

## **The Impact of Teachers' Self-evaluations through Rubrics versus Supervisors' Observation on their Performance in Setting Group-work Activities**

Elaheh Safari Pakrou, M.A.

Department of English, Garmsar Branch, Islamic Azad University, Garmsar, Iran  
*safarisalimeh@yahoo.com*

Bahram Bagheri, Assistant Professor

Department of English, Garmsar Branch, Islamic Azad University, Garmsar, Iran  
*bahram\_bagheri51@yahoo.com*

### **Abstract**

As a tool for self-evaluation and also self-regulation, rubrics can be very useful in equipping the language teachers with an assessment device. Moreover, teachers need feedback for their professional development, but how feedback is provided is a challenging question. This study was an attempt to find the effectiveness of self-evaluation through rubrics on setting up group work activities by EFL teachers as compared with the effect of supervisors' evaluation in this regard. Additionally, EFL teachers' attitudes with respect to the efficacy of rubrics in setting up group work activities were examined. To this end, 10 female EFL teachers with 2 years of experience were selected and further divided into two groups. One group used rubrics for setting up group work activities and the other used post-observation feedback. Their performances were video recorded and rated by two raters. The results of statistical analysis indicated that the group of teachers who used rubrics to set up group work activities and evaluate their performance significantly outperformed the group of teachers who received post-observation feedback. Teachers' responses to the attitude questionnaire also showed that overall teachers held a positive attitude towards the use of rubrics to self-evaluate their own performance.

**Keywords:** Self-evaluation, Self-regulation, Rubrics, Feedback, Observation

### **Introduction**

There are numerous studies on the topic of teacher evaluation in higher education settings. These studies address various aspects of teacher evaluations such as problems found in teacher evaluation and the various evaluation strategies used in teacher evaluation programs (Cardno, 1999; Stronge, Richard, & Catano, 2008). Researchers in higher education environments claimed that teacher evaluation should lead to improved teacher performance as well as professional growth and development (Marincovich, 1998, 1999; Smith, 2008). Supervising and evaluating teachers have developed over the past 300 years with the most notable changes happening in the last three decades. However, evaluation continues to be mediocre and limited in helping teachers change their pedagogy for the benefit of their students, despite the changes and increased focus on standards and supervision (Darling-Hammond, 2010; Harris, 1986; Bridges, 1990; Tucker and Stronge, 2005). Teachers and administrators are burdened with a process which is based on a limited number of observations in the classroom. Moreover, it demands a lot of time to complete the necessary paperwork and there is almost no evidence that the process actually better professional practice (Moss, 2010; Danielson & McGreal, 2000; Papay, 2012; Danielson, 2009). With regard to what was stated above, this study was an attempt to shed more light on the effect of

teachers' self-evaluation through rubrics on the quality of group work activities in English language classes.

## **Literature Review**

### **Rubrics and Self-evaluation**

Rubric is considered as a type of a criterion-specific performance scale which consists of a set of scoring criteria used to distinguish the achievement level of a performance on given tasks. As a matter of fact, a rubric breaks down a task into its components and gives detailed accounts and description of the performance levels of each component. The descriptions are presented in written mode so that learners can learn what need to be done to enhance their performances in the future (Panadero & Jonsson, 2013).

Panadero and Jonsson (2013) argue that criteria for assessment are the standards against which the implementation and the final outcome of a task are assessed. Students, especially those at the lower level of their education, in order to assess themselves appropriately need to clearly understand and digest these criteria. Rubrics and scripts contain these assessment criteria. Rubrics "a document that articulates the expectations for an assignment by listing the criteria or what counts, and describing levels of quality from excellent to poor" (Reddy & Andrade, 2010). Rubrics have been shown to enhance student performance and learning if used in combination with metacognitive activities (Panadero & Jonsson, 2013). Rubrics – by definition – are quite detailed scoring guides, they can be used in order to validly assess multi-dimensional performances (Andrade & Valcheva, 2009; Halonen et al., 2003).

