Effect of Two Alternative Self-reflection Strategies on Developing EFL Students’ Speaking Ability and Their Perceived Self-efficacy

Dr. Shayma’a A. Torky
National Ctre for Educational Research & Development
A.R.E.

ABSTRACT

Two strategies investigating the benefit of self reflection in developing speaking performance and speaking self efficacy are explored. The first strategy - employing a set of criteria as a basis for self assessment is compared with a reformulation based strategy. This is in sharp contrast with more typical studies which compare an approach using self reflection technique with an approach lacking this perspective. Both strategies include giving oral presentations, transcribing, self assessment, repetition and further reflection. Forty nine female students from College of Science at Kuwait University, comprising two experimental groups and one control group participated in the current study. The experimental groups were required to give three presentations followed by a cycle of either criteria led or reformulation-based self reflection techniques. The control group, however, was required to give the same presentations without any self-reflection instruction. Results confirm that both strategies proved effective in enhancing students’ oral performance as well as their self-efficacy. Nevertheless, while the criteria-based strategy could be effective as far as accuracy is concerned, reformulation-based strategy was more conducive when it comes to enhancing language complexity. Moreover, both strategies proved to hold promise regarding enhancing students’ speaking self-efficacy.

Introduction

Speaking instruction- as typically addressed by the communicative approach- has typically focused on exposing EFL learners to a plethora of purposeful authentic situations to stimulate the use of whatever language
repertoire they possess. Nonetheless, it has been noticed that under time pressure students allocate most of their attentional capacities to conveying the message, and in so doing they tend to pay no heed to the accuracy or complexity of what they are delivering, even if signals of misunderstanding or "recasts" are directed to them (Lynch, 2007: 311). Inevitably, this leads to the prolonged use of faulty forms or simplified structures that become fossilized or resistant to change and which adversely influence EFL learners' language proficiency development (Lynch, 2001). For that reason, the need for reappraising how EFL speaking proficiency is perceived and pedagogically tackled has increasingly grown.

One trend that has emerged to combat fossilization and enhance ESL/ EFL learners’ use of accurate and complex language is self-reflection. It is an approach which emphasizes the importance of redirecting learners’ attention to features that almost go unnoticed during production (Lynch, 2001). Self-reflection is underpinned by other concepts which have come into view recently, such as noticing, self-consciousness or awareness raising. All of which attempt to unveil and hence mobilize tacit cognitive mechanisms claimed to play a pivotal role in fostering SL/ FL language acquisition (Thornbury, 1997: 326).

Several empirical attempts in the domain of teaching ESL writing, reading and listening have been undertaken to get clearer insight into the impact of self-reflection (Riggenbach, 1990: 153). However, the time-bound nature of speaking has always been a stumbling block to embracing self-reflection practices in speaking instruction (Christianson and Hoskins, 2008). Nevertheless, some case studies have underlined the importance of encouraging EFL learners’ to scrutinize their own language use during or subsequent to oral performance in order to bring about a real change in their evolving inter-language system (Lynch, 2007; Patri, 2002; Mennim, 2003; and Stillwell et al., 2010).

Considerably, it was claimed that self-reflection can help students gain more confidence over the goal—means relation, that is, they will possess higher self-efficacy (Chan & Lam, 2010). Self-efficacy in turn is a motivational construct which is thought to help students invest more effort and persist, especially when they face setbacks during learning (Bandura, 1999). Although, a considerable degree of emphasis has recently been placed on investigating the role self-efficacy in foreign
language learning in general (Mahyuddin et al., 2006 & Idrus and Sivapalan, 2007), research on how to promote EFL speaking self-efficacy is still lacking.

**Context of the Problem**

One of the main goals of the general English Language Program at the College of Science, Kuwait University, is to develop the speaking skills necessary for participating effectively in future academic settings. In other words, the program is intended to raise EFL university students’ oral proficiency to cope with university studies. In particular, the program aims at enabling students to introduce topics, make presentations, and analyze opinions. Nonetheless, the oral activities currently utilized to pursue these goals-as proven by a pilot study- have fallen short of enabling students to achieve tangible progress, which is manifested in their inadequate grammar, lack of organization, limited vocabulary, and the tendency to use rather very basic utterances devoid of any sort of elaboration. This can be attributed, in part, to the prevailing teaching methods adopted that undermine the importance of raising learners’ consciousness of their own oral performance. Thus, for EFL students, speaking proficiency remains such an ambiguous entity assumed to be impossible to grasp, let alone master.

In addition, results of previous EFL research in the Arab milieu, such as the studies of Ghanem (1983), Hussein (1986), Al-Khuli (2000), and Torky (2006) consistently emphasized that EFL students cannot attain an adequate level of speaking proficiency. This was mainly attributed to the limited evaluation system, students’ inhibition, and lack of self-assessment opportunities.

The previous considerations triggered the researcher to conduct the current study as an attempt to examine the impact of incorporating self-reflection practices on developing university EFL students’ oral performance and perceived self-efficacy.

**Statement of the Problem**

The study problem could be identified as EFL university students inadequate spoken performance, which was partly attributed to the regular teaching methods that barely grant students the chance to reflect on their oral performance and devise strategies for self-improvement. The study revolved around the following main question:
"How will students’ oral performance differ if exposed to one of two self-reflection strategies: self-reflection against a set of criteria and self-reflection against a reformulated version of their output?

The following research questions were derived from the main question:

1 - To what extent will the two suggested self-reflection strategies have a differential effect on students’ speaking performance as measured by the performance measures of accuracy and complexity in the short term?

2 - To what extent will both strategies have a differential effect on students’ speaking perceived self-efficacy?

Purpose of the Study

The study aimed at comparing the effectiveness of two self-reflection strategies in developing EFL students speaking performance. Both strategies followed a sequence that commenced with students’ noticing of their own performance and ended with students’ production of an enhanced version of the original task. Yet, the two strategies adopted two distinct techniques: criteria-led versus reformulation-led self-assessment. The study thus intended to demonstrate how oral performance measures such as accuracy and complexity could be distinctly targeted using the suggested strategies. Furthermore, the study provided insights into how self-reflection oriented instruction could bear upon FL speaking self-efficacy- a construct conducive to foreign language learning.

Study Hypotheses

The following null hypotheses were stated:

- There are no statistically significant differences between the mean scores of the first experimental (criteria-led) group as well as second experimental (reformulation-led) group on the first performance and repeat one of the three presentations.

- There are no statistically significant differences between the mean scores of the first experimental group as well as the mean scores of the second experimental group on the pretest and posttest on the speaking performance measures.
- There are no statistically significant differences between the mean scores of the first and second experimental groups on the post test on the speaking performance measures.

- There are no statistically significant differences between the mean scores of the first as well as the second experimental group on the pretest and post-test of the speaking self-efficacy scale.

- There are no statistically significant differences between the mean scores of the first and second experimental groups on the posttest of the speaking self-efficacy scale.

**Delimitations of the Study**

This study was confined to:

1 - Two performance measures (accuracy and complexity). Two sub-measures were chosen to assess accuracy and five for assessing complexity, two for syntactic complexity and three for lexical complexity. Fluency was not tackled in the current study since it was thought that a limited time treatment as such can barely provide real change in fluency which would require self-observation over a longer period of time. In addition, it was thought that enhancing accuracy and complexity would exhaust most of the students’ attention that focus on fluency would be better postponed after some gains could be achieved in accuracy and complexity to avoid fossilization.

2 - A sample of forty nine first year low-intermediate EFL university students, College of Science, Kuwait University.

3 - Transactional monologic oral production. Thus, other interactional conversational genres were not addressed in the current study.

4 - A limited duration for implementing the treatment (one semester, i.e., nearly three months).

**Definition of Terms**

The operational definitions of terms as utilized in the current study are as follows:

**Self-reflection:** in the current study, self-reflection is defined as improving students’ capacity to recognize their performance during and subsequent to speaking so that they become able to determine how distant
their actual performance is from the level they had to achieve, and so they
can act accordingly (Riggenbach, 1990). In the current study, criteria-led
and reformulation-led self-reflection were incorporated.

**Speaking performance measures:** These are measures that provide
counts of specific linguistic features occurring in the spoken discourse.
These include fluency, accuracy and complexity (Ellis, 2005: 166). In the
current study only measures of accuracy and complexity were ad-
dressed.

**Speaking self-efficacy:** It refers to individuals’ beliefs in their ability
to organize and execute the courses of action required to speak the
target language. It includes three dimensions: ability perception, activity
perception and aspiration (Bandura, 1999).

