43	13	2.66	0.02
Post		185.68	16.11			

Furthermore, Table 3 indicates there was no significant difference in the tribal telemetric measurements in the control group. Thus, the group that underwent the programme was the only group affected; the control group was not affected.

Table 3: T-test of pre and post measurements in the control group

statistics Variable	Sample size	average	Std. deviation	Degrees of freedom	T- value	Significance
Pre	11	173	18.65	10	2.15	0.06
post		159	12.53			

Moreover, Table 4 indicates the success of the proposed programme as the experimental group was more a function of the control group in the telemetric measurement. This finding suggests the effect of the proposed training programme in the development of the level of knowledge in relation to the strategic skills of formative assessment among Saudi science teachers in primary schools.

Table 4: Differences between the experimental and control groups in the telemetric based on a T-test

statistics Variable	Sample size	average	Std. deviation	Degrees of freedom	T- value	Significance
Experimental	14	185.86	16.11	23	4.55	0.001
Control	11	159	12.53			

b) Presentation of Qualitative Findings

Observations were made both before and after the implementation of an intervention (the proposed training programme was based on the strategic skills of formative assessment).

b1) Observations prior to Intervention

During the classroom visit, all teacher/student interactions were captured during the lessons to determine whether Saudi teachers use formative assessment in their classrooms. The results demonstrate that all five Saudi science teachers in primary schools often focused on transferring only scientific information to their pupils. Therefore, pupils are viewed as empty containers, in which teachers deposit information into their minds. Pupils are treated as objects that are acted upon rather than as active and knowledgeable participants in the educational process. Moreover, they tend to use chalk-and-talk methods to explain their lessons, with pupils spending most of their time listening to the teacher and copying from the board, as opposed to being allowed to express their ideas and reflect their attitudes. At the end of the lessons, the teachers allow one pupil to answer when they ask them one or two questions to assess the pupils' understanding. However, if the pupils do not understand, the teachers cannot explain the entire lesson again; therefore, none of the questions appear to seek information unknown to the teacher. This result confirms that the teachers do not use formative assessment.

They also concentrate on improving only good pupils' performances, which requires minimal work, rather than focusing on less-able pupils, who require more time and hard work during lessons. Thus, during the classroom observations, the pupils had a lack of good support and positive feedback from their teachers when they lacked the confidence to be successful in the tasks set by their teachers. Furthermore, Saudi teachers pay close attention to the curriculum and follow textbooks that are extensive and very detailed, particularly in primary schools. Thus, the role of teachers is only concerned with covering the entire lesson, with the teacher often moving from one point to another point without assessing or checking pupil understanding. This is mainly because teachers tend to consider what they will teach rather than what pupils should learn and will understand at the end of lessons.

b2) Observations after Intervention

Five themes may be identified in relation to the formative assessment strategies used by the teachers during the interactions with the pupils, namely, sharing learning goals, open questions, assessment-based evidence, classroom dialogue (teacher–child dialogue) and classroom dialogue (teacher–child dialogue).

Sharing Learning Goals

This transcript indicates that the teacher wanted to share learning goals in terms of thinking and learning, rather than only knowing what they have to do:

Transcript 1/Classroom (5^{th} Grade):

The teacher began to write the following on the board:
'What we will learn today includes:
How does lightning occur?
How does thunder occur?
What is the difference between lightning and thunder?'
T: Today, we will have a very important and interesting lesson, and you can see on the board the objectives that we are going to learn.
T: What will we learn today?
Students: How lightning occurs (choral answer).
T: What else?
S: Why thunder occurs and the difference between lightning and thunder (choral answer).

The teachers used the learning goal to describe what the pupils should know, understand, or be able to do by the end of the lesson. Leaning goals are important for children because they are only able to assess their progress when they have a clear notion of what they aim to achieve. However, Clarke (1998) indicates that all too often, children do not have a clear idea of the purpose of their activities; therefore, classroom activities appear to children as a collection of disconnected and meaningless exercises. The key to solving this issue is through pupils sharing learning goals in terms of thinking and learning, rather than only knowing what they have to do. For example: *'What will we learn today?' 'Today, we will have a very important and interesting lesson, and you can see on the board the objectives that we are going to learn.'*

Open Questions

This transcript indicates that the teacher asked open questions to encourage children and provoke a range of answers; this approach provides access to their views regarding things and their feelings, as well as promotes enquiry by the children.

