An Investigation into Annual Faculty Evaluation Process of EFL Teachers in the Middle East: Challenges for Leadership & Management

Asif Ahmed

English Language Institute, King Abdulaziz University, Saudi Arabia

Jabreel Asghar

English Language Institute, King Abdulaziz University, Saudi Arabia

Abstract—**This study examined the effectiveness of the Annual Faculty Evaluations Process at an English language centre in the Middle Eastern context. It sought to determine to what extent this evaluation process helped the English language teachers improve their teaching practice. A mixed-method design was used to collect quantitative and qualitative data. The data were collected via teacher questionnaire & teacher interviews and triangulated via cross examination of the documents pertaining to the teacher evaluation policy and procedures used by the language centre. The findings of this case study revealed that the teachers believe that the process of annual faculty evaluation (AFE) has had a very positive impact on their professional development. However, the AFE process is quite rigid and does not take cultural diversity of the teachers into account. The report makes various recommendation in terms of teacher autonomy as well as participation in the evaluation process and using the three-track evaluation system (Danielson and McGreal, 2000) by the administration.**

Index Terms—**observation, teaching evaluation, professional development**

I. INTRODUCTION

According to Boyd (1989 "a teacher evaluation system should give teachers useful feedback on classroom needs, the opportunity to learn new teaching techniques, and counsel from principal and other teachers on how to make changes in their classroom" (p.1). A recurring theme in almost all teacher evaluation systems is the importance of establishing clear understanding of the objective of the system which must then be aligned with procedures involved.

This study examined the effectiveness of the annual faculty evaluation (the AFE) Process initiated in the academic year 2010-11 at a PYP run by an English Language Centre (the ECL) at a Middle Eastern university. The ECL was originally established about three decades ago, to teach English as a foreign language (EFL) to prospective students of various disciplines at the universality. The ECL is comprised of 12,000 male and female students, studying in two gender-segregated campuses, and approximately 500 culturally-diverse faculty across both campuses. This study focuses on the men's campus faculty that includes about 200 EFL teachers.

All the faculty members are evaluated by means of classroom observations on annual basis. This evaluation is summative and directly linked to their contract renewal at the end of the academic year. If a teacher is rated low on classroom performance, he immediately figures out that his job is at risk. The AFEs have been introduced five years ago at the ECL, for the quality assurance and performance evaluation of teachers. Since then, they have become the most debated practice because of their rigid nature and their main reliance on only one mode i.e., announced classroom observations. As an after-effect of these observations, teachers started leaving the organization which became a serious challenge for the ECL administration. Meanwhile, some observers/evaluators received no professional training prior to the practice whereas others received an orientation for only one or two days before taking up the job. Apart from this, the formal evaluation at the ECL primarily required 2 formal and one informal observation for performance evaluation every semester. In other words, every teacher had to undergo.

In addition to 5 observations per academic year i.e., two formative and four summative observations (later in 2011-12, the number of observations was reduced to one formal observation per semester), all teachers have to submit annual teacher development plans (TDP). Several in-house training workshops and focused group discussions were also created to help teachers discuss certain areas they wanted to develop amongst themselves. This was a positive change and created an opportunity to learn new teaching techniques, as well as counsel from other teachers on how to make changes in their classrooms.

Nevertheless, the instrument being used for classroom observations is still a checklist of behaviors encompassing 13 teaching standards with five ratings where 5 is outstanding and 1 is unsatisfactory. Any teacher who is rated 3 or below on the AFE is generally believed to be in trouble. Sadly, majority of the teachers somehow end up with a 3 on the AFE

rating, hence having stress and job insecurity. This is the context of the problem in which this paper evaluated the AFE process in order to propose certain recommendations to improve the situation both for teachers and the administration.

II. LITERATURE REVIEW

There are two purposes of performance evaluation i.e., formative and summative. The choice of approach greatly depends on the purpose of evaluation. Screening out unsuitable candidates and terminating incompetent teachers are summative functions. On the contrary, providing constructive feedback and providing direction for teacher development are formative purposes.

Bailey (2006) defines that the purpose of formative evaluation is 'to provide feedback on an ongoing project, to improve the program' whereas the summative evaluation 'is a final assessment, a make-or-break decision at the end of a project or funding period' (p. 184). While formative evaluation is associated with helping, summative evaluation is associated with "terminating".

An evaluation system that fosters teacher learning will obviously differ from the one whose aim is to measure teacher competence. Teacher evaluation systems have not accurately measured teacher quality because they have failed to do a good job of discriminating between effective and ineffective teachers and have not aided in developing a highly skilled teacher workforce (Weisberg, Sexton, Mulhern, & Keeling, 2009). Measuring teachers and developing teachers are different purposes with different implications. An evaluation system designed primarily for measurement will look quite different from a system designed primarily for development.

The relative importance of these two aspects of evaluation is significantly different for different groups of people. Legislators and policymakers tend to value the summative purposes. They make the point that public schools are, after all, public institutions, supported by taxpayer's money, and that the public has a legitimate interest in the quality of the teaching that occurs there. It is through the system of teacher evaluation that members of the public, through their legislators and local boards of education ensure the quality of teaching. A parent, in other words, is entrusting the education of his child to the public schools, has a right to expect a certain minimum level of performance. Educators, on the other hand, tend to think that teacher evaluation should be designed for the purpose of professional development and the improvement of teaching. Experienced practitioners argue that professional dialogue about teaching, in a safe environment, managed and led by teachers, are the only means by which teachers can improve their practice (Marzano, 2012).

A. Classroom Observations

The present study focuses on the AFEs by means of classroom observations. Therefore, it is important to begin with the rationale for these classroom observations. Bailey (2001) enlists four broad functions of classroom observation in English Language Programs: (1) Pre-service teachers in the practicum observed by teacher educators; (2) Practicing teachers observed by colleagues for professional development; (3) Practicing teachers observed by supervisors/coordinators to judge their performance; and (4) Teachers observed for collecting data in classroom research.