**Theoretical Framework and Background**

**(Effect of Self-reflection on speaking Performance and Self-Efficacy)**

In the following section, it will be argued that self-reflection can be
incorporated as an instructional method to help students develop their
spoken performance. This section is further divided into four sub-
sections that set the framework for the present study. The first sub-
section gives an overview of two quite distinct methods of assessing
EFL spoken performance-rating scales and performance measures- and
provides a rationale for adopting performance measures in the present
study. The second sub-section highlights the definition and instructional
implications of adopting self-reflection to develop oral performance. It
describes two techniques that contribute to enhancing EFL learners’
self-awareness of their oral performance, transcription and task repeti-
tion. The third sub-section sets the assumptions upon which the
comparison between the two self-reflection strategies adopted in the
current study, namely criteria-led and reformulation-led self-reflection-
are based. The fourth sub-section deals with the concept of speaking
self-efficacy and pinpoints how it can be pedagogically addressed
through incorporating self-reflection practices.

**I- Assessing Speaking Performance**

To assess speaking performance, two methods adopting entirely
distinct perspectives have been incorporated: rating scales and perfor-
manace measures. On the one hand, rating scales describe qualitatively
various levels of spoken ability along a continuum ranging from relatively inadequate to adequate performance (Foster, Tonkyn, & Wigglesworth, 2000). Rating scales measure language aspects such as the use of correct grammar, comprehensibility, pronunciation, pragmatic competence, etc. Mostly, it is the discretion raters must use in assigning a score to students’ oral performance; this raises some concerns about possible inconsistencies between raters or subjectivity of judgment (North, 2000: 206).

On the other hand, performance measures- sometimes called discourse analysis measures- are methods for assessing overall language proficiency, rather than a certain language skill in particular. The key performance measures are fluency, accuracy and complexity. These measures attempt to quantify aspects of spoken or written language by providing counts of certain features in learners’ production to facilitate comparison of different performances (Koizumi, 2005, Ellis, 2005 and Inoue, 2010). For instance, the number of subordinate clauses is considered a measure of written and spoken output complexity (Koizumi, 2005).

Performance measures are often assessed using ratio scales (e.g., number of correct AS units, a single speakers utterance, per total AS units to measure accuracy) (Inoue, 2010). Therefore, unlike rating scales, discourse analytic measures sound more objective, as little disagreement can ensue among raters (Riggenbach, 1990: 153).

With this in mind, both rating scales and performance measures can be perceived of as complementary rather than contradictory views of speaking assessment, as they integrate the qualitative and quantitative aspects of language performance. However, for the sake of practicality and preciseness, the current study limited its scope to two performance measures as indictors of spoken proficiency, namely accuracy and complexity. Fluency was not tackled in the current study to provide ample chance for students to hone and expand their linguistic performance without being distracted by enhancing speed of delivery.

II-Self Reflection

Self-reflection is mainly substantiated by four closely intertwined SLA conceptual foundations: the output hypothesis, the noticing theory, language related episodes and peer interaction (Gass & Mackey, 2007). First, Swain (2005) gives precedence to the learners output, since,
according to his viewpoint, a learner’s written or spoken performance can provide multiple chances of deeper language processing compared to those provided by the input he/she receives. The output helps learners to notice or be cognizant of the gaps in their inter-language system or make comparison between their current linguistic level and the target performance (Ellis, 2005), so that they will be motivated to put some effort into bridging these perceived gaps. Noticing of mismatch in ones language can be manifested in a set of Language Related Episodes (LRE) - a concept which further supports self-reflection (Ibarrola, 2009). According to this concept, students have to engage in negotiating the adequacy of their output, which pushes them to consciously practice what is called “meta-linguistic reflection” or talking about their language use including points of strength and weakness (Mennim, 2012: 56). Therefore, the fourth construct, peer interaction, appears to play a paramount role in mediating noticing, meta-linguistic reflection, bridging the gap in the inter-language system and subsequently long term language acquisition (Tocalli-Beller & Swain, 2005).

As far as speaking instruction is concerned, self-reflection implies handling speaking from a process rather than a product perspective where concepts like drafting, editing and redrafting are implemented to bring about a real change in the learners performance (Howarth, 2001). In essence, self-reflection can serve as one of the best instructional strategies for promoting spoken proficiency, given the observed inefficiency of the teachers feedback whether in the form of recasts, overt correction or repetition in altering learners inter-language system (Riggenbach, 1990: 159).

Techniques sustaining self-reflection can range from those inducing indirect noticing of one’s own performance, such as task repetition (Finardi, 2008; Hawkes, 2009 and Ahmadian, 2011), and transcribing (Lynch, 2007, Mennim, 2012) to more overt techniques, such as checklists, learners diaries, self-ratings, etc. (Lim, 2007). In the following section the techniques of transcribing and repetition are investigated in depth.

A. Transcribing:

Transcribing is a technique that enables learners to make their spoken production visible and tangible by listening carefully to their own recorded performance and writing exactly what they hear, including
pauses, dysfluncies (interruption or repetition), and errors. It was termed "proof listening" and was thought to allow students to subject their performance to close scrutiny or monitoring (Lynch, 2001: 125). Inevitably, this monitoring facility-unleashed by transcribing- can hardly be benefited from during the stressful experience of real-time production that consumes the speakers almost all attentional resources and renders various language aspects entirely unnoticed (Cappi, 2011). Despite being a time consuming activity, however, it was argued that transcribing can have a dramatic effect on improving learners future oral performance (Stillwell et al., 2010). Transcribing helps learners check their hypotheses about the target language, notice points of strength and weakness and set goals for bridging the gaps in their inter-language system (Mennim, 2012; Stillwell et al., 2010). If required to modify their performance, students can also be asked to transcribe both their original spoken output as well as the modified one to trace development (Stillwell et al., 2010). Transcribing can be carried out by the teacher, yet students transcription is argued to be more influential (Lynch, 2007).

Alternatively, although transcribing is typically viewed as a technique that sustains mainly incidental noticing, the current study postulates that transcribing would fall short of improving speaking skills unless followed by more structured activities that mold students' seemingly intuitive remarks. Such activities can take the form of a detailed self-report by the students on their self-perceived difficulties and aspects of improvement (Christianson and Hoskins, 2008). Another suggestion is to conduct teacher-student conferencing to provide further corrections on points that students missed and guide their future plans (Mennim, 2003). Students can also be provided with a model performance or a set of detailed criteria to guide their error correction endeavor (see section III).

B. Task repetition

Howarth (2001: 43) suggests that repetition can be dealt with as one component of speaking process-oriented instruction adopting a plan perform analyze repeat sequence.

Even though research suggests that task repetition positively affects oral performance on the same task (Bygate, 2001; Finardi, 2008; Hawkes, 2009& Ahmadian, 2011), only few studies could show how this effect extends to unseen novel tasks. For instance, Matsumura, Kawamura and
Affricano (2008) tracked the long-term changes in the productive skills brought about by repeating either a narration or a decision making task. Results show that, in the long term, learners gained in accuracy and certain aspects of fluency but not in complexity. Likewise, Ahmadian (2011) revealed that “massed”, or frequent, repetitions of the same task assisted students to achieve long term gain in complexity and fluency but not in accuracy on new tasks. Thus, results of both studies can be described as contradictory, each showing that repetition influences different aspects of language performance on new tasks.

In the present study, instead of examining the effect of repetition in isolation as previous studies did, repetition was viewed as a componential element in overall self-reflection cycles comprising other techniques. Repetition was utilized both as a noticing triggering device to improve long term oral performance and as a means of assessing oral performance short term gain.

**III-Criteria versus Reformulation-Led Self-assessment**

Criteria-based self-assessment implies encouraging students to grade their own performance or transcripts according to a set of standards or a checklist designed in accordance with certain criteria (Patri, 2002 & Lim, 2007). Leger (2009) conducted a longitudinal study of French students who self-assessed their speaking skills with reference to self-rating scales, including vocabulary, fluency, turn taking, etc., and set learning goals accordingly. It was concluded that students’ self-perception evolved particularly in relation to fluency, vocabulary, and overall speaking confidence. Riggenbach (1990, 158) sets forth important features to be used as targets of self-assessment, such as organization, content, delivery, grammar and vocabulary. Similarly, Christianson and Hoskins (2008) investigated the use of self-analysis sheets subsequent to self-recording of oral performance produced by first year university EFL students. Alternatively, Stillwell et al. (2010) stimulated EFL advanced students to evaluate their spoken performance using the simplified criteria of fluency, accuracy and complexity.