Transcript 2/Classroom (5th Grade):

T: Based on your thinking, why does thunder occur?
S 2: Voices in the sky when the storm comes.
S 5: When the rain falls.
S 7: The sound is caused by static electricity.
T: Who has any other views?
S 13: God wants it to occur.
S 4: It is the air moving.

Thus, the teachers avoid asking their pupils subject-centred questions that focus on only the subject matter; instead, the teachers wanted to consider using person-centred questions, such as '*Based on your thinking, why does thunder occur?'*. The use of the phrase 'Why do you think' encourages a pupil with an opinion or a pupil who has been thinking about this concept or has ideas regarding it to attempt to answer, regardless of whether they are right or wrong (Harlen & Qualter, 2007). In relation to the question content, Elstgeest (2001) states that productive teachers are more useful in helping children's investigation and thinking. Although the pupil provided the scientific answer, *'The sound is caused by static electricit*y', the teacher attempted to direct the pupils to provide unexpected answers, not merely scientific answers, *'Who has any other views*?'. This approach led the children to provide different types of thoughts, such as religious thinking, *'God wants it to occur*', as well as everyday thinking, *'It is the air moving*'. This is mainly because the pupils received the message that they are required to conceptualise their own answers.

Assessment-based Evidence

This transcript indicates that the teacher used evidence of the pupils' mastery status to determine whether it was necessary to make adjustments in what they are doing in relation to their instruction.

Transcript 3 /Classroom (4th Grade):

T: Do lightning and thunder occur at the same time?
T: The teacher asked the students to express their opinions through the strategy of thumb, which enabled the teacher to obtain an index survey of the students' understanding; thus, when the student indicates it does not occur at the same time, he puts the finger of a thumbs-up in front of his chest, whereas the students who disagree direct their thumbs to the bottom. This approach enabled the teacher to see the students' answers, whereas the students cannot see each other.

The majority of the students put their thumbs up.
T: Let us move on to another task.

The teacher used the thumb strategy to establish immediate feedback in relation to the children's understanding of lightning and thunder. Thus, it is clear from the previous transcript that formative assessment supplied assessment-based evidence as an indication of whether the teachers and students 'got it right'; this may play an important role in helping teachers and students to monitor the effectiveness of their efforts. The majority of the students put a thumbs-up in relation to the question *'Do lightning and thunder occur at the same time*?', which indicates there is no need for the teacher to adjust the currently instructional approach. This allows the teacher to know what he planned to do next instructionally: *'Let us move on to another task*'. Furthermore, the teachers ask questions for various purposes, such as to initiate a discussion, engage pupils in effective thinking regarding a topic (refer to the previous transcript), close a discussion (the current transcript), or quickly determine whether children follow what has been explained or whether there is confusion.

Assessment methods that assess pupils and provide feedback during teaching are essential. Teachers should gather different information in different forms at different times to make their decisions regarding pupil learning (Bekiroglu, 2007). To improve student learning, teachers must constantly support student participation, assess their understanding and provide regular feedback (Cimer, 2007).

Thus, the need to think deeply about the purposes of posing questions in the classroom should be considered in-depth, with these questions enhanced for use in the assessment of children's performance for learning. This analysis suggests that there is not only one answer in relation to whether closed questions represent a good or bad thing.

***Classroom Dialogue (Teacher–Child Dialogue*)**

This transcript indicates that the teacher allowed every pupil to think, with pupils instructed not to raise their hands if they know the answer to a question.

Transcript 4/Classroom (4th Grade)

T: In this lesson, there is no correct or wrong answer. All answers are very important in enriching the lesson and purpose of the lesson. The most important point is that everyone expresses their opinion and way of thinking. I hope everyone cannot raise their hands; it is necessary for everyone to participate.
Student 3: Lightning comes from God.
Student 4: Lightning comes from God who created everything.
Student 5: Lightning divides the two halves of trees.
Students 6: Lightning comes from the wind.
Students 7: Lightning comes because of the air that charges the water.

The teacher wanted to teach the children to raise their hands at the same time and consider only their thinking, regardless of whether they were correct. *'In this lesson, there is no correct or wrong answer. All answers are very important.'* Thus, the teachers created a supportive environment to help pupils to be comfortable with providing a wrong answer (Black et al., 2003), particularly with very young children, who tend to not answer because they are afraid of receiving negative feedback. This approach enabled the children to want to talk about the lightning from different aspects, views and thinking, such as religious thinking ('*Lightning comes from God*'), everyday thinking ('*Lightning divides the two halves of trees*') and scientific thinking ('*Lightning comes because of the air that charges the water*').