Rubrics are self-assessment tools with three characteristics: "a list of criteria for assessing the important goals of the task, a scale for grading the different levels of achievement and a description for each qualitative level" (Reddy & Andrade, 2010, p 387). Evaluation criteria are the factors that an assessor considers when determining the quality of a student's work (Reddy and Andrade, 2010). When students use rubrics, they can compare their projects against the criteria or "standards" in the rubric, and then self-grade their work based on the rubric (Reddy and Andrade, 2010, p 387). Although rubrics are designed to analyze the final product of an activity, it is recommended that they are given to students before they start a task in order to help them establish appropriate goals (Alonso-Tapia & Panadero, 2010).

### **Group work**

Group work is understood here as a generic term covering a multiplicity of activities in which three or more students are assigned a task that involves collaboration and self-initiated language. It implies groups of up to six students (Brown, 2001, p 177). Interaction in communicative classes is a must and group work is a good way to promote interaction. Since the focus of the study is on group work, the benefits and some teacher's concerns about group work activities are presented. Firstly, group work can increase students' involvement in the lesson. Bejarano (1987) and Fushimo (2010) highlight that group work provides students with the opportunity to be actively involved in a communicational interaction in the target language, developing linguistic competence. This involvement could be oriented to develop cooperative learning, in which "the learning task is based on interaction and reciprocal interdependence among the members of the group and requires mutual help" (Bejarano 1987: 485). Secondly, some studies have suggested that group work promotes a positive affective climate in the classroom, reducing anxiety especially through interaction among students (Long & Porter 1985; Davis, 1997; Brown, 2001; Hess, 2001). Consequently, group work

could provide opportunities to use the target language in a safe environment, allowing students to support each other. Fushino (2010) found that the "Communication Confidence in the L2 Group Work was a strong predictor of the WTC (willingness to communicate) in the L2 Group Work" (Fushino, 2010, p, 715). This means that the more confident students feel in their group the more willing they are to communicate in the target language during groupwork.

However, despite all these benefits for group work activities, some English teachers are concerned about setting group work activities. The reason most commonly cited by teachers why they do not use group work is classroom management. For example Davis (1997) and Brown (2001) reported that teachers have difficulty controlling class discipline during group work. This could be related to a lack of instruction or preparation of the strategy. For some teachers group work means putting students together without planning or developing a rationale to use group work for that activity or task. For example Davis (1997) and Brown, (2001) point out that teachers may implement group work superficially, and that many of them do it without a thorough understanding of the underlying purposes for the technique. As a result, the outcomes tend to be limited due to a poor design.

### **Research Questions**

Based on the purpose of the study outlined above, following research questions were formulated:

**Q1:** Does teacher's self-evaluation through rubrics lead to setting better group work activities than evaluation done by supervisors?

**Q2:** what are the teachers' attitudes regarding the efficacy of rubrics in setting group work activities?

### **Research Hypothesis**

In line with the above research question, the following null hypothesis was formulated:

**H01:** Teacher's self-evaluation through rubrics does not lead to setting better group work activities than evaluation done by supervisors.

### **Participants**

In line with the purposes of this study, 10 EFL female teachers were selected from one of the distinguished private English institutes in Tehran (Zabansara Language Institute). The teachers were chosen based on availability and convenience and their age ranged between 18 to 35. The element of purposiveness was also included. In other words, only female teachers teaching at intermediate level with at least 2 years of experience were selected to control the effect of level of qualification and gender. The reason behind the selection of this age group of teachers was based on the assumption that age may affect the behavior under question. Convenient sampling was used for choosing the participants; that is, the participants were chosen on the basis of their availability at the time of data collection (of course, if they were willing to participate). Therefore, there was no random sampling or stratified random sampling. After selecting the legitimate participants of the study, they were divided into two groups. One group was rubric group and the other one was the observation group. Rubric group of teachers was given the rubrics for setting up group work activities and to evaluate

their performance and observation group of teachers was observed by a supervisor and in the post-observation conference they were given feedback for their performance.