Teacher evaluation by means of classroom observation emerged as a career track in later decades of twentieth century. By then, language teaching had begun to be a profitable commercial enterprise and supervisors were needed to ensure that customers were satisfied and were getting what they were paying for. Bailey laments, 'very few language teachers ever receive any formal preparation for carrying out supervisor's responsibilities' (2006:3). Most of the language teachers are promoted to supervisors because of reasons other than their professional preparation for the job.

Classroom observations are mainly conducted for teacher evaluation. Teacher trainers and educational researchers believe that classroom observations can provide useful feedback to teachers, and they can improve the overall effectiveness of teaching-learning process. But combining professional development with evaluation is a real challenge faced by the ELP administrators.

Sheal (1989) enlisted four obstacles that do not let teachers benefit from observations. First, most classroom observations are conducted by administrators rather than by practicing teachers. Peer observations are not very common. Consequently observations tend to be seen as judgmental. Secondly, much of the observation that goes on is unsystematic and subjective. Administrators and teachers are generally not trained in observation. Consequently they tend to use themselves as a standard. Thirdly, most observations are for teacher evaluation purposes, with the result that teachers generally regard observations as a threat. This leads to tension in the classroom and between teacher and observer at any pre or post-observation meetings. Finally, post-observation meetings tend to focus on the teacher's behavior - what he/she did well or might do better - rather than on developing the teacher's skills. Teachers tend to react in defensive way, and given this atmosphere, even useful feedback is often 'not heard'.

He concludes that if classroom observation is to be used for staff development and to improve the quality of instruction then the focus needs to shift towards teacher development rather than teacher evaluation. The issue of class observation is remarkably similar to the issue of performance management in the world of business and industry. Just as traditional performance reviews fail to improve employee performance, in the same way traditional subjective classroom observations fail to improve teacher performance.

Sheal (1989) approves of classroom observations provided they are developmental; on the contrary, Quirke (1996) is strictly against any classroom observations at all. He believes that both teacher and students are usually on their good behavior, so observing lessons is a charade. Whether such observations should take place and why is a crucial question, and the answer is rather political and strategic, but logic makes it easier. A teacher's main responsibility is to teach effectively. How can decisions about his performance, promotion or termination be fairly made without first-hand knowledge of his teaching. Therefore, to evaluate teachers fairly, it is important to see them teach. How it should be done is another debate.

Those who agree with Quirk (1996) that classroom observations are of little value believe that both teacher and students put on a show. Therefore, they choose Surprise Observations because they want to see a normal lesson which is not specially prepared, however, this practice too has serious repercussions because it leads to mistrust and defensiveness.

Amidst this debate, Powell (1999) suggests an alternative idea of 'Unseen Observations'. He calls it 'unseen' because the supervisor neither visits the class nor watches a video tape of the lesson. Instead, the teacher submits a detailed lesson plan and then candidly reports what occurred during the lesson. Powell believes 'self-observation' is a very useful tool for encouraging teachers to increase their awareness of their own teaching. This system too has serious drawbacks. In self-observation one loses the opportunity of getting likely professional feedback from an expert observer.

B. Three-track System of Teacher Evaluation

Danielson & McGreal, (2000) suggest three-track system for teacher evaluation. Track 1- The Beginning Teacher Program is for making teacher retention decisions and moving the probationary teachers to tenured position with professional and personal development opportunities. It encourages teachers and administrators to understand the importance and usefulness of evaluation, and supports the practice of reflection and professional learning. Classroom observations remain the most common tool for teacher performance evaluation with a pre- observation conference, the observation itself and the post-observation conference for feedback. The observation itself is not evaluation, rather, it is a source of collecting evidence to be used for professional discussion and reflection on teaching and learning. The reliability and usefulness of classroom observation is linked to the amount and type of information observer has before the observation. Similarly, observations are reliable if the observers are trained and observation and conferencing skills. Post-conferences are a time for reflection, review and constructive feedback, and reinforcement. Those teachers who demonstrate competency in meeting the required standards are moved to track 2 which is professional growth track.

Track 2 – Professional Development Track "professional development is the process by which competent teachers achieve higher professional competence and expand their understanding of self, role, context, and career (Suke & Stiggins, 1990 quoted in Danielson & McGreal, 2000, p. 99). Professional Development Track implies an acceptance of alternative forms of assessment and a commitment to change the evaluation system. Danielson & McGreal (2000) claim that "it is not only the teacher's own motivation, awareness, and imagination, but also the policies and practices of the school in which she teaches (p. 99). Both quality assurance and professional learning can be incorporated in one evaluation system – and Track 2 is the place to do it. The purpose of this track is to provide a structured, supportive, and collaborative environment for professional learning that will further enhance student learning. An organization that chooses this format for evaluation, places all tenured and experienced teachers automatically in track 2. They remain fully involved in this track unless the supervisor begins to have concerns about the teacher's performance. At this point, the supervisor contacts the teacher, raises the concerns, and determines with the teacher how they should address the concern. This could involve scheduling classroom observations, but only if observation would fit the type of concern being raised. If the teacher positively addresses or clarifies the concern, involvement in track 2 continues normally. If the problem is real and cannot be resolved during the initial contacts between the teacher and the supervisor, then exactly the same action occurs as would in a cycled summative model: the supervisor or peer review team moves him to assistance track which is track 3.