Classroom Dialogue (Teacher–Child Dialogue)

This transcript indicates that the teacher initiated a dialogue; groups of pupils were subsequently given time to discuss issues with one another and were invited to report their ideas to the class.

Transcript 5/Classroom (5th Grade):

The teacher divided the students into groups, and each group was provided with a small white paper to write on.

T: Thunder and lightning occur at roughly the same time; however, you see the flash of lightning before you hear the thunder. Why?

The teacher provided the students with three minutes to think and asked a member of each group to speak and provide their answers to encourage each group to discuss with one another.

Group A: It cannot occur. This task is wrong. The lightning and thunder cannot occur together. Lightning occurs first, followed by thunder.
Group B: We can only see the lightning, and we cannot see the thunder.
Group C: Lightning and thunder are voiceless voices.
Group D: Lightning is faster than thunder.
Group A: Light is faster than sound.
Group B: Light comes first; then comes the sound.
Group C: Light is faster than sound.
Group D: Light is faster than sound.
T: What do we conclude from this?
Group D: Light is faster than sound, which indicates the lightning occurs first and then the thunder; however, they occur at the same time.

The teacher used a group discussion strategy because it is recognised as a good approach to practice pupils' cooperation and communication knowledge (think, talk and discuss with one another). The advantage of using this strategy is that pupils who have trouble in calculations or understanding the concept may learn from group members and receive advice from one another. This is clear from Group A, who completely refused the task: *'It cannot occur. This task is wrong. The lightning and thunder cannot occur together. Lightning occurs first and then the thunder.*' However, following the interactions of the pupils in the different groups, such as B, C and D, and without intervention from the teacher, the pupils in Group A modified their knowledge and provided a correct answer: *'light is faster than sound'.* The same situation and process occurred in Groups B and C. The opportunities for the interactions between the groups included an immediate assessment of the pupils' learning and understanding regarding a topic and provided dynamic and live feedback. This work tends to be formatively assessed; thus, the pupils feel able to develop their ideas over time. This is clear from the pupils in Group D with respect to how they developed their answers and justifications for the task provided by the teacher: *'Light is faster than sound, which indicates the lightning occurs first and then the thunder; however, they occur at the same time'.* In terms of assessment, the high level of interactivity in the different groups of teaching (A, B, C and D) indicates that teachers may provide a formative (or ongoing) assessment of progress.

DISCUSSION and IMPLICATIONS

This section discusses and synthesises the findings from a view to address the research question: What are the effects of the proposed training programme based on formative assessment to improve the instruction among Saudi science teachers in primary schools? From the analysis of the quantitative findings, it may be interpreted that there was a statistically significant difference between the experimental and control groups in favour of

the experimental group. Thus, the proposed training programme for two months (60 hours) was successful in developing knowledge related to strategy skills of formative assessment among Saudi science teachers in primary schools. This finding is consistent with Wiliam et al. (2004), in which teachers who participated in professional development based on formative assessment for 18 months significantly changed their perceptions of themselves as professionals.

However, this finding is inconsistent with Yin et al. (2008), who provided science teachers with a short-term professional development session based on formative assessments. The results of the Yin et al. (2008) study demonstrated that the impact of the formative assessment treatment on these outcomes was not statistically significant. Thus, this finding was unanticipated and therefore suggests that teachers who attend a one-off professional development workshop or who are merely provided with new instructional materials make few changes to their practice (Meiers & Beavis, 2005). Darling-Hammond et al. (2009) demonstrated that professional development programmes that provide 30 to 100 hours (an average of 49 hours) over a period of 6 to 12 months had a "positive and significant effect" on student achievement (p. 9). Professional development that provided a limited number of hours (5–14) produced no significant impact on achievement. This finding has important implications for the Ministry of Saudi Education that providing support for science teachers is critical to implement what they are learning.

Teachers should be trained on assessment programmes to be updated with new moves (Alotabi, 2014). However, professional development requires strong, highly visible, organisational support and may be implemented in many ways (Almazroa & Alshamrani, 2015). Sufficient time, resources, and professional assistance should be provided to support Saudi teachers to integrate new knowledge and skills into everyday practices (ibid). Organisational support may represent a key to the success of professional development efforts (ibid) when Saudi teachers participate in a professional development programme on formative assessment, in which teachers must implement their training in schools where students are only graded on a summative assessment (ibid).