### **Instruments**

The instruments utilized in the current investigation were group-work rubric, Observation Schedule, Attitude Questionnaire, Rating Scale, and Face to Face 2<sup>nd</sup> edition intermediate book a description of which follows:

### **Rubric**

In line with the purpose of the study, a rubric was adopted from Pearson Education Teaching Tools to be given to rubric group of teachers. The purpose of this rubric was to provide a tool for teachers to carry out the self-evaluation process without any assistance. The rubric contained 8 criteria with each one ranging on a scale of one to five. To validate and modify the content of the rubrics to fit in foreign language classes, a team of experts was formed. The members of this team included the researcher and three Ph.D. graduates in TEFL. They had obtained their Ph.D. in TEFL. The original rubric contained 8 categories which were revised and therefore some items were deleted and some items were added. After revision the number of items was increased to 12 items.

### **Observation Schedule**

The regular observation being carried out by supervisor of the institute was conducted again with the aim of providing feedback to teachers in order to promote their awareness and skill regarding setting up group work activities. It needs to be further noted that in post-observation sessions teachers were guided and informed about how to set up group work activities and what are considered the characteristics of a good group work activities. The feedback provided to teachers contained the criteria in the rubrics. These observations were carried out to provide feedback to observation group of teachers and rubric group of teachers were not observed for providing feedback.

### **Attitude Questionnaire**

Based on the criteria in group activity rubrics a questionnaire was designed to survey students' attitudes toward the efficacy of rubrics. The questionnaire was constructed drawing on Dorniyev's (2005) guidelines for questionnaire construction and was based on three major sub-constructs of attitude (Fishbein & Ajzen, 1975) including convenience, usefulness and preference. This questionnaire employed a 5-point scale for participants to indicate their answers (5-strongly agree, 4-agree, 3-neutral, 2-disagree and 1-Strongly disagree).

After developing the first draft of the questionnaire, in order to assure its appropriacy in the current research context, it was piloted twice. Once on five participants with characteristics the same as participants of the study to gain insights in terms of the appropriate wording and grammar of the items and once to run Cronbach's Alpha for the purposes of establishing the required internal consistency. After the first piloting phase of the study, the questionnaires were collected and analyzed. Two days after that, 20 minute sessions were held with the participants to gather their viewpoints in terms of the appropriacy of the items. The sessions unfolded following these steps:

- The respondents were given the questionnaire and asked to respond to the items again.
- The questionnaires filled out in this session were compared with the previously filled out ones and any differences were spotted.

- The participant was asked to answer why there was a change in the answers provided. Most of the changes were found to be rooted in the ambiguities involved in the wording.
- The findings were used to make revisions to the questionnaire. In this regard, the participants' suggestions concerning the right wording and easy grammar were taken into account as well.

After the revisions were carried out, the questionnaire was piloted again on 10 teachers with the same characteristics of the participants. The data gathered in this pilot study underwent Cronbach's Alpha to assure the internal consistency of the questionnaire. The first version of the questionnaire contained twenty items. Through the procedure of Cronbach's Alpha four questions which were found to affect this index negatively were deleted in a stepwise manner. Table 1 displays the amount of Cronbach's Alpha with all the items (20) included.

Table 1

*Cronbach's Alpha with all the 20 questions considered*

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.569	.578	20

In the next step those items whose deletion impacted the reliability index of Cronbach's Alpha positively were deleted one at a time until the index reached an acceptable level of .720. Table 2 illustrates the results of Cronbach's Alpha after deletion of four items.

Table 2

*Results of Cronbach's Alpha after deleting four of the items*

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.720	.727	16

### **Rating Scale**

To rate teachers' performance in setting up group work activity, the same rubrics used by teachers were given to two raters. One rater was the researcher and the other one was a member of expert team. Therefore, the same rubric served as the rating instrument in evaluating the teachers' performance in setting up group work activities. It needs to be noted that all categories in teachers' rubric and all the rating bands were present in the raters' rubric. Since the rubric would be used as a tool by observers for assessing performance of teachers in terms of quality of students' group activities, it was piloted for reliability issue. In this pilot the rubrics were given to two observers in the institute to assess the quality of group work activities in classes. They observed five classes independently. The classes were the same classes for the observers. In order to find the reliability of the rating interclass

correlation coefficient was conducted. Table 3 shows the results of a Two-Way Random Consistency ICC Calculation in SPSS.

Table 3

*Interclass Correlation Coefficient*

	Intraclass Correlation	Value	Sig
Single Measures	.114	11.538	.043
Average Measures	.606	11.538	.043

Therefore, 0.60% of the variance in the mean of these raters was “real”. The scale enjoyed acceptable level of consistency.