Track 3 – Teacher Assistance Track with a purpose to provide organizational support and assistance to teachers who are not meeting organization's teaching standards. This track demonstrates organization's commitment to quality teaching by providing a supported and focused system of assistance to ensure that every staff member is meeting the organization's standards. This track is for tenured and experienced teachers who are experiencing problems in meeting the teaching standards and are identified as marginal teachers. It is not for the probationary or new teachers because such teachers should be eliminated through the established procedure in track1. There are usually two dominant groups involved in the assistance track. The first group is of those teachers who are experiencing stress in their lives such as divorce, serious illness, the death of a loved one, bankruptcy, and other forms of depression – all these events affect classroom performance. This group of teachers tends to respond favorably to any support provided in track 3. The second group is usually a small number of experienced teachers who never had structured professional assistance – such teachers are marginal, but not incompetent. They have the capacity to improve their skills. But the desired improvement will not take place through teacher's individual initiative, rather it is the structured assistance which can help such teachers and it is the organization that should cater for such structured assistance. After all this support, if a teacher still fails to meet the teaching standards then the final resort will be termination of his services.

Bailey (2006) suggests five options between the supervisor control and the teacher control when it comes to decisions on evaluation format (figure 1). The extreme opposite of the Unseen Observation is the Surprise Observation.

| Surprise Observations | Announced Observations | Negotiated Observations | Invited Observations | Unseen Observations |
| **Greater Supervisor Control** | | | | **Greater Teacher Control** |

Figure 1. Supervisor Control

The surprise visit is to see a normal lesson, rather than one the teacher has specially prepared whereas an unseen observation is an alternative to conducting classroom observations. The concept of unseen observation is based on structured self-evaluation conducted by an individual and later discussed with the supervisor. It is called unseen because the supervisor neither observes the class nor sees a video tape of the lesson; it is descriptive rather than prescriptive. It encourages practitioners to articulate their general principles then analyze them for contradictions. Teachers have a chance to see what they had planned and what actually happened in the lesson. Announced observations can also be negotiated simultaneously. If teachers are allowed to choose the time they deem convenient for observations then they get more control over scheduling process and hence more confidence and less stress, but it may not always be possible or feasible for evaluators especially where there are few observers and a large number of teachers to be observed in a tight timeframe. It should be convenient for both parties involved. Another alternative is to allow teachers to decide on their own and invite their supervisors to visit their classes. It is a proactive approach which allows more autonomy to teachers but of course with an expectation that teachers come forward.

III. METHODOLOGY

There were 194 male teachers at Men's campus of ELC at the time of data collection. The faculty included 79 Arab non-native English speakers, 53 non-Arab non-native English speakers whereas 62 native speakers of English. The data was collected from 45 teachers taking 5 teachers from the each category above. The sample included all ratings such as outstanding, above average, satisfactory and needs development as per the AFE for academic year 2012~13. It was ascertained that each category accommodated veteran and novice teachers (Figure 2).
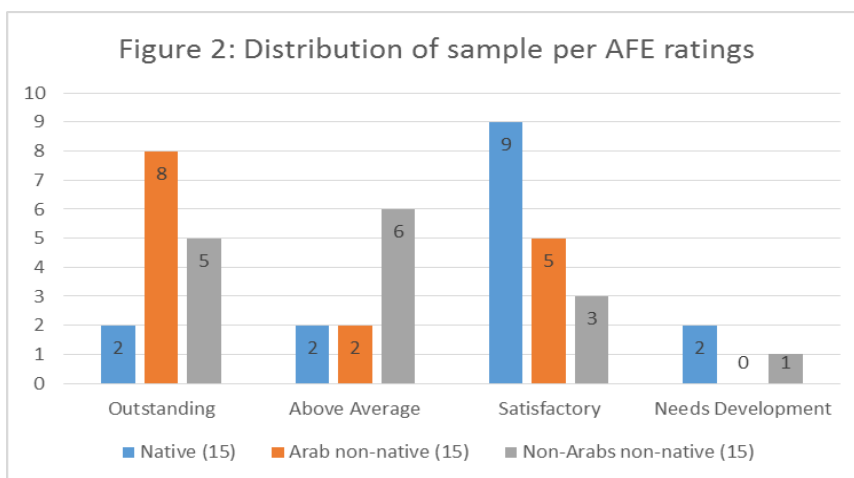

Figure 2. Distribution of sample per AFE ratings

Questionnaires solicited data in 2 broad categories: Facts & Opinions. Factual information required general information of the respondents such as age, nationality, qualification etc. The opinion questions require respondents to reveal their impressions rather than mere reporting of facts. There were7 sections (A-G) of the questionnaire designed for this study. Section A solicited demographic information; sections B & C solicited rating attributes; section D solicited teachers' perceptions of the most recent observer; sections E-G solicited the attributes of the procedure, feedback and evaluation context during the AFE process. After administering the questionnaire, 9 questionnaire respondents were interviewed to gain deeper insight into their responses about the AFEs.

Quantitative data was obtained through a five-point Likert scale and then the raw data was transformed into groups and tables. Next, percentage of responses was calculated for each category. A step further, the five point scale was reduced to three point scale for clarity and accuracy in the analysis; hence strengths and weaknesses of the AFE process were identified. On the other hand, the qualitative data was obtained through semi-structured interviews. All interviews were recorded in English and important chunks were transcribed. For triangulation purposes recurrent themes were identified and further crosschecked against quantitative data whether the two sources complemented or contradicted each other. The following section analyses the data as per parts in the questionnaire. The first part of the questionnaire recorded demographic information of the respondents which do not have impact on the data acquired from the following sections.

For qualitative data 3 teachers were selected from each category in the form of semi structured interviews in cognizance with the questionnaire. Interviews were recorded with interviewee's consent and relevant parts of the recordings were transcribed for data analysis.