Furthermore, the education programmes of Saudi teachers should address higher-order thinking skills in a concrete way to help teachers observe what they look like at different levels and determine how to teach pupils the thinking skills required to complete projects. In addition, teachers need instruction and practice in the skills important for assessment, such as recording anecdotal observations, providing written and oral feedback, and analysing information collected from various types of assessments (The National Foundation for the Improvement of Education (NFIE),1996).

Furthermore, for the qualitative findings, it is somewhat unexpected that due to the effective of the proposed programme the teachers used different types of formative assessment strategies, such as sharing learning goals, open questions, assessment-based evidence, classroom dialogue (teacher-child dialogue) and classroom dialogue (teacher-child dialogue). These strategies play a crucial role in helping teachers to improve their instruction. For example, prior to the intervention, the Saudi teachers asked very few closed questions to seek the correct answers.

In contrast, after the intervention, the Saudi teachers asked different types of opened and closed questions, and the target was to encourage more students to take turns to think, talk and argue. This interaction helped the Saudi teachers to view the learning as an active rather than passive process and attempt to practice a student-centred approach instead of the teacher-centred approach that is common in Saudi schools. These findings are consistent with Buck and Trauth-Nare (2009) who demonstrated that following the implementation of formative assessment strategies within the classrooms, successful science teachers asked open questions to gain the trust of their students so the students would not shut down because they do not

know the correct answers. The current findings indicate that the teachers used the expression "*there is no correct or wrong answer. All answers are very important in enriching the lesson and purpose of the lesson. The most important point is that everyone expresses their opinion and way of thinking*". Thus, children are important as individuals, social beings and inquirers. This is crucial to shift the learning from traditional teaching and constructivist learning (ibid).

In another example, a group discussion strategy helped Saudi teachers to assess the students in a way that enabled them to actively engage in making cumulative progress during the process of teaching and learning, rather than viewing the assessment as a product of the learning experience, even when final/modular examinations dictate formal results (Black & Wiliam, 1998b). Formative assessment aims to support student development rather than provide definitive and 'summative' measures of progress. A group discussion strategy improves the Saudi teachers' instructions because they do not emphasise providing a didactic lecture, which covers the subject matter; instead, they adopt the role of a facilitator, who "provides guidelines and creates an environment" for children to arrive at their own conclusions (Alao, et al., 2010, p.56). When formative assessment is conducted during teaching, it provides "both students and teachers with feedback concerning the learning and the teaching of the lesson" (Bulunuz & Bulunuz, 2013, p.130).

This finding runs parallel with Panizzon & Pegg (2007) who indicated that when teachers attended a series of professional development sessions based on formative assessment, they had a major shift in their perceptions of assessment from a focus on the accumulation of student marks to diagnosis as a means of directing teaching to enhance students' scientific and mathematical understandings.

Moreover, Black & Wiliam (1998a) demonstrated that meaningful formative assessment is not a key priority for many teachers because teachers meet challenges to implement formative assessment in their classrooms and schools. This study begins to highlight several implications for the Ministry of Saudi education, which is ultimately responsible for the educational system and practices. Many of the Saudi curriculum materials commonly in use do not provide the type of curriculum-embedded assessments that evoke pupil thinking and do not provide interpretive frameworks for teachers to use. "There is need to increase the number of formative assessment probes in science textbooks and other resource materials" (Bulunuz & Bulunuz, 2013, p.132).

The implementation of formative assessment may be limited by other conditions in the classroom. Carless et al. (2003) determined that teachers at all levels tend to experience heavy workloads; under these circumstances, it may be a challenge to expect them to engage in learning-centred discussions with individual pupils. When teachers are overloaded with numerous responsibilities, they may perceive formative assessment as something of a luxury. Moreover, Furtak (2005) indicated that teachers may become satisfied too quickly with their assessment conversations and may also become frustrated by the time and effort necessary to conduct an effective formative assessment.

Furthermore, Alotabi (2014) states that a substantial number of students in a classroom imposes limitations on the classroom activities. Alotabi adds that educators in Saudi Arabia have discovered that oversized classes may affect the application of the new form of assessment in primary schools. High numbers of students in classrooms prevent the teachers from following student progress and providing suitable feedback (ibid). To conclude, this study indicates that training programmes in Saudi Arabia should provide science teachers with the skills and knowledge necessary to implement formative assessment in classrooms because many teachers understand assessment only in terms of providing pupils with marks, indicating their mistakes or moving from one level to the next level. The implementation of this approach represents a substantial task and novel direction in the Saudi and Middle East context, particularly within the existing teaching traditions.