**Teaching Materials**

The course book studied by students was Face to Face 2<sup>nd</sup> edition intermediate book. This book was the regular course book of the institute in which research was conducted. Students had passed the previous courses or had been put in intermediate course based on placement test scores. The placement test given to the students was the standard placement test published by Cambridge University Press. The placement standards to discriminate students were based on the instructions in the placement test manual. The placement test could be accessed through going online at [www.cambridge.org](http://www.cambridge.org). A collection of 20 group work exercises was chosen for the research purposes. Out of 20 exercises, 10 was selected to be included in the experiment. The selection criteria were as follows:

- Being interesting and engaging
- Requiring at least 3 people
- Familiarity of students with the topics
- Correspondence between the level of difficulty of the exercise and students’ level of cognitive development and English proficiency.

The selection was performed through attaining agreements of all expert team members. An example of group work exercise can be found in Appendix 1. Again the ten exercises were further divided randomly into two groups. Each group consisted of five exercises and the first group of exercises was used in the first five sessions and the second group in the second five sessions of the experiment. As was stated earlier the first five sessions served as the training sessions. For the rubric group, teachers conducted the group work exercises by following and evaluating the group works through rubrics. And for the observation group, teachers were observed during handling group work activities and were given feedback based on the rubric criteria.

**Procedure and Data Collection**

After selecting the participants of the study, they were divided into two groups of teachers. Next, teachers’ performance in conducting group work activities in both groups were video recorded for five sessions. Raters watched the films several times and rated the quality of group work activities based on the rubrics. The two groups were compared with each other to ensure that two groups are not significantly different from each other in terms of

their performance based on rubric scores. One group was given the rubrics and asked to conduct the group work activities and evaluate their performance.

The group receiving rubrics called rubric group conducted the group work activities and evaluated their performance based on given rubrics. The other group of teachers received feedback about their performance based on regular observation carried out by the supervisor. Totally teachers received feedback regarding their performance in conducting group work activities for five sessions and accordingly rubric group used the rubrics for five sessions. In each session one group activity was performed by the teacher as specified in the syllabus by the researcher and expert team. Teachers taught for five more sessions after the initial five sessions which served as the evaluation and scoring sessions. In these five sessions a close-circuit camera recorded the teachers at work and used by the raters to assess teachers' performance in terms of setting group activities. Scoring procedure as mentioned earlier was carried out based on the identified criteria and scoring system. Each criteria or component was rated on a scoring band of 1 to 5. The researcher and inter rater scored the recorded performances and the average score was counted as the final score.

## Results

### Normal Distribution of Data

The normal distribution of data was conducted for those data used to detect the difference between rubric group and observation group in terms of their performance in setting up group work activities. This was to make sure if parametric tests were suitable for data analysis. To this end, One-Sample Kolmogorov-Smirnov Test was utilized to check the normal distribution of data. Table 4 shows the results One-Sample Kolmogorov-Smirnov Test.

Table 4

*One-Sample Kolmogorov-Smirnov Test*

		Rater1	Rater2
N		10	10
Normal Parameters <sup>a</sup>	Mean	31.9000	32.8000
	Std. Deviation	4.81779	4.56557
Most Extreme Differences	Absolute	.174	.183
	Positive	.174	.183
	Negative	-.130	-.121
Kolmogorov-Smirnov Z		.551	.577
Asymp. Sig. (2-tailed)		.922	.893

a. Test distribution is Normal.

As seen in Table 4 the data collected from two raters are normal. Speaking statistically, the significant value was 0.92 for rater number one and 0.93 for rater number two. Since the values are greater than confidence interval of 0.05, then it was concluded that data enjoyed the needed normality. To further make sure about the normality of data, average

of scores obtained from two raters were also calculated and One-Sample Kolmogorov-Smirnov Test was run. Table 5 shows the result of One-Sample Kolmogorov-Smirnov Test on average scores.

Table 5

*One-Sample Kolmogorov-Smirnov Test*

		Average scores
N		10
Normal Parameters <sup>a</sup>	Mean	32.3500
	Std. Deviation	4.63111
Most Extreme Differences	Absolute	.173
	Positive	.173
	Negative	-.147
Kolmogorov-Smirnov Z		.546
Asymp. Sig. (2-tailed)		.926

a. Test distribution is Normal.