IV. QUANTITATIVE DATA ANALYSIS

Responses to questions 6-8 in part B seek information on the participants' perceptions towards overall quality of evaluation, the impact of evaluation on professional practice, and the impact of evaluation on professional growth at the ECL (Figure 3). Generally the responses in this part reflect positive inclination of the respondents towards the quality of evaluation and its impact on their professional practice and growth. The average score of the three items in this part in terms of 'agree' and 'strongly agree' is 51.84%. On the contrary, the average score of the three items in terms of 'disagree' and 'strongly disagree' is 10.36%. Hence, the results of part 2 indicate that about 52% participants are of the opinion that the quality of evaluation at the ECL is very high and it has positive impact on their professional practice and professional growth. Fifty-eight percent respondents said that the evaluation process had positive impact on their professional growth whereas 47 % respondents believed that the evaluation process had positive impact on their professional practice.
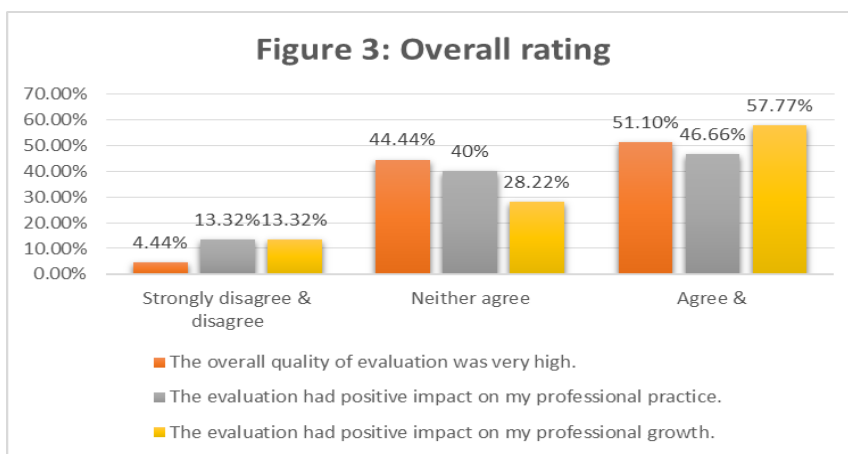

Figure 3. Overall rating

A. Rating Attributes of Evaluation

The results of part C show a positive response of the respondents towards the rating attributes of evaluation. The positive average score of the eight items in this part is 56.65%. On the contrary, the average score of the eight items in terms of 'disagree' and 'strongly disagree' is 10.82%. The results of part C showed that around 36% participants did not have enough knowledge of the technical aspects of teaching. The highest reported positive response is for question 9 which means 60% of the respondents are optimistic about the strength of their professional expectations whereas the lowest response, in the same category, is 47 % denoting that respondents believe that the evaluation process had positive impact on their professional practice.
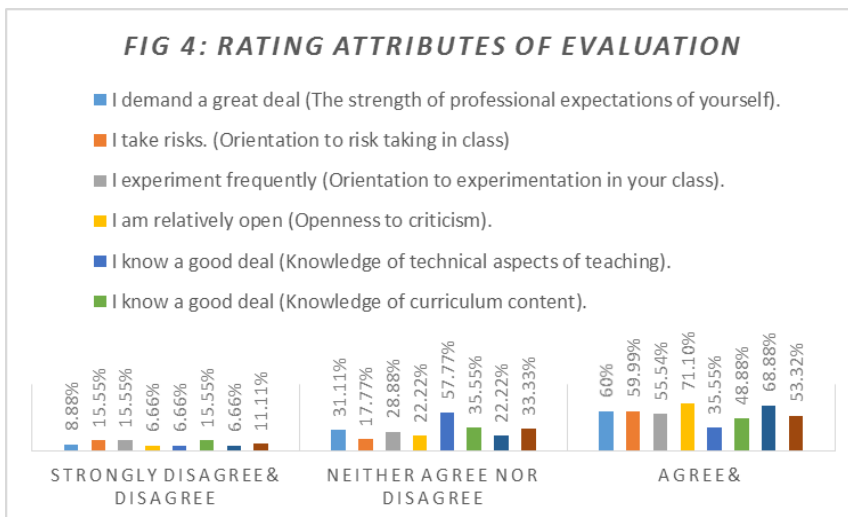

Figure 4. Rating attributes of evaluation

B. Perceptions of the Most Recent Observer

Part 4 seeks information about the participants' perceptions of the most recent observer in terms of the usefulness of the feedback provided, helpfulness of the observer, patience of the observer, flexibility of the observer, and usefulness and persuasiveness of the suggestions provided after the observation. The overall results of part 4 show a positive response of the respondents towards their most recent observer (Figure 5). The average score of the five items in this part in terms of 'agree' and 'strongly agree' is 43.10 % On the contrary, the average score of the five items in terms of 'disagree' and 'strongly disagree' is 20.88%. The highest reported response in terms of agreement ('agrees' and 'strongly agree') is for question 19 which means 53.33% of the respondents are of the opinion that their most recent observer was very patient whereas the lowest response, in the same category, is for question 18 which means around 35.55 % respondents believe that their most recent observer was very helpful.
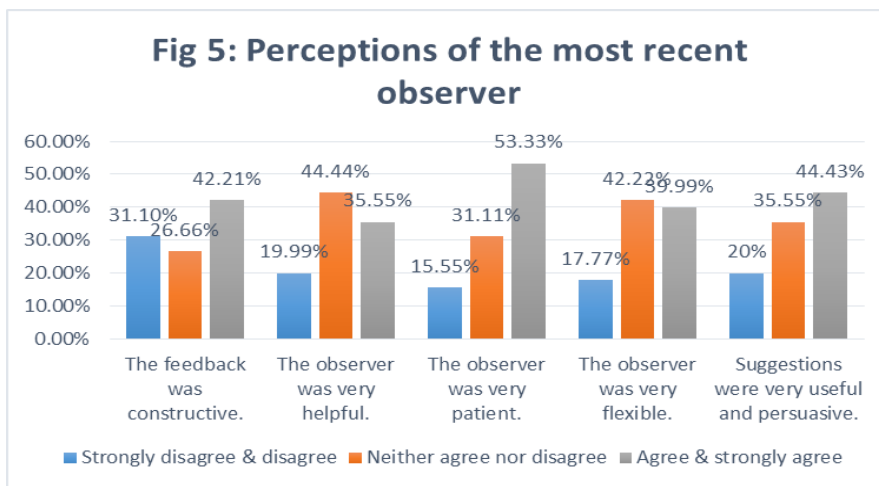