REFERENCES

Alao, K. A., Kobiowu , S. V., &Adebowale, O. F. (2010). *Fundamentals of educational and counselling psychology.* United Kingdom: Strategies Insight Publishing.

Alaudan, R. (2015). *Saudi student teachers' perceptions of formative assessment*. Unpublished thesis (PhD), University of York.

Al-Abdulkareem, S.A. (2004). *Investigating science teachers' beliefs about science and science teaching: struggles in implementing science education reform in Saudi Arabia.* Unpublished thesis (PhD). West Virginia University, Morgantown, USA.

Alkharusi, H. (2011). Teachers' classroom assessment skills: influence of gender, subject area, grade level, teaching experience, and in-service assessment training. *Journal of Turkish Science Education*, 8 (2), 39-48.

Alotabi, K.A. (2014). Student assessment strategies in Saudi Arabia: a case study of pre and post classroom practices. *Literacy Information and Computer Education Journal (LICEJ)*, 3 (1), 1786-1763.

Almazroa, H., & Alshamrani, S. (2015). Saudi science teacher professional development: Trends, practices, and future directions. In N. Mansour and S. Alshamrani (Eds), *Science education in the Arab Gulf States: Visions, sociocultural contexts and challenges* (pp.3-21). Rotterdam: Sense Publishers.

Al-sadaawi, A.S. (2007). *An investigation of performance-based assessment in science in Saudi primary schools.* Unpublished thesis (PhD), Victoria University.

Bekiroglu, F.O. (2007). Performance based assessment: theory and practice. *Journal of Turkish Science Education.* 5(1), 113-131.

Bennett, R.E. (2011). Formative assessment: a critical review. Assessment in Education: Principles, Policy & Practice, 18 (1), 5 –25.

Berry, R. (2008). *Assessment for learning. Hong Kong*: Hong Kong University.

Black, P., & Wiliam , D. (1998a). Inside the black box: raising standards through classroom assessment. London: King's College.

Black, P., & Wiliam, D. (1998b). Assessment and Classroom Learning. *Assessment in Education*, **5,** (1), 7-74.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning Putting into Practice*. Maidenhead: open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2004). working inside the black box :assessment for learning in the classroom . *Phi Delta Kappa*n, 86 (1), 8-21. Bloom, B. (1984). The search for methods of group instruction as effective as on-to-one tutoring. *Educational Leadership*, 41 (8), 4-17.

Buck, G., & Trauth-Nare, A. (2009). Preparing teachers to make the formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education*, 20 (5),.475-494.

Bulunuz, M., & Bulunuz, N. (2013). Formative assessment in science teaching and demonstration of its effective implementation. *Journal of Turkish Science Education*, 10 (4),119-135.

Briggs, D.C., & Ruiz-Primo, M.A. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. Educational Measurement: Issues and Practice, 31(4), 13-17.

Carless, D. (2003, November 24-25). Learning-oriented assessment. *a paper presented at the Evaluation and Assessment Conference,*. Adelaide, South Australia.

Cimer, A. (2007). Effective teaching in science: a review of literature. *Journal of Turkish Science Education*, 4 (1), 20-41.

Clarke, S. (1998). *Targeting assessment in the primary classroom*. London: Hodder and Stoughton.

Clarke, S. (2000). Getting it right-distance marking as accessible and effective feedback in the primary classroom. In S. Askew (Ed.), *Feedback for Learning* (pp.32-45). London: RoutledgeFalmer..

Cole, M.W. (2010). *Influence of assessment for learning professional development in rural Georgia publics schools*. Unpublished thesis (PhD). Liberty University.

Cohen, P.A., Kulik, J.A., & Kulik, C.L.C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journa*l, 19 (2), 237-248.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, A., & Orphanos, S. (2009). *Professional learning in the learning profession: a status report on teacher development in the United States and abroad*. Oxford, OH: National Staff Development Council.

Doran, R. L., Lawrenz, F., & Helgeson, S. (1994). Research on Assessment in Science. In D. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning* (pp.388-442). New York: MacMillan.

Dunn, K., & Mulvenon, S. (2009). A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education. Practical Assessment, Research, & Evaluation, 14(7), 1-11.