Again the One-Sample Kolmogorov-Smirnov Test proved that data enjoyed normality. Significance level was 0.92 indicating that the test distribution was normal.

**Difference Between Rubric Group and Observation Group in Setting up Group Activities**

Based on the scores obtained from the two raters, mean score for each individual calculated. These scores were related to their adherence to group activity criteria as determined by the two observer raters. The Score were added up and final scores were computed for each student. Table 6 shows the descriptive statistics related to final scores of the participants of the study.

Table 6

*Group Statistics*

	group	N	Mean	Std. Deviation	Std. Error Mean
Final scores	Rubric	5	36.0000	3.50000	1.56525
	Observation feedback	5	28.7000	1.64317	.73485

Table 6 shows that rubric group had a mean score of 36 and observation feedback group a mean score of 28.700. Standard deviation of rubric group was 3.5 and that of observation group was 1.64.

Since it was confirmed that data were normal, parametric test of independent samples t-test was employed to find the statistical difference between rubric group and observation feedback group. Table 7 shows the results of independent samples t-test.



Table 7

*Independent Samples Test*

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	Sig. (2- tailed)	Mean Difference
TotalAverage	Equal variances assumed	3.615	.094	4.222	.003	7.30000
	Equal variances not assumed			4.222	.006	7.30000

As seen in the table 7, test of Levene had a significant value of 0.09 indicating equal variance of the groups. T value was 4.22 with significant value of 0.003. Therefore, it was concluded that there was statistical significant difference between rubric and observation feedback groups of teachers. As mention earlier rubric had higher mean score which means that rubric group of teachers performed better in setting up group work activities.

**Teachers' Attitudestowards Rubrics**

Teacher's attitudes towards rubrics were examined through asking them to complete a questionnaire. On the whole it can be claimed that teachers who experienced use of rubrics hold a positive attitude towards the use of rubrics. Teachers' responses to the each item of the survey can be found in Table 8. As evident in Table 8 most of the teachers agreed with statements of the survey. There were two statements with which two teachers out of five disagreed. One was the statement:*Relying on Rubrics does not hinder teacher's creativity*. One of the criticisms that can be put forward for the use of rubric is the fact that following rubric may not provide the teachers with enough space for creativity. For this statement one teacher strongly disagreed and one teacher held neutral attitudes.

Table 8

*Teachers' Responses to Each Item of the Attitude Questionnaire*

		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	A rubric is a positively challenging learning opportunity.	2	2	1		
2	I believe that an ongoing, long- term, continuous, and constant assessment which promotes learning is obtained through application of Rubrics.		3	2		
3	I believe Rubricsbrings more order and organization to teaching performance.	3	2			
4	Relying on Rubric does not hinderteacher's creativity.	1		1	2	1

5	It is useful to implement rubrics for purpose of setting up group work activities.	1	3	1
6	I feel more motivated when I use rubrics		1	3
7	I believe rubrics makes a lot of difference to the group work outcome.		2	2
8	Rubrics can be helpful tool for teachers in various aspects of teaching.		1	2
9	I found rubrics very exciting and interesting.	1	1	2
10	I was very anxious while I had to handle group work activities but rubrics helped me feel better when setting up group work activities.	1	1	3
11	I prefer the use of rubric in setting up group work activities to classroom observation.	2	3	
12	I felt more confident when setting up group work activity by using rubrics.	1	2	2
13	It was much easier to handle group activities using rubrics.		2	2
14	Students seemed to be happier when I set up group work activities through using rubrics.		1	3
15	I found it helpful when I could monitor my performance through using rubrics.	1	2	2
16	I believe use of rubrics could lead to more accurate and quality performance in class.	1	2	1

Similarly two teachers out of five teachers disagreed with the statement *Rubrics can be helpful tool for teachers in various aspects of teaching*. Two teachers chose the choice *Neutral* and one teacher chose *Disagree*. However it needs to be taken into account that this item ask for teachers' attitudes towards the use of rubrics in other aspects of language teaching and not group work activities.