Figure 5. Perceptions of the most recnt observation

C. Attributes of the Procedure Used during the Most Recent Evaluation

Part D seeks information about the participants' perceptions about the attributes of the procedure used during their most recent evaluation regarding communication of the standards prior to the observation, clarity of the standards, and appropriateness of the standards for the current teaching assignment. The overall results of part 5 show a positive response of the respondents towards the attributes of the procedure used during the most recent evaluation (Figure 6). The average score of the three items in this part in terms of 'agree' and 'strongly agree' is 57.77 % On the contrary, the average score of the three items in terms of 'disagree' and 'strongly disagree' is 10.36%. The highest reported response in terms of agreement ('agrees' and 'strongly agree') is for question 22 which means 66.66% of the respondents acknowledge that standards were communicated in great details whereas the lowest response, in the same category, is for question 24 which means 44.44 % respondents agree with the appropriateness of the standards with their current teaching assignment at the ELI.
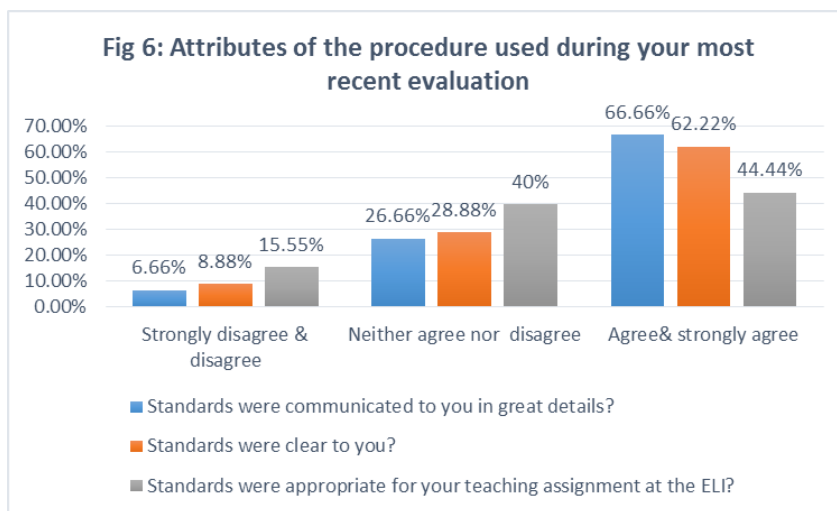

Figure 6. Perception of the procedure used during your most recent evaluation

D. Describe the Attributes of the Feedback You Received during Your Last Evaluation Experience?

Part E seeks information about the participants' perceptions about the feedback they received during their last evaluation experience towards the depth of information provided, quality of ideas and suggestions contained in the feedback, and timing of the feedback. The overall results of part 6 show a positive response of the respondents towards the attributes of the feedback they received during their last evaluation experience. The average score of the three items in this part in terms of 'agree' and 'strongly agree' is 48.88% On the contrary, the average score of the three items in terms of 'disagree' and 'strongly disagree' is 22.22%. The highest reported response in terms of agreement ('agrees' and 'strongly agree') is for question 25 which means 53.33% of the respondents acknowledge that information was provided in depth during feedback sessions whereas the lowest response, in the same category, is for question 27 which means 42.22 % respondents agree with the appropriateness of the timing of the feedback.

E. Describe the Attributes of the Evaluation Context

Part F seeks information about the participants' perceptions about the attributes of the evaluation context in terms of time spent on the evaluation process, time allotted for professional development during the academic year, availability of training programs and models of good practices, clarity of the policy statement, and Intended role of evaluation.

The overall results of part 7 show mixed responses of the respondents towards the attributes of the evaluation context. The average score of the five items in this part in terms of 'agree' and 'strongly agree' is 34.66% On the contrary, the average score of the three items in terms of 'disagree' and 'strongly disagree' is also 34.21% (Figure 5).

The highest reported response in terms of agreement ('agrees' and 'strongly agree') is for question 28 which means 62.22% of the respondents acknowledge that good deal of time was spent on the evaluation process including their time and that of all other participants such as students, evaluators and academic coordinators whereas the lowest response, in the same category, is for question 29 which means only 17.77% of the respondents appear to be satisfied with the amount of time allotted for professional development during the academic year.

Part G: Describe the attributes of the evaluation context.	Strongly disagree & disagree	Neither agree nor disagree	Agree & strongly agree
28	Good deal of time was spent on the evaluation process including your time and that of all other participants such as students, evaluators and academic coordinators.	15.55%	22.22%	62.22
29	Good deal of time was allotted during the academic year for professional development at the ELI.	46.66%	35.55%	17.77%
30	Good deal of training programs and models of good practices were made available.	26.66%	46.66%	26.66%
31	Policy statement regarding purpose of evaluation was very clear.	40%	31.11%	28.88%
32	Intended role of evaluation was teacher development.	42.22%	20%	37.77%

Figure 7. Attributes of the evaluation context

F. Summary of Quantitative Data

The overall quantitative data (part 2 through 6 except part 7) suggests that the teachers who participated in this case study perceived the teacher evaluation process at the ELI to be a positive experience which had a positive impact on professional practice and growth. Participants reported positive perceptions of the overall quality of evaluation, rating attributes of evaluation, the most recent observer, the attributes of the procedure used during the most recent evaluation and the attributes of the feedback received during last evaluation experience. However, Participants reported mixed perceptions of the attributes of the evaluation context. Respondents do not seem to be satisfied with the allotted time for professional development. Similarly, majority of respondents believes that the intended role of teacher evaluation is accountability rather than teacher development.