Noticeably, the criteria adopted by both Christianson and Hoskins (2008) as well as Stillwell et al. (2010) are rather vague. In particular, the former study required students to answer a few yes/no questions, utilizing very limited performance indicators that fall short of stimul-
ing in depth self-observation. Alternatively, the latter immersed students in the labor intensive task of counting words and pauses without providing them with clear interpretation of why and how these measures can account for their oral performance. Taking into consideration the previous pitfalls, the current study presented clear criteria in the form of a detailed analysis tool handling grammar, vocabulary, organization and content. Students were also informed of how these criteria can bear on measures of accuracy and complexity.

Another means of inducing self-assessment is reformulation which was claimed to predispose students “to notice features of the modeled behaviors that they themselves had found problematic in the initial trial run” (or first draft) (Thornbury, 1997: 328). Reformulation derives its origin from the Community Language Learning method (Richards & Rodgers, 2001). According to Thornbury (1997: 327), reformulation within a speaking instruction milieu comprises recasting what the students say spontaneously in a manner that resembles as much as possible the target language, so that students can compare their "flawed" performance with this recast model (Sachs and Polio, 2007: 71). In this way, various individual linguistic needs can be well catered for (Adams, 2003). Evidently, the reconstructed output presents students with cognitive conflicts and hence initiates self-criticism, gap noticing and subsequent long term improvement (Tocalli-Beller and Swain, 2005).

Although reformulation has been used to address EFL/ESL writing skills (Lynch, 2001; Sachs and Polio, 2007; Santos et al., 2010; Rahim and Riasati, 2011), using a noticing triggering device as such to improve speaking performance is looked at with a lot of suspicion. This might be due to the fact that unlike writing which is thought of as a permanent record of language performance, speaking is considered a spontaneous or on-the-spot activity that can hardly lend itself to retrospective evaluation. Nonetheless, in their study, Christianson and Hoskins, (2008) argue that teachers can create recorded examples of what effective speaking is like to help students reflect on their oral performance.

Only one experimental study examined the effect of using reformulation on developing pronunciation (Smith and Beckmann, 2005). In this study, students read and recorded a provided text. They were then provided with a native speaker recording of the same text and were
asked to analyze it for some targeted phonetic features. Results showed that students felt their pronunciation had improved and continued to use the strategies of noticing and reformulation later on.

IV- Self Efficacy

Perceived self-efficacy is an individual’s belief in her/his capabilities to carry out the necessary steps to realize a certain goal (Bandura, 1999: 286). Self-efficacy is generally defined as a domain-specific performance-oriented construct; hence measures of self-efficacy are more accurate when they gauge specific, rather than general ability beliefs (Bandura, 1999). Self-efficacy can provide a foundation for motivation and achievement (Huang and Chang, 1996; Mahyuddin et al., 2006; Schunk & Pajares, 2002). That is to say, if individuals believe that their actions will produce an outcome they desire, they will be able to overcome problems and remained unaffected by threats which weaken those with low self-efficacy (Yough, 2011). Basically, perceived self-efficacy can be positively affected by students mastery experiences, positive feedback, observing others doing the same task and personal emotions (Idrus and Sivapalan, 2007). Furthermore, research to date suggests that self-efficacy is positively associated with language proficiency, strategy use, internal attributions of ability, and negatively associated with communication anxiety (Hsieh and Kang, 2010). Inevitably, also, self-efficacy in a particular language skill is significantly related to proficiency in this skill; research could substantiate this relation in reading (Mills et al., 2006 & Shang, 2010), writing (Klassen, 2002) and listening (Rahimi & Abedini, 2009).

Self-efficacy bears linkage to Gardener’s concept of intrapersonal intelligence (Gardner, 1993, 2006). From one perspective, it is argued that the concept of self-efficacy is one component in Garden’s concept of intrapersonal intelligence, since without being aware of their own self, learners will not be able to build confidence in their abilities or have a high sense of self- efficacy (Shepard et al., 1999). From another perspective, research has proven that there is a mutual relationship between self-efficacy and intrapersonal intelligence in that individuals high in intrapersonal intelligence are likely to be more self-confident, and intrinsically motivated (Yough, 2011). On the other hand, it was argued that the higher the perceived self-efficacy, the higher the learner’s
motivation and the better he will be able to activate intrapersonal intelligence or put it into action by moderating motivation and skills to reach his goals (Pajares, Johnson, & Usher 2007 and Sellars, 2010).

In spite of these similarities between the concepts of self-efficacy and intra-personal intelligence, there are also some discrepancies. Notably, self-efficacy is perceived more as an acquired evolving cognitive facility that can be subjected to change in the short term by self-reflection practices, a positive learning atmosphere, and reduced level of risk (Schunk & Pajeres, 2002). On the other hand, even though it is argued that students’ intra-personal intelligence can potentially be stretched, there is a predominant trend that alludes to the fact that intrapersonal intelligence development is a long term process and that educational intervention has to be tailored to students relatively fixed intra-personal profile (Sellars, 2010: 54). In addition, in contrast with intra-personal intelligence, self-efficacy theories are accused of falling short of embracing an encompassing view of the cognitive processes underlying learning. Nonetheless, it is claimed that self-efficacy has a more immediate influential role in shaping students’ thinking and hence academic achievement compared to the role played by intrapersonal intelligence (Bandura, 1999).

Observably, only little research could be found on how to boost students’ self-efficacy through manipulating the teaching and evaluation conditions in academic and EFL classrooms. For instance, Schunk & Pajeres (2002) put much emphasis on the role of self-reflection in promoting self-efficacy. Likewise, Chan and Lam (2010) found that EFL students experienced less of a decrease in self-efficacy if they received formative self-referenced feedback, rather than summative norm referenced feedback. It is also argued that self-reflection which fosters both goal-setting and proactive use of strategies enable learners to attribute success or failure to their strategy use, rather than factors outside their control (Leger, 2009).

* From the preceding literature review, the following points can be concluded:

Self-reflection, in EFL speaking instruction contexts, can be conceptualized as an overarching concept comprising a variety of techniques all aiming at raising students’ awareness of their own oral performance. Some techniques to have self-reflection include repetition, transcribing
and self-correction which help learners put more conscious effort in bridging current linguistic gaps and hence achieve long term linguistic gain. Study results imply that self-reflection against a reconstructed version of a learner’s own output can be more effective than self-reflection against a set of norms. Moreover, self-reflection inducing goal setting can play a critical role in augmenting EFL learners’ sense of efficacy or confidence in their language ability. Nevertheless, there are issues that were not addressed adequately by previous research as follows:

* Though few studies were carried out on the effect of criteria-led self-assessment, no single study exists which adequately handles the technique of reformulation as a self-assessment tool for promoting oral performance. In all previous studies, except for Lynch (2007) - which is a case study-scarce reference, was made to the procedures of reformulation. The study of Smith and Beckmann (2005) was solely focused on pronunciation and hence reformulation was conceptualized as providing students with an original listening text, rather than a reconstructed version of learners own performance. Thus, the current study attempted to operationalize reformulation as a self-assessment tool to compare its effectiveness to criteria-based self-assessment in developing EFL speaking performance.

* Most of the previous studies measured the short term effect of transcription and self-noticing as manifested by students’ repeated performance of the same task. Therefore, it cannot be inferred whether such self-reflection tasks might have a cumulative effect on inter-language development. Accordingly, the current study attempted to measure the long-term effect of self-reflection on students oral performance.

* The research to date has tended to focus on one or two self-reflection techniques at a time and no clear model was suggested that views self-reflection as an integrative multi-stage process. Thus, the current study sought to incorporate several techniques into one strategy (self-reflection) that commences with oral performance, proceeds to transcribing and self-correction and eventually gets to the stage of making informative decisions for further improvement.
* No studies to date have measured the effect of self-reflection on EFL speaking self-efficacy; thus further research is warranted in that domain.

Methodology

Design and sample of the study:

The study utilized a quasi-experimental nonequivalent pretest and posttest comparative group design including one control and two separate experimental groups. This design is most effective for this type of research since all groups were intact classrooms that could not be divided and separated for random assignment.