The teachers' answers to item number 11 support teachers' better performance in rubric group in setting up group work activities. Two teachers *Strongly Agreed* and 3 teachers *agreed* with the statement *I prefer the use of rubrics in setting up group work activities to classroom observation*. This means that teachers in rubric group were happy with use of rubrics in setting up group work activities. Three teachers also strongly agreed that use of

rubrics brought order and organization to their teaching performance. Two other teachers *Agreed* with item number 11.

### Discussion

The study aimed at investigating the effectiveness of use of rubrics on setting up group work activities. A group of teachers used rubrics while setting up group work activities in classroom and another group of teachers received post observation feedback regarding group work activities. Results showed that teachers who had used rubrics while setting up group work activities had significantly better performance. The results of attitude questionnaire also proved that teachers had positive attitude towards the use of rubrics in setting up group work activities in language classrooms. The results of the study is consistent with other previous studies in terms of the positive effect of rubrics (e.g. Goodrich, 1997; Hafner & Hafner, 2003).

The better performance of language teachers could be attributed to many benefits of rubric and self-assessment as was discussed in chapter two. The researcher believes that the clear scheme and goal oriented quality of rubrics help teachers to better focus on their performance. Besides, rubric is considered a non-intrusive action of assessment which by itself eliminates the anxiety due to presence of an observer in the classroom. In this way teachers have more freedom and focus regarding their performance. In addition rubric as a self-evaluation device provides the opportunity to reflect on their teaching. Airasian and Gullickson (1997) explain that reflective thinking is an important part of self-evaluation and that self-evaluation is the most prevalent form of teacher evaluation. Another reason that can be put forward is the continuous nature of rubrics. As teachers always have access to the rubrics, this provides the ground for ongoing reflection and assessment. Reflection should be an ongoing activity. This fact is not common knowledge because often self-evaluation is performed informally (Airasian and Gullickson, 1997). Teachers' response to the attitude questionnaire also shows that rubrics provide the ground for continuous assessment. For instance, the item number 2 clearly states; *I believe that an ongoing, long-term, continuous, and constant assessment which promotes learning is obtained through application of Rubrics.* And 3 out of 5 teachers agreed with this statement.

Rubrics also provide clearly stated goals which do not allow teachers to get overwhelmed and confused. This gives more confidence to the teachers and allows them stay more focused on tasks. Teachers' responses to attitude questionnaire also proved this. Three teachers responded *Strongly Agree* and *Agree* to the statement; *I felt more confident when setting up group work activity by using rubrics.* Hafner and Hafner (2003) conducted a study on 107 students in a biology college which lasted for 8 years. The purpose was to look how students could use rubrics. In other words, they wanted to see if rubric could be used as an assessment tool by students to grade their peers' performance. It was found out that rubric is a useful tool for both peer assessment and self-assessment. The score peer determined for each other was not significantly different from instructor's scores. Andrade, Du, and Wang (2008) investigated the use of rubric for assessing writing assignment of third and fourth-grade level. They wanted to see the effect of using a model to create rubric and then use the rubric for self-assessment during assignment. By the help of two way ANOVA they examined the effect of treatment. Results indicated a significant difference between treatment group and comparison group.

The findings of the present study suggest that teacher's self-evaluation program need to be incorporated in language teaching program. In this way more quality teaching

performance can be expected. Based on the findings of this study rubrics can be used as viable self-evaluation tool for language teachers. Therefore it is suggested that teachers get more familiar with the use of rubrics and design of rubrics. Various rubrics can be designed by experienced teachers and applied in language classrooms. People involved in language material development can also include rubrics in the course books so that both teachers and students get equipped with clearly defined procedure and goals to accomplish language tasks and also evaluate their performance. Teachers trainers also need to promote use of rubrics among language teachers as a self-assessment tool. In case they are fully familiar with rubrics and rubric design, they must keep themselves up-to-date in this matter. Rubrics can also be used for students as well. They can be equipped with rubrics that clearly define procedures and goals for various language tasks and activities including group work activities.