V. QUALITATIVE DATA ANALYSIS

Of the 45 respondents of the questionnaires, 9 instructors were randomly selected for scheduled one-on-one interview. Each respondent was asked 8 questions: (1) what do you perceive as the purpose of faculty evaluation process at the ECL? (2) What are the STRENGTHS of the faculty evaluation process at the ECL? (3) What are the WEAKNESSES …? (4) How has the faculty evaluation process improved your teaching practice? (5) How has the faculty evaluation process made institutional improvements? (6) What is the post-evaluation meeting like? (a) focussed on the teacher's behaviour; (b) what he/she did well or might do better or; (c) on developing the teacher's skills? (7) Regarding the evaluator: (a) Does he take adequate time to observe your performance? (b) Does he know and understand the standards and the rubrics? (8) How would you improve the evaluation process so that it became more effective for a) you, & b) for the entire ECL faculty?

	Outstanding	Above Average	Satisfactory	Needs Development	Total
Native (3)		1	1	1	3
Arab non-native (3)	1	1	1		3
Non-Arabs non-native (3)		1	1	1	3
TOTAL	1	3	3	2	9

Figure 8. Qualitative sample breakdown

A. Purpose: Accountability or Teacher Development?

Of the 9 interview participants, 7 reported their perceptions regarding the purpose of evaluation to be teacher development whereas the other two reported it to be only teacher accountability. The quantitative data does not support this finding. 42.22 % of the questionnaire respondents believed the intended purpose of the evaluation was not teacher development and 20% remained neutral whereas 37.77 % believed that it was teacher development rather than accountability.

One instructor who underwent a number of observations at the ECL during academic years 2012-13 & 2013-14, moved from lower rating (needs development) to higher rating (above average) reported:

"I believe the purpose of the AFE process is quality assurance. I improved my teaching practice after every observation. However, the AFE process was painful and almost no professional develop opportunity was offered from the ECL administration. The post observation feedback helped me reflect on my teaching. All subsequent professional development was self-directed, self-disciplined and self-sponsored".

Another participant reported:

"The evaluation process has encouraged me to reflect on my teaching which was an important step for my professional development".

Of the two who perceived the AFE process to be non-developmental, one reported:

"I don't see any benefit of these evaluations and I didn't learn anything at all. For me, every year is a probationary year".

Overall, the qualitative data suggests that teachers hold positive perceptions towards the purpose of the AFE process as 'teacher development' unlike the findings of quantitative data. According to the qualitative data, the teacher development and institutional development are taking place as joint outcome of the AFE process.

B. Effective Aspects of the AFE Process

In addition to eliciting the respondents' perceptions of the purpose of the AFEs, we elicited their perceptions of the effective aspects of the AFE process as well. The cycle of pre-observation conference, the observation itself followed by post observation conference resulting into reflective teaching and hence professional development was reported as the perceived effective aspect of the AFE process. In this respect, one participant reported:

"Teachers are becoming reflective".

Another participant admitted:

"I was under the impression that I was an outstanding teacher and did not approve of the idea of class observation for teacher evaluation. I did not see any benefit of it and saw it as a complete waste of time. I resisted and challenged the credibility of the observer. Later, observation after observation, feedback after feedback, I started thinking about my own teaching practices and had to unlearn a number of habits and even techniques which were pedagogical wrong and de-motivating and demoralizing for students. Now, unlike two years ago, I welcome any kind of feedback. I can distinguish between the valid and invalid feedback. I know one thing for sure – I need feedback". Three more participants expressed the same; they admitted that they had resented in the beginning, now they welcome feedback, however. The other participants expressed similar feelings with regards to reflective practice in their own words. Also, all of the participants reported that they had become interested in their professional development after these evaluations. 8 out of the 9 participants (88%) admitted that they had enrolled onto CELTA after the AFEs they underwent which is further confirmed by the quantitative data as well because 20 out of 45 (around 45 %) respondents earned CELTA during academic years 2012-13 & 2013-14.

C. Improvements to the AFE Process

Respondents were given the opportunity to express their perceptions to any changes that would improve the AFE process at the ELI. The dominant themes emerged were the need to train the observers and post evaluation professional development opportunities. One respondent reported:

'Having an evaluation policy, a list of standards and just one week training is not enough. I think there needs to be more training on the use of evaluation tools and how they can be used most effectively and objectively'.

A similar view emerged that:

'Observers should stay in the job for some time. We've seen too many observers joining the department of professional development and then leaving it within few months'.

A couple of participants complained that the AFE process heavily relies on only one mode of evaluation that is class observations. Some teachers can never perform well when there is an observer; they are excellent teachers, nevertheless. The alternative modes of evaluations were suggested such as Self-evaluation (Wajnryb, 1986) Peer Observations & Unseen Observations (Bailey, 2006) to encourage reflective teaching leading to professional development. All respondents expressed their serious concerns about the lack of professional development opportunities at the ELC especially in the post evaluation scenario. Teachers have reportedly taken initiatives for professional development such as earning professional certifications such as CELTA and DELTA, but the contribution on the part of the ELC is almost non-existent. Teachers want to see post-evaluation professional development opportunities facilitated and funded by the ELC administration. The quantitative data suggests that questionnaire respondents hold positive perceptions towards their most recent observer although they are not satisfied with the training and experience the observers have as per the

qualitative data. Both questionnaire respondents and the interview participants were in agreement that the ELI should facilitate post evaluation professional development opportunities before making any strategic decisions related to non-renewal of contracts or termination of employment.