The participants were forty nine first year low-intermediate EFL female university students taking an intensive general English course at Kuwait University, College of Science, to enable them to pursue their degree courses. Regularly, students have ten hours of general English teaching each week, following a new general English syllabus called “Q: Skills for Success”, Book 2 (Craven et al., 2010). The current syllabus gives special emphasis to presentation skills; thus, it was decided that these skills can easily be adapted to include the suggested self-reflection component.

Students were between the ages of eighteen and twenty. The sample comprised three classes representing two experimental groups both including seventeen students and one control group of fifteen students. They all had been learning English as a foreign language at Kuwait schools for twelve years. Therefore, they could be considered a fairly homogenous group in terms of their English proficiency. In addition, the three groups were considered linguistically equivalent as determined by a placement test administered at the beginning of the study. Moreover, students’ scores on the pre-test indicated no statistically significant differences among them in terms of speaking measures and self-efficacy, so any gain could be attributed to participating in the current study.

The first experimental group was exposed to the criteria-led strategy, while the second group was exposed to the reformulation-led strategy. Both experimental groups were taught by the researcher while the control group was instructed by another teacher, a Ph.D. holder,
under the researcher supervision, following the regular method. Another teacher taught the control group because the treatment did not require a great deal of expertise on the part of the teacher; instead most of the effort was exerted by the students who selected and gave PowerPoint presentations subsequent to each unit.

**Instruments:**

The current study utilized the following tools:

1. A speaking performance measures checklist.
2. A pre-posttest speaking test (designed by the researcher).
3. The self-efficacy scale for learning speaking (modified by the researcher).

1 - The speaking performance measures checklist:

The checklist was designed to determine the most important speaking performance measures that suit the purpose of the current study. The performance measures included in the checklist in its primary form, were determined through reviewing literature and empirical studies pertinent to assessing language performance in general and speaking performance in particular such as Ahmadian (2011), Inoue, (2010), Koizumi, (2005), Lynch, (2007), and Stillwell et al., (2010). The current study adopted AS units for analyzing students speech. An AS unit is defined as “a single speakers utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (Foster et al., 2000: 365).

The checklist was composed of 19 speaking performance measures subsumed under the headings of accuracy and complexity. A seven-member jury comprising specialists in the field of TEFL were asked to determine the most important and feasible measures for the present study. The jury members also were asked to choose measures which lent themselves to a single interpretation in order to minimize disagreement among raters. According to their advice, seven performance measures were selected to be the focus of the current study as follows:

**Accuracy:**

1. A1 = No. of error-free verb forms per total verb forms (i.e. tense, modality, and subject -verb agreement).
2 - A2 = No. of error-free AS-units per total AS-units (including grammatical and lexical errors).
   Complexity:
   a. Syntactic complexity:
3 - S1 = Subordination- No. of subordinate clauses per total AS-units
4 - S2 = The mean length of AS units- No. of tokens (words) per total AS-units
   b. lexical complexity:
5 - L1 = Lexical density- No. of lexical tokens (words) per No. of total tokens.
6 - L2 = lexical diversity (Giraud index) - No. of types (different words) divided by the square root of total No. of tokens.
7 - L3 = lexical sophistication- No. of sophisticated word types per token.

2 - The speaking test:

To measure whether students’ short term progress brought about by self-reflection practices could transfer to new contexts, a pre-posttest was designed to assess students’ speaking performance. The test consisted of two main sections. The first focused mainly on exchanging personal information to set the stage for the next task. The second section, in turn, was “a telling story” task assigned after giving students ten minutes to plan for their performance. The students were required to recount a happy or sad event that happened to them or to someone else they knew.

After test administration, the researcher transcribed students recordings on the pre and posttests. Then, two versions of the transcripts were produced. The first was segmented into AS units to calculate accuracy and subordination. The second transcribed version was entered into the Vocabprofile program (Cobb, 2002)- a text analysis program based on Laufer’s Lexical Frequency Profile (LFP) (Laufer and Nation, 1995), which calculates the proportion of word types belonging to the first 1,000 most frequent words, the second 1,000 most frequent words, and off -list words. For the purpose of the current study, all words beyond the first 1000 were considered quite sophisticated for the targeted students. Counts of words, types, lexical density,
and sophisticated words were also obtained through the Vocabprofile program. Scoring did not require another rater as all test measures were objective. However, a sample of 30% of the total transcripts was checked by two other raters-who received training- and inter-rater reliability was calculated. Cronbach Alpha results showed the correlation coefficient ranged from 0.80 to 0.92 on all performance measures. Furthermore, intra-rater reliability was calculated through repeating the correction process by the researcher after one month and calculating the correlation between the two ratings. Cronbach Alpha results proved to be adequate, ranging from 0.75 to 0.90 for almost all spoken samples. Moreover, content validity was obtained through submitting the test to a panel of jury of 10 TEFL specialists.

3- The Self- Efficacy Scale:

This scale aimed at determining the impact of the two suggested self-reflection treatments on dimensions of self-efficacy. A twenty-seven item survey questionnaire was adapted from questionnaires by Idrus and Sivapalan (2007), Rahimi and Abedini (2009), and Yough (2011); however, the researcher modified the scale to suit the study purpose by adding, modifying and deleting some items. Substantially, the questionnaire comprised three dimensions. The first included eleven items assessing students perception of their speaking competence. This dimension was labeled "ability perception". The second dimension, “activity perception”, included eight items measuring students attitudes towards oral activities, such as role play, oral presentations and in-class discussion. The third dimension- aspiration- comprised eight items, measuring students future expectations and goals relevant to speaking proficiency. To obtain Crohnbach Alpha reliability estimates, the internal consistency of the entire scale was measured. The alpha level of the three factors proven to be high ranging from 0.75 to 0.81.

Items comprising the first dimension were assigned a score of 1 to 5, 5 assigned to the “Absolutely confident” and 1 to the “not at all confident” response. Activity perception and aspiration items were also given a score of 1 to 5, with 5 assigned to “strongly agree” and 1 to “strongly disagree”.

The suggested self-reflection strategies

The two suggested self-reflection strategies can be considered a sort of formative evaluation that helped students focus on the process rather
than the product of their spoken performance. In other words, a process-oriented approach to speaking was adopted. Prior to the treatment, students in both experimental groups—over five introductory sessions—were exposed to training on the transcription process. Subsequently, students were engaged in a self-reflection cycle as follows:

1 - Power point presentations and recording. First, throughout the semester, students in both experimental and control groups were assigned seven main topics; comprising subtopics pertinent to the units they had studied in their textbook (Q: Skills for Success, Book 2)—to choose three of their interest throughout the treatment. Experimental groups were asked to give and record their PowerPoint presentations in class utilizing the recording facility on their cell phones. A backup was recorded on the teacher’s cell phone. The presentations’ topics were as follows:

Presentation 1:

a. Do students have to wear school uniforms? OR b. Describe a school uniform design.

Presentation 2:

a. My dream building; b. Favorite colors of my family members; OR c. Animals that use camouflage

Presentation 3:

a. Bad manners (in school, for children, at table, while driving.) OR b. Describing a process such as (cooking a meal, fixing up a machine, etc.)

2 - Transcribing. Students in both experimental groups were requested to complete the initial transcript of the first presentation. The teacher had to check its consistency with the audio recording; noticeable discrepancies were discussed with the students. Luckily, the time element was not problematic as students did most of the time consuming task—transcription—at home.

3 - Self correction (distinct for both experimental groups). To sustain incidental learning, students— in both experimental groups—were asked to look freely at their performance on each of the three
presentations and express their general perspective. After that, students had to assess their performance after being exposed to one of two distinct treatments as follows:

3 - 1 - Criteria-led strategy

Students were handed a checklist of a set of criteria in light of which they had to determine how much they were satisfied with their oral production. They had to give their answers on a three point rating scale (yes, no, to some extent). If the answer was “no” or “to some extent”, they edited their mistakes or added modifications. Students were first guided to focus on the content and elaboration of their ideas; then, they were directed to address grammatical and vocabulary mistakes. Subsequently, a student-teacher conference was held to induce students to rethink the parts of the transcript that were linguistically inadequate. The outcome of this revision was a modified transcript (2).

3 - 2 - Reformulation strategy

a. Teacher reformulation. Students were given a reconstructed text-modified by the teacher of their original presentation. The reformulated version maintained the students’ original ideas. However, it reconstructed the content in a native-like style through editing aspects untypical of the English language, elaborating on ideas, and incorporating more subordination. The reformulated text was checked by a native speaker for accuracy and authenticity.

b. Noticing the gap. Students were provided with a CD recording of the teachers model presentation accompanied by a transcript of the reformulated text. Learners-guided by a sheet provided for that purpose compared their own performance to the model one. Students modified their original transcripts; the output of this process was transcript (2).