Earl, L. (2006). Assessment – a powerful lever for learning. Brock Education, 16 (1), 1-15.

Elstgeest, J. (2001). The right question at the right time. In W. Harlen, (ed.), primary science: *Taking the Plunge, (2* nd edn, pp. 25-47) Portsmouth, NH: Heinemann.

Flick, U. (1998). *An Introduction to qualitative research*, London: SAGE.

Fox-Turnbull, W. (2006). The Influences of teacher knowledge and authentic formative assessment on student learning in technology education. *International Journal of Technology and Design Education*, 16(1), 53-77.

Frowe, I. (2005). The formative use of assessment information in planning-the notion of contingent planning**.** *British Journal of Educational Studies*, 53 (1), 54-65.

Furtak, E.M. (2005). *Formative assessment in K-8 science education: A conceptual review*: Commissioned paper by the National Research Council for Science Learning K-8 consensus study.

Gibson, K., & Shaw, C.A. (2010). Assessment of active learning. Retrieved from http://webs.wichita.edu/depttools/depttoolsmemberfiles/carolynshaw/Gibson%20Shaw%20compendium.pdf. Accessed: 3th August 2014]

Gioka, O. (2009). Teacher or examiner? The tensions between formative and summative assessment in the case of science coursework. *Research in Science Education*, 39(4), 411-428.

Kingston, N., & Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. Educational Measurement. Issues and Practice, 30(4), 28-37.

Harlen, W., & James. M. (1996, April). Creating a positive impact of assessment in learning. *Paper presented American Educational Research Association*, New York.

Harlen W., & Qualter. (2007). *The teaching of science in primary schools*. (4th Ed) London: David Fulton Publishers.

Hyde, A., Lohan , M., & Mcdonnell, O. (2004). *Sociology for health professionals in Ireland*. Dublin: Institute of Public Administration.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives*, 13(10), 1-28.

Lambert, D., & Lines, D. (2000).*Understanding assessment: Purposes, perceptions, practice.* New York: Routledge Falmer.

Meisels, S., Atkins-Burnett, S., Xue, Y., &Bickel, D. D. (2003). Creating a system of accountability: the impact of instructional assessment on elementary children's achievement scores. *Educational Policy Analysis Archives*. 11(9), 1-18.

Melmer, R., Burmaster, E., & James, T. K. (2008). Attributes of effective formative assessment. Washington, DC: Council of Chief State School Officers.

National Foundation for the Improvement of Education, (NFIE). (1996). *Teachers take charge of their learning: transforming professional development for student success* Washington, D.C.: NFIE.

Rodriguez, M.C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Educatio*n, 17(1), 1-24.

Ruiz-Primo, M. A., & Furtak, E.M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. Educational Assessment, 11, (3-4), 205-235.

Sadler, D. (1989). Formative and the design of instructional systems. *Instructional Science*, **18 (**2), 119-144.

Shepard, L.A. (2006). Classroom assessment. In R. Brennan (Ed.), Educational measurement (4th ed., pp. 624–646). Westport, CT: Praeger.

Stephens, A. (2011). *Formative Assessment: Understanding Learning in the Middle School Classroom*. Unpublished Thesis (MS Edcuation). John Fisher College.

Task Group on Assessment and Testing, (TGAT). (1988). *National curriculum: a report Task Group on Assessment and Testing.* London: Department of Education and Science/Welsh Office.

Trumbull, E., & Gerzon, N. (2013). *Professional development on formative assessment: Insights from research and practice*. San Francisco, CA: WestEd.

Panizzon, D., & Pegg, J. (2008). Assessment practices: empowering mathematics and science in rural secondary schools to enhance student learning. *International Journal of Science and Mathematics Education*, 6 (2), 417-436.

Phelan, J., Choi, K., Vendlinski, T., Baker, E. L., & Herman, J.L.. (2009). *WWC Review of the Report "The Effects of POWERSOURCE Intervention on Student Understanding of Basic Mathematical Principles"* Retrieved from *https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=WWCSSR10051*[Date Accessed: 3th August 2014]

Wiliam, D. (2011). Embedded formative assessment. Bloomington, IN: Solution Tree.

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles Policy and Practice*, **11**(1), 49-65.

Yin, Y. (2005). *The influence of formative assessments on student motivation, achievement, and conceptual change*. Unpublished thesis (PhD), Stanford University.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., & Young, D.B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. Applied Measurement in Education, 21(4), 335-35.