VI. DISCUSSION

This study is guided by the following research questions:

1. What is the perceived purpose of faculty evaluation process at the ECL?

2. What is the perceived effectiveness of faculty evaluation process in improving teaching practice and professional development of teachers at the ECL?

3. What are the strengths and weaknesses of faculty evaluation process at the ECL?

4. Does the faculty evaluation system take account of the cultural diversity at the ECL – with a focus on what leadership and management role is played to make it happen?

A. Research Question 1

What is the perceived purpose of the faculty evaluation process at the ELI?

The quantitative data in this study suggests that teachers perceive the AFE process at the ECL as a measure for accountability rather than professional growth. In other words, it is summative in nature. On the contrary, the findings of qualitative data suggest that the perceived purpose of the evaluation is teacher development i.e. formative.

The discrepancy stems from the fact that a great deal of professional development took place among the ELC faculty during academic years 2012-13 & 2013-14. This professional development is linked to the AFE process because most teachers decided to enhance their professional knowledge and skill after the AFEs and they did achieve it by enrolling onto professional courses such as CELTA / DELTA and other short courses. So, the evaluation did instil spirit into the faculty members the passion for professional development and growth. On the contrary, all such initiatives were self-directed, self-disciplined and self-sponsored and took place during vacation period. Teachers invested their earned vacation and money and did not get any such support from the ELC. Those who perceive it as accountability perhaps mean that the ELC left them in a 'sink or swim' situation after the AFEs with little help offered. Others saw it as a factor that 'triggered professional learning'. The purpose and outcome of classroom observations, according to The ECL policy is performance improvement. It is the practice, not the policy which seems to be perceived as daunting. Teachers seem to prefer professional development as an outcome of evaluations and organizations seem to target accountability to ensure certain standards are being adhered to. Middlewood (2001) believe that both of the aforementioned purposes can certainly be achieved but there must be a context of professionalism, openness, and the involvement of both managers and staff in the evaluation process.

B. Research Question 2

What is the perceived effectiveness of faculty evaluation process in improving teaching practice and professional development of teachers at the ELI?

Danielson (2011) noted that a commitment to professional learning is important, not because teaching is of poor quality and must be fixed, rather because teaching is so difficult that we can always improve it. Similarly, Shinkfield and Stufflebeam (1995) noted that 'as professional people, teachers themselves must engage in evaluation for both professional development and accountability (p.8).The quantitative data suggests that teachers perceive that the teacher evaluation process at the ELC has had huge impact on their professional practice as well as professional growth. During interviews, 8 out of 9 participants admitted that they had enrolled onto CELTA after undergoing their summative evaluations.

C. Research Question 3

What are the strengths and weaknesses of faculty evaluation process at the ELC?

The quantitative data show several areas of perceived strengths. The quality of ideas and suggestions contained in the feedback was very high and in-depth information was provided which had reportedly positive impact on professional practice and professional growth. This is supported by experts in the field who maintain that teachers highly regard evaluators with deep knowledge of curriculum, content, and instruction who can provide suggestions for improvement (Stiggins & Duke, 1988). Also, the feedback was perceived to be very useful. The value of constructive feedback promoting improved teacher practice and professional growth cannot be underestimated (Marzano et al., 2001; Marshall, 2005). The qualitative findings also confirm that the post-evaluation feedback had positive impact on professional development of the respondents. The information provided was in-depth, that is why it was convincing enough for the respondents to invest their earned vacation and a good deal of money in pursuing professional qualifications and certification in order to enhance their professional knowledge and skill.

The areas that emerged as perceived areas of weaknesses are; little time allotted during the academic year for professional development, and the intended role of evaluation being teacher accountability rather than teacher development. The qualitative data also support this finding. Nearly all respondents were of the opinion that: (1) The ELC administration should consider facilitating on-campus professional development opportunities, and (2) and also

allow them to pursue out of the Kingdom self-sponsored professional development opportunities with a privilege of paid leave up to one month per annum so that they can avail their earned vacations and do not get burnt out.

D. Research Question 4

Does the faculty evaluation system take account of the cultural diversity at the ELC – with a focus on what leadership and management role is played to make it happen?

The ELC faculty is culturally diverse coming from 22 different countries. The quantitative data suggests that around 50% of the teachers were either never observed prior to the ELC experience or just had two or three observations altogether in the entire career. It must be intimidating and stressful for such people to undergo evaluations which are mainly meant for accountability with little opportunity for professional development. At present, the evaluation system in place does not make any difference whether a teacher is new or experienced, outstanding or struggling with teaching standards. Whatever teachers have achieved is their self-motivated and self-directed professional development. Danielson's suggested framework of three tracks can be a good starting point to improve the existing AFE process at the ELC.

Secondly, the system of teacher evaluation via classroom observation is imported from the Western society and implemented with little attention paid to cultural sensitivities. The context in question is an Asian-Arab country with 100% Muslim population. The Muslim and Asian traditions should not be ignored. In Islam, the teacher is respected like one's father. Duties and respect owed to the father are transferred to the professional site, making a detailed or challenging assessment of performance not only difficult and unlikely, but possibly anti-Islamic (Shah 1998). The observation system has not been a complete success even in the western society. One very good example is the New York City school reform movement. Shulman Sullivan and Glanz (2008) concluded that high stakes accountability leaves little room for reflective practice. The major conclusion they drew from analysis was that administrative instructional supervision as best practice was mostly absent in the schools in which the educators they surveyed worked. The dominance of Anglo-American theory, policy and practice denies or understates the influence that culture, societal culture in particular, may have on successful implementation of policy (Walker and Dimmock 2002, p. 77). The ECL is a culturally diverse workplace situated in Saudi Arabia with 100 % Islamic and Asian way of life. Any evaluation system which is imported and implemented without necessary adaptations will be seen as a threat.