## References

- Airasian, P., & Gullickson, A. (1997). *Teacher self-evaluation tool kit*. Thousand Oaks, California: Corwin Press.
- Alonso-Tapia, J., & Panadero, E. (2010). Effect of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje*, 33(3), 385–397.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12–19.
- Andrade, H., Du, Y., & Wang, X. (2007). *Putting rubrics to the test: A study of the effects of rubric-referenced self-assessment on students' writing*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices*, 27(2), 3–13.
- Bejarano, Y. (1987). *A Cooperative Small-Group Methodology in the Language Classroom*, *TESOL Quarterly*, 21:483-504.
- Bridges, E. (1990). Evaluation for tenure and dismissal. In J. Millman & L. Darling Hammond (Ed.), *the new handbook of teacher evaluation* (147-157). Newbury Park, CA: Sage.
- Brown, H.D. (2001). *Teaching by Principle: An Interactive Approach to Language Pedagogy*, 2nd ed. (Essex: Longman)
- Cardno, C. (1999). Appraisal policy and implementation issues for New Zealand schools. *The International Journal of Educational Management*, 13, 87–97.
- Danielson, C (2009). *Talk about teaching! Leading professional conversations*. Thousand Oaks, California: Corwin Press.
- Danielson, C. & McGreal, T. (2000). Teacher evaluation to enhance professional practice. (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 337-353).

- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.
- Davis, R. (1997). 'Group Work is NOT Busy Work Maximizing Success of Group Work in the L2 Classroom', *Foreign Language Annals*, 30:265-279.
- Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fushino, K. (2010). Causal Relationships between Communication Confidences, Beliefs About Group Work, and Willingness to Communicate in Foreign Language, *TESOL Quarterly*, 44:700-724.
- Goodrich, H. (1997). Understanding rubrics. *Educational Leadership*, 54(4), 14-17.
- Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509–1528.
- Halonen, J. S., Bosack, T., Clay, S., & McCarthy, M. (with Dunn, D. S., Hill, G. W., IV, McEntarffer, R., Mehrotra, C., Nesmith, R., Weaver, K. A., & Whitlock, K.). (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30, 196–208.
- Harris, B. (1986). *Developmental Teacher Evaluation*. Newton, MA: Allyn and Bacon.
- Hess, N. (2001). *Teaching Large Multilevel Classes, Cambridge Handbooks for Language Teachers*. Cambridge: Cambridge University Press.
- Long, M.H and Porter, P.A. (1985). Group Work, Interlanguage Talk, and Second Language Acquisition. *TESOL Quarterly*, 19:207-228.
- Marincovich, M. (1999). Using student feedback to improve teaching. In P. Seldin (Ed.), *Evaluating faculty performance: A practical guide to assessing teaching research and service* (pp. 45–67). Bolton, MA: Anker.
- Marincovich, M. (1998). *Teaching teaching: The importance of courses on teaching in training programs*. In M. Marincovich, J. Prostko, & Frederick Stout, (Eds.), *The Professional Development of Graduate Teaching Assistants*, 145-162. Boston: Anker Pub Co.

- Moss, P. (2010). Thinking systemically about assessment practice. In M.M.Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 355-374). San Francisco: Jossey-Bass.
- Panadero, E., &Jonsson, A. (2013). The use of scoring rubrics for formativeassessment purposes revisited: A review. *Educational Research Review*, 9(0), 129–144.
- Papay, J.P. (2012). *Refocusing the debate: Assessing the purposes and tools of Teacher assessment and the quest for teacher quality: A handbook* (pp. 69-132). San Francisco, CA: John Wiley & Sons.
- Reddy, Y, M, & Andrade, H (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*. 35(4), 435–448.
- Smith, C. (2008). Building effectiveness in teaching through targeted evaluationand response: Connecting evaluation to teaching improvement in higher education. *Assessment and Evaluation in Higher Education*, 33, 517–533.
- Stevens, D. D., & Levi, A. J. (2005). *Introduction to rubrics: An assessment toolto save grading time, convey effectivefeedback, and promote student learning* Sterling, VA: Stylus.
- Stronge, J. H., Richard, H. B., &Catano, N. (2008). *Qualities of effective principals*. Francisco, CA: John Wiley & Sons.
- Tucker, P.D. and Stronge, J.H. (2005). *Linking Teacher Evaluation and StudentLearning*. Alexandria, VA: Association for Supervision and Curriculum Development.