4 - Group discussion. A class session was conducted to help students in both groups to work collaboratively with their partners to report back on the modifications they were planning to undertake.
5 - Task repetition and further reflection. Upon reviewing their corrections, learners presented again in front of the class and recorded their performance for the second time. It is worth mentioning that the reformulation group was instructed to avoid mimicking the model presentation as much as possible. The transcript for the repeat presentation was written by the teacher and a copy was given to the students. Later, students in both groups had to compare their first and second transcripts, utilizing the checklist designed for that purpose. Then, in pairs and whole class sessions, comments on improvements and recurrent errors were elicited, and goals for future performance were reported by each student. To assess students’ short-term gains, the presentations first and repeat transcripts were graded by the teacher utilizing the same measures incorporated for analyzing the pretest and the posttest.

Observably, the main difference between the two strategies incorporated lies in the source of self-assessment. In the first strategy, the criteria-led self-reflection, the assessment was student-initiated, as students were responsible for transcribing and modifying their speech in their own way. In the second strategy, however, the corrections were somehow teacher-initiated, in that although the students self-transcribed their output, the teacher was the one in charge of reconstructing students’ oral performance in a native-like style.

**Duration of the treatment**

For both experimental groups, the five introductory sessions comprised five hours. The recording, transcription and correction of each single presentation spanned approximately two weeks. Then, the repetition and further reflection was conducted in one week. Taking into consideration that there were other skills in the language syllabus to cover, each presentation required approximately four weeks to revise, reflect on and repeat. The time spent was distributed evenly for each presentation. As a result, the three presentations took approximately three months to undergo the entire cycle. The treatment commenced on the 11th of October 2010 and ended on the 3rd of January 2011. The timelines for the self-reflection cycle for each presentation-adopted by both experimental groups- were as follows:
### Table (1)
The self-reflection cycle for a single presentation

<table>
<thead>
<tr>
<th>Week</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental group 1 (Criteria group)</td>
<td>Experimental group 2 (reformulation group)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Presentation/recording (5-7 students daily; 5-10 minutes each)</td>
<td>Transcribing by students</td>
<td>Submitting transcripts &amp; revising by teacher</td>
<td>Presentation/recording (5-7 students daily; 5-10 minutes each)</td>
<td>Transcribing by students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35 min in class</td>
<td>35 min in class</td>
<td>35 min in class</td>
<td>Open time at home</td>
<td>35 min in class</td>
<td>35 min in class</td>
<td>35 min in class</td>
<td>Open time at home</td>
<td>35 min in class</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Editing transcript by students</td>
<td>overview of mistakes by the students</td>
<td>Handing the self-assessment checklist</td>
<td>Individual Error analysis</td>
<td>Editing transcript by students</td>
<td>overview of mistakes by students</td>
<td>Handing in audio and written reformulation</td>
<td>Individual gap noticing</td>
<td>Peer group analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45 min</td>
<td>45 min</td>
<td>45 min</td>
<td>90 min in class (extra time at home)</td>
<td>45 min</td>
<td>45 min</td>
<td>45 min</td>
<td>40 min</td>
<td>90 min in class (extra time at home)</td>
<td>60 min</td>
</tr>
<tr>
<td></td>
<td>Student-teacher conference (10 students)</td>
<td>Task repetition &amp; in class recording</td>
<td>Student teacher conference (10 students)</td>
<td>Task repetition &amp; in class recording</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>In class &amp; after class 15 min per student</td>
<td>In class and after class 15 min per student</td>
<td>In class and after class 15 min per student</td>
<td>In class and after class 15 min per student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50 min in class</td>
<td>50 min in class</td>
<td>50 min in class</td>
<td>50 min in class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transcribing the second performance by the teacher</td>
<td>Comparing performance 1&amp;2</td>
<td>Setting goals</td>
<td>Transcribing the second performance by the teacher</td>
<td>Comparing performance 1&amp;2</td>
<td>Setting goals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(home assignment)</td>
<td>In class (60-70 minutes)</td>
<td>20 Min</td>
<td>(home assignment)</td>
<td>In class (60-70 minutes)</td>
<td>20 Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The regular Instruction followed with the control group

The control group had also to choose three presentations; nonetheless, no self-reflection practices were provided. The control group spent the same time as both experimental groups on the presentations, yet the extra time given for repetition, transcribing or self-correction in the experimental groups was compensated for by engaging students in discussions and answering questions pertinent to each presentation prepared ahead of time by the teacher.

Data analysis and results

In this section, a general review of the students’ short and long term gains in terms of accuracy, complexity and self-efficacy will be discussed. Moreover, the accuracy-complexity tradeoff is demonstrated. The following results are presented with reference to the study hypotheses:

* Hypothesis 1: There are no statistically significant differences between the mean scores of the first as well as the second experimental group on the first performance and repeat one of the three presentations.

For the first group, recordings and transcripts 1—3 along with their repetitions were analyzed to trace any short term progress in accuracy and complexity as shown in table (2).

| Performance measures | Indicators | Presentation I | | Presentation II | | Presentation III | |
|----------------------|------------|----------------|----------------|----------------|----------------|----------------|
|                      |            | First | Repeat | t-test | Sig  | First | Repeat | t-test | Sig  | First | Repeat | T-test | Sig |
| Accuracy             | A1         | .58   | .70    | 3.5    | 0.02*| .69   | .86    | 4.6    | 0.00 | .62   | .79    | 7     | -.01 |
|                      | A2         | 0.2   | 0.52   | 5.3    | 0.03 | 0.33  | 0.55   | 3.4    | 0.0  | 0.45  | 0.56   | 1.3   | 0.19 |
| Syntactic complexity | S1         | .46   | .55    | 1.7    | 0.17 | 0.43  | 0.42   | 2.3    | 0.2  | 0.34  | 0.42   | 1.4   | 1.4  |
|                      | S2         | 10    | 10.5   | 0.00   | 0.12 | 10.3  | 9.4    | 0.7    | 0.4  | 10.4  | 10.5   | 1.08  | 0.80 |
| Lexical complexity   | L1         | 0.51  | 0.52   | 3.4    | 0.41 | 0.48  | 0.48   | 0.8    | 0.2  | 0.49  | 0.47   | 1.20  | 0.24 |
|                      | L2         | 6.03  | 6.09   | 0.22   | 0.82 | 6.3   | 6.2    | 0.5    | 0.5  | 5.7   | 6.7    | 2.8   | 0.01 |
|                      | L3         | 0.086 | 0.068  | 1.6    | 0.125| 0.08  | 0.074  | 0.083  | 0.93 | 0.09  | 0.06   | -2.1  | 0.04 |

Table (2)

T-test results comparing the criteria-led group mean scores on the first performance and repeat one of the three presentations
Evidently, table (2) shows that there were statistically significant differences at 0.05 level between students’ first performance and repeat one in favor of the repeat one on both accuracy measures in the three presentations, except for presentation III, where no statistically significant differences as found on the second accuracy measure (A2) (error-free AS-units) (p = 0.19). Regarding syntactic complexity, clearly, in the three presentations, no statistically significant differences were noticed between the first and second performance on both measures. As far as lexical complexity measures are concerned, no statistically significant differences between students’ mean scores on the original performance and repeat one were found on the three presentations (p > 0.05) and even lexical sophistication (L3) dropped in presentation III from M = 0.09 to M = 0.06. On presentation III, however, only lexical diversity (L2) had improved.