VII. RECOMMENDATIONS FOR THE ELC MANAGEMENT

The ELC administration should consider giving more autonomy to teachers in deciding on evaluation system such as a five-point evaluation format suggested by Bailey (2006) which covers two extremes from greater supervisor control to greater teacher control. Only announced observations are taking place at present which clearly marginalizes the instructors because they do not have any other choice. The implementation of Negotiated Observations, Invited Observations and Unseen Observations could increase teacher control, hence – more motivation and less teacher stress. It is a fact that there are always some very good teachers who cannot perform well when it comes to formal observations. Similarly, there are teachers who are willing to improve, but formal observations stress them out. They are not willing to share their weaknesses with anyone and eventually are never able to unlearn or relearn. Invited Observations or even Unseen Observations might be very helpful for such introvert teachers.

Whereas the ELC administration should also consider introducing a multitier evaluation program that considers years of experience and teacher status such as the three-track system for teacher evaluation as proposed by Danielson & McGreal, (2000). Track 1 for Beginning Teachers: evaluations in this track will be summative in nature, but once the probationary period of a year or two is over, successful teachers should not suffer from the lack of job security at all. Track 2 for Professional Development: for existing teachers on the basis of their ongoing evaluations and development needs. Track two observations /evaluations should be strictly formative in nature and the teachers under observation/evaluation should be provided ample opportunities to develop. Enough time and budget should be allocated to this track and it is envisaged that the majority of the faculty will fall in this category. Finally, Track 3 Post Evaluation Teacher Assistance: for underperforming teachers. Once again, track three evaluations will be summative in nature – make or break decisions will be made in this Track. Firing decisions should be made either after track 1 when a new teacher fails to fulfill the required conditions or after track 3 when a teacher is offered enough assistance but he does not show any signs of improvement.

REFERENCES

[1]   Bailey, K. (2001). Observations in Carter, R. and Nunan, D. (eds.) *The Cambridge Guide to Teaching English to Speakers of Other Languages*. Cambridge: Cambridge University Press, 114-119.
[2]   Bailey, K. (2006). Language Teacher Supervision. Cambridge: Cambridge University Press.
[3]   Boyd, R. T. C. (1989). Improving teacher evaluations. *Practical assessment, Research & Education,* 1(7) (http://PAREonline.net/getvn.asp?v=1&n=7) Online accessed Feb 10, 2014.
[4]   Bolam, R. (2002). 'Professional Development and Professionalism', in Bush T. and Bell, L. (eds.), *The Principles and Practice of Educational Management*. London: Paul Chapman Publishing.

[5]   Danielson, C. and McGreal, T. L. (2000). Teacher Evaluation to Enhance Professional Practice. Princeton, NJ: Educational Testing Service.

[6]   McGreal, T. (1983). Successful teacher evaluation. Alaxandria, VA: Association of Supervision and Curriculum Development.

[7]   Marshall, M. (2005). It's time to rethink teacher supervision and evaluation. *Phi Delta Kappan*, 86, 727-730.

[8]   Marzano, R., Pickering, D., and Pollock, J. (2001). Classroom instruction that works: Research-based strategies for increasing student achievement. Alexandria, VA: Association for Supervision and Curriculum Development.

[9]   Marzano, R. (2012). Teacher Evaluation: What's Fair? What's Effective? *Educational Leadership* 70. 3, 14-19.

[10]   Middlewood, D. (2001). Appraisal and Performance in the UK, in Middlewood, D. and Cardno, C. (eds.) *Managing teacher Appraisal and Performance: A Comparative Approach.* London: RoutledgeFalmer.

[11]   Powel, Garry. (1999). How to avoid being the fly on the wall. *The Teacher Trainer*, 13.1, 3-4.

[12]   Quirck, Phil. (1996). Using unseen observations for an IST development program. *The Teacher Trainer* 10.1, 18-20.

[13]   Shah, S. (1998). Gender perspectives in principalships in Pakistan. Unpublished PhD thesis, University of Nottingham.

[14]   Sheal, Peter. (1989). Classroom observation: Training the observers. *English Language Teaching Journal* 43.2, 92-103.

[15]   Shinkfield. J., and Stufflebeam, L. (1995). Teacher Evaluation: Guide to effective practice. Boston, MA: Kluwer Academic Publishers.

[16]   Shulman, V., Sullivan, S., and Glanz, J. (2008). The New York City school reform: consequences for supervision of instructions. *International Journal of Leadership in Education* 11, 407-225.

[17]   Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Brooklyn, NY: New Teacher Project. Retrieved from (http://widgeteffect.org/downloads/TheWidgetEffect.pdf) (accessed 15/01/ 2014).

[18]   Wood, C. and Pohland, P. (1983). 'Teacher Evaluation and the "hand of history". *Journal of Educational Administration* 21.2, 169-181.

Asif Ahmad is a lecturer in English at King Abdul Aziz University, Kingdom of Saudi Arabia. He earned an MA in English Language & Literature (1994) from Gomal University, Pakistan; an MA in TEFL (2000) from AIO University, Pakistan; an MSc in Educational Leadership and Management (2014) from University of Leicester, UK. Currently he is enrolled onto a PhD program at University of Leicester, UK. Mr. Ahmad's research interest includes language teacher identity, teacher accountability and professional development.

Jabreel Asghar is a faculty member in the English Language Institute at King Abdulziz University, Saudi Arabia. He acquired his MA in ELT and doctorate in Applied Linguistics from the University of Warwick. He has taught in the British, Asian and Middle Eastern higher education sector as an educator, teacher educator, material developer and in leadership roles. His research interests include critical discourse analysis, language and identity, academic discourse, educational leadership and material development.