For the second group, similarly, transcripts of the presentations 1-3 along with their repetitions were analyzed as in table (3):

**Table (3)**

<table>
<thead>
<tr>
<th>Performance measures</th>
<th>Indicators</th>
<th>Presentation I</th>
<th>t-test</th>
<th>Presentation II</th>
<th>t-test</th>
<th>Presentation III</th>
<th>t-test</th>
<th>Sig</th>
<th>Sig</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>A1</td>
<td>0.7</td>
<td>0.88</td>
<td>2.3</td>
<td>0.0*</td>
<td>0.69</td>
<td>0.74</td>
<td>0.7</td>
<td>0.48</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>0.3</td>
<td>0.58</td>
<td>3.5</td>
<td>0.00</td>
<td>0.42</td>
<td>0.56</td>
<td>2.2</td>
<td>0.00</td>
<td>0.4</td>
</tr>
<tr>
<td>Syntactic complexity</td>
<td>S1</td>
<td>0.2</td>
<td>0.45</td>
<td>2.6</td>
<td>0.02</td>
<td>0.23</td>
<td>0.30</td>
<td>3.3</td>
<td>0.00</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>9.8</td>
<td>11.5</td>
<td>2.9</td>
<td>0.01</td>
<td>9.6</td>
<td>10.4</td>
<td>3.9</td>
<td>0.00</td>
<td>9.5</td>
</tr>
<tr>
<td>Lexical complexity</td>
<td>L1</td>
<td>0.5</td>
<td>0.50</td>
<td>0.5</td>
<td>0.61</td>
<td>0.51</td>
<td>0.52</td>
<td>1.0</td>
<td>0.33</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>5.8</td>
<td>6.9</td>
<td>3.6</td>
<td>0.00</td>
<td>6</td>
<td>6.9</td>
<td>2.9</td>
<td>0.01</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>L3</td>
<td>0.0</td>
<td>0.74</td>
<td>0.5</td>
<td>0.57</td>
<td>0.08</td>
<td>0.09</td>
<td>0.1</td>
<td>0.86</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table (3) shows that there were statistically significant differences at 0.05 level between the mean scores of the reformulation group on the first performance and the repeat one of the first presentation on the
second accuracy measure (A2) (error free AS units) in favor of the repeat one (p≤0.05). Yet, no statistically significant differences were identified on the first accuracy measure (A1) (error free verb forms). Notably, no statistically significant differences on both accuracy measures, however, could be identified on presentation II and III. As far as syntactic complexity is concerned, there were statistically significant differences between the first performance and the repeat one of the three presentations on both measures in favor of the repeat one (p≤0.05). Regarding lexical complexity measures, no statistically significant differences were found between the first and repeat performance of presentation I and II on the first measure L1 (lexical density), yet some statistically significant differences were found on this measure in presentation III in favor of the repeat performance (p≤0.03). On the second measure L2 (lexical diversity), there were statistically significant differences between the original performance and the repeat one of the three presentations in favor of the repeat one. Nevertheless, on the third measure L3 (lexical sophistication), no significant differences could be recognized on presentations I and II (p≤0.05); however, statistically significant differences were found on this measure in presentation III in favor of the repeat presentation (p≤0.01).

**Hypotheses 2:** There are no statistically significant differences between the mean scores of the first experimental group as well as second experimental group on the pretest and posttest.

To evaluate both groups’ long-term gains, pretest and post tests were compared for each group. Paired samples t-test was used as in table (4).

**Table (4)**

*T*-test results comparing the pre- and post-test means for criteria and reformulation-led groups on speaking performance measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>Experimental group (A)</th>
<th></th>
<th>Experimental group (B)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>T-test</td>
<td>sig</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>0.52</td>
<td>0.22</td>
<td>0.72</td>
<td>0.14</td>
</tr>
<tr>
<td>A2</td>
<td>0.24</td>
<td>0.12</td>
<td>0.61</td>
<td>0.29</td>
</tr>
</tbody>
</table>

*Volume 28*
Cont / Table (4)
T-test results comparing the pre- and post-test means for criteria and reformulation-led groups on speaking performance measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>Experimental group (A)</th>
<th>Experimental group (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>Syntactic complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>.33</td>
<td>.15</td>
</tr>
<tr>
<td>S2</td>
<td>8.8</td>
<td>2.03</td>
</tr>
<tr>
<td>Lexical complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>0.46</td>
<td>0.04</td>
</tr>
<tr>
<td>L2</td>
<td>4.7</td>
<td>0.98</td>
</tr>
<tr>
<td>L3</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>

According to table (4), it is apparent that there were statistically significant differences at 0.05 level between the mean scores of the first experimental group on the pretest and posttest on both accuracy measures, as well as on the second measure of lexical complexity L2 (lexical diversity) in favor of the posttest (p ≤ 0.05). On the other hand, no statistically significant differences between the mean scores of the first experimental group on the pre-test and the post-test on both measures of syntactic complexity as well as the first and third measure of lexical complexity L1 and L3 were found (p ≤ 0.05).

As for the reformulation group, there were no statistically significant difference at 0.05 level between the mean scores of the pretest and posttest on both accuracy measures (p ≤ 0.05). So, hypothesis two was partly refuted.

**Hypothesis 3**: There are no statistically significant differences between the mean scores of the first and second experimental groups on the posttest on the identified performance measures.

To compare the speaking performance of both experimental groups and the control group on the post test, the one way ANOVA was utilized as shown in table (5).
Table (5)
One way ANOVA comparing the experimental and the control groups on the posttest on speaking performance measures

<table>
<thead>
<tr>
<th></th>
<th>Means (SD)</th>
<th>Location of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>control</td>
<td>Criteria</td>
</tr>
<tr>
<td>A1</td>
<td>.46 (.14)</td>
<td>.72 (.14)</td>
</tr>
<tr>
<td>A2</td>
<td>.37 (.23)</td>
<td>.61 (.44)</td>
</tr>
<tr>
<td>S1</td>
<td>.20 (.136)</td>
<td>.37 (.15)</td>
</tr>
<tr>
<td>S2</td>
<td>7.73 (2.1)</td>
<td>9.5 (1.4)</td>
</tr>
<tr>
<td>L1</td>
<td>.31 (.04)</td>
<td>.42 (.05)</td>
</tr>
<tr>
<td>L2</td>
<td>4.01 (1.1)</td>
<td>6.2 (.92)</td>
</tr>
<tr>
<td>L3</td>
<td>.04 (.06)</td>
<td>.06 (.03)</td>
</tr>
</tbody>
</table>

Note: CON = control group, CRI = criteria- led group, and REF = reformulation- led group

As shown in table (5), the results of the one-way ANOVA indicate that there were overall statistically significant differences among the three groups on all measures. The post hoc test shows that the almost all statistically significant differences were located between the control and criteria- led groups as well as between the control and reformulation- led group in favor of experimental groups, which demonstrates that both experimental groups outperformed the control group on accuracy and complexity measures. However, noticeably, no statistically significant differences were found between the control and criteria group on the third lexical complexity measure (L3) (lexical sophistication). Likewise, no statistically significant differences were found between the control and reformulation group on the second accuracy measure (A2) (error free AS units.)

Comparing the criteria and reformulation group, it becomes evident that statistically significant differences existed in favor of the criteria group on both accuracy measures. As for syntactic complexity, it is apparent that the reformulation group outperformed the criteria- led group on the first measure (S1) (mean length of AS units) (p≤0.01).
However, no statistically significant differences were found on the second measure (S2) (subordination) \((p \leq 0.62)\). Somewhat different results were obtained for lexical complexity. Regarding lexical density (L1), statistically significant differences were found between both groups in favor of the reformulation group \((p \leq 0.04)\). Likewise, the reformulation group could surpass the criteria group on lexical diversity (L2) \((p \leq .002)\). Yet, no statistically significant differences were found on lexical sophistication (L3) \((p \leq 0.08)\). So, hypothesis three was partly refuted.

Turning now to the experimental evidence on self-efficacy, the hypotheses stated were as follows:

**Hypothesis 4:** There are no statistically significant differences between the mean scores of the first as well as the second experimental group on the pretest and posttest of the self-efficacy scale.

First, to test hypothesis 4, paired samples t-test was used as in table (6).

**Table (6)**

<table>
<thead>
<tr>
<th></th>
<th>Criteria-led group</th>
<th></th>
<th>Reformulation-led group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>T-test</td>
<td>Sig</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>M</td>
<td>S.D.</td>
<td>M</td>
<td>S.D.</td>
</tr>
<tr>
<td>Dimension Ability</td>
<td>1.6</td>
<td>0.44</td>
<td>0.69</td>
<td>4.8</td>
</tr>
<tr>
<td>Activity</td>
<td>1.9</td>
<td>1.16</td>
<td>3.8</td>
<td>0.87</td>
</tr>
<tr>
<td>Aspiration</td>
<td>1.8</td>
<td>0.92</td>
<td>3.3</td>
<td>1.14</td>
</tr>
</tbody>
</table>

According to table (6), there were statistically significant differences at 0.05 level between the mean scores of the first experimental group on the pretest and posttest on the three self-efficacy dimensions \((p \leq 0.05)\) in favor of the posttest. As for the second group, similarly, there were statistically significant differences at 0.05 level between the mean scores of the pretest and posttest on ability perception \((p \leq 0.00)\), activity perception \((p \leq 0.04)\) and aspiration \((p \leq 0.03)\) dimensions in favor of the post test. Thus, hypothesis four was refuted.
Hypothesis 5: There are no statistically significant differences between the mean scores of first and the second experimental groups on the posttest of the speaking self-efficacy scale.

To compare the three groups (control, criteria and reformulation) on the posttest, the one way ANOVA was conducted as in table (7).

Table (7)
One way ANOVA comparing the experimental and the control groups on the posttest on self -efficacy dimensions

<table>
<thead>
<tr>
<th></th>
<th>Means (SD)</th>
<th>Location of significance</th>
<th>Sig</th>
<th>CON-CRL</th>
<th>CON-REF</th>
<th>CRI-REF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>1.7(.54)</td>
<td>2.9(.69)</td>
<td>4.1 (.9)</td>
<td>38</td>
<td>0.00</td>
<td>0.01*</td>
</tr>
<tr>
<td>Ability</td>
<td>2.3(.72)</td>
<td>3.8 (.87)</td>
<td>2.9 (1.2)</td>
<td>12.6</td>
<td>0.00</td>
<td>0.00*</td>
</tr>
<tr>
<td>Activity</td>
<td>1.8(.72)</td>
<td>3.3(1.1)</td>
<td>3.1 (.8)</td>
<td>15.9</td>
<td>0.00</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

Table (7) shows that self-efficacy differed significantly across the three groups, F (2, 46) = 38, 12.6 and 15.9, p≤.05 for the three dimensions respectively. Tukey post-hoc indicates that the criteria-led group surpassed the control group on all self-efficacy measures, p≤.05. Similarly, there were statistically significant differences between the control and reformulation groups in favor of the reformulation group on both ability perception and aspiration dimensions (p≤0.00); yet no statistically significant differences were found on activity perception, p≤.05, between the reformulation and control group. Similarly, the reformulation group (M=4.1) significantly outperformed the criteria-led group (2.9) on the ability perception dimension (p=.00). In contrast, statistically significant differences were found in favor of the criteria led group (M =3.8), (p≤.03) on the dimension of activity perception. Nonetheless, the comparison between the two groups on the aspiration dimension reveals no statistically significant differences. So, hypothesis five was partly refuted.

Discussion

This study was set out to explore the effect of criteria-led and reformulation led self-reflection on complexity and accuracy measures
as well as on the self-efficacy of low- intermediate EFL Learners. By and large, throughout qualitative observation, an exciting result that could be reached was the students positive response to the notion of noticing the gap which was reflected in their enthusiasm for recording and transcribing their spoken performance. Nonetheless, boredom experienced by some students during repetition could not be handled adequately by the teacher. This is supported by previous results (Lynch, 2007, Christianson and Hoskins, 2008, Rahim and Riasati, 2011 and Mennim, 2012). Quantitative data, however, yielded more insights into both experimental groups’ performance as shown in the following section.

**Effect on Accuracy:**

By and large, both groups relatively gained in terms of accuracy as they both outperformed the control group. This supports the assumed effect of the self-reflection treatment techniques including transcribing, repetition and self-correction. However, some discrepancies between the two experimental groups were discernible. In the short term, for the criteria-led group, a notable progress could be observed, on both accuracy measures, which was not the case with the reformulation group. A closer qualitative look revealed that about 70% of the errors corrected by students in the criteria led group compared to approximately 30% for the reformulation group were grammatical mistakes. A possible explanation for criteria led group’s improvement might be that self-reflection against clear-cut criteria had led students to prioritize the correction of their mistakes and although students’ attention was directed to all aspects of spoken performance, they tended to view their original flaws as mainly grammatical or lexical. Lynch (2001, 2007) drew similar conclusions and showed more gain in accuracy after self-transcription. This focus on accuracy had transferred to new tasks, which gives indication that subsequent to assessing their performance against certain criteria, students attention might have become totally channeled to producing accurate language in the long term.

As far as the reformulation group is concerned, the students’ short-term progress was not significant on both accuracy measures, except on the first presentation where some progress could be discerned on the second accuracy measure "error free AS unit". Possibly, these results
are due to the fact that students’ attentional resources were mainly exhausted in content related aspects at the expense of attaining accuracy. That is to say, being occupied with the model reconstructed presentation, students seemed to have given priority to ensuring a thorough coverage of their topic and supporting their ideas with enough details to emulate the model performance. Hence, they seemed to have not achieved the same progress as the criteria-led group a tendency which unconsciously was transferred to the unseen posttest task.

**Effect on Complexity:**

By and large, there was indication that complexity grew for both groups in comparison to the control group which attests to the fact that transcribing, noticing and self-correction can contribute to stretching students intra-language system. Nevertheless, the progress for the reformulation group in terms of syntactic complexity both in the short and long term was more remarkable. This was also clear regarding lexical density and diversity, though in the short term gains in lexical density was not dramatic.

Actually, throughout the treatment, attending to a reformulated version of their output, the reformulation groups noticing efforts became highly focused on how ideas are expressed in more elaborate ways. This gives indication to the effectiveness of reformulation in enhancing deeper level of processing that goes beyond the peripheral noticing of grammatical errors. This short term progress seemed to have transferred to new tasks as well. This was sustained by previous research in favor of using reformulation (Adams, 2003 & Ibarrola, 2009).

However, the criteria-led group, in spite of the slight gain in the short term in lexical density, could not achieve much progress in terms of lexical density and lexical diversity in the long term. Throughout the treatment, students tended to stick to their original presentations and directed their focus to correcting grammatical or lexical mistakes, and hence none of their suggested elaborations or content editing undertaken on the self-assessment sheet appeared in their repeat spoken performance. In the long term, also, students syntactic and lexical complexity remained unaltered. This supports the contention that noticing the mismatch at the complexity level cannot be fully realized unless a model for comparison-derived from learners’ output-
provided. Alternatively, leaving students to their own devices or just providing them with self-assessment criteria might confine their abilities or deprive them from the chance of exploring new ways of delivering their own thoughts by considering new expressions and words or by opting for new ways to show the interconnectedness of their ideas.

Notably, the results gave indication that sophisticated words did not grow much for the criteria-led group. On the other hand, a moderate progress could be observed in the short term for the reformulation group, which could barely transfer to new posttest tasks. An implication of this is the possibility that under time pressure students tended to rely on their available active vocabulary to cope with time constraints and so they avoided trying out words or expressions they felt not quite familiar with. That is to say, developing sophistication might have been hindered by students’ over-reliance on their fossilized vocabulary repertoire. In other words, passive vocabulary could hardly be activated during such a limited-time treatment for both experimental groups.

Effect on Self-Efficacy:

To some degree both groups experienced a significant positive gain in self-efficacy from the pre-test to the post test and also surpassed their counterparts in the control group. This progress lends itself to the interpretation that subsequent to self-reflection treatment, the students were more likely to believe that they could speak English better. Repeating the same task might have given students the chance to experience mastering speaking skills which contributed to their enhanced self-efficacy. In addition, taking the responsibility of transcribing, assessing and modifying their own performance might have elevated their sense of self-regulation, so that they began to attribute success to their effort rather than external factors such as chance. This study corroborates other studies that show how self-reflection can positively affect learners’ self-efficacy (Huang & Chang, 1996; Pajares et al., 2007; Hsieh, & Kang, 2010).

Nevertheless, the reformulation group had shown more gain in terms of ability perception which might be attributed to the fact that they—instead of being left to their own discretion—were engaged in noticing a model linguistic output which served as a referential frame-
work for learning and modifying their own performance. This is confirmed by other researchers who point to the fact that observing models can powerfully influence students' self-efficacy (Pajares et al., 2007). Furthermore, this supports the role of self-referenced assessment (reformulating the speakers own output), rather than norm referenced self-assessment (using external criteria), in fostering EFL learners' internal attribution of behavior (Chan and Lam, 2010). Reformulation seemed to have acted as a form of positive feedback that induced error noticing without undermining students self-esteem.

In terms of activity perception, it can be assumed that both groups seemed to have acquired positive attitudes towards class oral activities. This can be due in part to the possibility that students started to realize how oral activities acted positively as a stimulant for testing hypotheses about their language proficiency which in turn paved the way for improvement. Observably, the criteria-led group seemed to have surpassed the criteria-led group in this respect. This might be accounted for by the fact that assessing oneself against a checklist did not make such demands on students' cognitive ability as is the case with comparing oneself to a model presentation which might have sounded somehow more ambitious, exhausting and cognitively challenging.

As for the third dimension, equally both groups expressed high aspiration to become good English speakers. A possible explanation might be that most of EFL learners perceive English spoken proficiency as an indispensable element in their academic success. Another interpretation might be that the technique of setting goals following the self-reflection cycle might have inspired students, in both groups, to set higher yet reachable expectations for the future. This might also be attributed to the emerging perceived self-confidence in managing oral progress created by the self-reflection experience in its entirety. This was corroborated by Hawkes (2009) and Leger (2009).

**Conclusion**

The current study supports the contention that both self-reflection strategies have a positive effect on students' accuracy and complexity. Nonetheless, given the learners' limited attentional resources, it was found that students who were guided by certain criteria gave priority to correcting their recurring errors, whereas students who analyzed a
reformulated version of their production had a tendency to focus more on elaborating their ideas. It can, hence, be assumed that reformulation disposes students to deeper level of processing the content.

Furthermore, taken together, these results suggest that students’ gain in accuracy can be transferrable to new tasks as shown in the criteria-led group performance. Short term gains in syntactic complexity could also be transferred to new tasks, but not in terms of subordination. At the lexical level, however, students could only achieve progress regarding lexical density and diversity on new tasks; nevertheless, sophistication seemed resistant to change. The study also suggests that inducing students to self-reflect on their performance can have a positive effect on their level of confidence to speak the target language, especially if self-reflection is carried out with reference to a modified version of their own output.

Based on the previous results, some implications can be drawn. First, EFL curriculum designers and teachers should adopt self-reflection practices to enhance students’ oral proficiency. Substantially, it stands to reason that in order to address accuracy and complexity, EFL instructors are advised to employ a combined self-reflection approach, by encouraging students to first evaluate their performance against a set of criteria and subsequently expose them to a reconstructed version of their output. Remedial lessons are recommended to be added to the self-reflection cycle to help students work on expanding their vocabulary repertoire by stimulating them to experiment with synonyms for the fossilized set of words they keep resorting to under time pressure.

The results of this study suggest possibilities for future research. One suggestion is utilizing self-reflection to promote pragmatic competence, fluency and discourse competence. Further, studies need to investigate if students learning style, motivation and proficiency level can shape their response to self-reflection. Moreover, given the impracticality of reformulating each student’s performance, needs emerge to investigate the possibility of incorporating a unified model text that addresses common students’ weakness without compromising individual differences. Substantially, also, further work needs to explore the interaction between students self-efficacy-as an independent variable-on the one hand and self-reflection -as a dependent variable-on the other hand.
أثر استخدام استراتيجيات بديلتين في تنمية مهارات المحادثة والثقة بالنفس باللغة الإنجليزية

د. شيماء عبدالفتاح تركي
المركز القومي للبحوث التربوية والتنمية
ج. 3/0/4

الملخص
تهدف الدراسة الحالية إلى مقارنة أثر استراتيجيات تعليمية، كلتا الاستراتيجيات تطبقاً مفهوم التأمل الذاتي لتنمية مهارات التحدث باللغة الإنجليزية والثقة بالنفس.
وتعد الاستراتيجية الأولى على استخدام مجموعة من المعايير المحددة كأساس للتقييم الذاتي بينما تعتمد الاستراتيجية الثانية على إعادة صياغة اللغة المحدثة للمتعلم وتقديمها له ليعتمد عليها في عملية التقييم الذاتي. أي أن المقارنة تتم بين استراتيجيتين تشتمل على نموذج أو نسخ تتم إعادة صياغته واستراتيجيته أخرى تخلو من هذا النموذج. تشمل كلتا الاستراتيجيتين على مجموعة مشتركة من الأساليب: هي تقديم عرض شفهي، نسخ النص المحدث إلى نص مكتوب، التقييم الذاتي، إعادة العرض والتحقيق الذاتي مرة أخرى. تكون عينة الدراسة من 49 طالبة بجامعة الكويت تم توزيعهن على مجموعتين تجريبيتين ومجموعة واحدة ضابطة. قامت المجموعتان التجريبيتين بتقديم عرض شفهي ينبع واحدة من استراتيجيات التأمل الذاتي، إجاهما تركز على التقييم الذاتي في ظوء مجموعة من المعايير، بينما تعتمد الاستراتيجية على التقييم في ظوء إعادة صياغة العرض الذي قدمته الطالبات. على الوجه المقابل، قامت المجموعة الضابطة بتقديم نفس العروض الشفهية دون أن يعقبها أي من استراتيجيات التأمل الذاتي. وقد أسفرت النتائج عن فعالية كلتا الاستراتيجيتين في تنمية كل من الأداء الشفهي والثقة بالنفس في تحمل اللغة الإنجليزية.
كما أن الدراسة أظهرت أن التأمل الذاتي المركزي على المعايير كان ذا فاعلية أكبر في تنمية الصحة اللغوية بينما كان لأستراتيجية التأمل الذاتي المركزي على إعادة صياغة العرض الشفهي أثر أكبر في تعقيد التراكيب اللغوية المستخدمة من قبل الطلاب. وافقت النتائج أن كلتا الاستراتيجيتين لهما تأثير إيجابي في تنمية الثقة بالنفس لدى الطلاب.
REFERENCES


from: http://web.icu.ac.jp/lrb/vol_23_files/Christianson%26Hoskins2.pdf


reading and listening proficiency. *Foreign Language Annals*, 39(2), 276-293.


48 - Sellars, M. (2010). Intrapersonal intelligence, executive function and stage three. (Doctoral dissertation, Australian Catholic Uni-
theses/public/adt-acup281.02032011/02whole.pdf

49 - Shang, H. (2010). Reading strategy use, self-efficacy and EFL

Intelligence: Affective Factors in Thinking. Education, 119, 4,
633-648.

through Noticing-Reformulation Tasks. Paper presented at the
Phonetics Teaching and Learning Conference Programme. Re-
trieved from: http://www.phon.ucl.ac.uk/home/johnm/ptlc2005/
pdf/ptlc46.pdf

52 - Stillwell, C., Curabba, B., Alexander, K., Kidd, A., Kin, E.,
Noticing fluency, accuracy, and complexity, ELT Journal, 64 (4),
445-455. doi:10.1093/elt/ccp081

In E. Hinkel (Ed.), Handbook of research in second language

tive conflict and L2 learning it generates. International Journal of

55 - Torky, S. (2006). The effectiveness of a proposed program in the
light of task based instruction in developing English Language
speaking skills among secondary stage students. (Doctoral dis-
sertation, Women’s College, Ain Shams University. (ERIC Docu-
ment Reproduction Service No. ED420967)


(Doctoral dissertation). Ohio State University Retrieved from
ProQuest, 3477045.
تهدف إلى معالجة المشكلات المعاصرة والقضايا المستجدة من وجهة نظر الشريعة الإسلامية.
تشمل موضوعاتها معظم علوم الشريعة الإسلامية: من تفسير، وحديث، وفقه، واقتصاد وتربية إسلامية، إلى غير ذلك من تقارير عن المؤتمرات، وراجعات كتب شرعية معاصرة، وفتاوى شرعية، وتعليقات على قضايا علمية.
تنوع الباحثون فيها، فكانوا من أعضاء هيئة التدريس في مختلف الجامعات والكليات الإسلامية على رقعة العالمين: العربي والإسلامي.
تخضع البحوث المقدمة للمجلة إلى عملية فحص وتحكيم حسب الضوابط التي التزمت بها المجلة، ويقوم بها كبار العلماء والمختصين في الشريعة الإسلامية، بهدف الارتباط بالبحث العلمي الإسلامي الذي يوجد الأمة، ويعمل على رفعة شانها، نسال المولى عز وجل مزيدًا من التقدم والازدهار.

جميع المراسلات توجه باسم رئيس التحرير
صبب 174/23 – رقم البريد: 72465 - الكلية: 298/81 - 05 - 09
fax: 744
E-mail - jsis@ku.edu.kw
issn: 1029 - 8908
http://pubcouncil.kuniv.edu.kw/JSIS
عنوان المجلة على شبكة الإنترنت: Social and Human Sciences Documentation Center
في شبكة الإنترنت تحت الموقع
www.unesco.org/general/eng/infoserv/db/darc.html