The Journal of Modern Thoughts in Education
Vol 4, No 4, Autumn 2009, pp. 73-90

The Effectiveness of Reflective Teaching Tools in English Language Teaching

Hamidreza Fatemipour[1]

Abstract

The purpose of the present study was to determine whether different reflective teaching tools obtained the same kind of data. If not, which tool provided the reflective teachers with more reliable data? In order to investigate the research question, a researcher-made questionnaire related to the teachers' decision making was designed. The population of the study consisted of all the Teachers (i.e., around 60 teachers) and students (i.e., around 2500 students) in the English Department of Islamic Azad University, Roudehen Branch in the academic year of 2006-2007. The method of selecting the research sample was simple random sampling. The research sample was comprised of 10 teachers and 251 students. The design of the study was descriptive (non-experimental correlational). The statistical analysis of data included the application of ANOVA and Pearson Coefficient of Correlation. The ANOVA test showed that the F observed (8.12) was significant at the 0.05 level; therefore, there were significant differences among the kind of data that different reflective tools obtained. Moreover, further analysis emphasized that some reflective teaching tools provided the teachers with more reliable data.

Keywords: reflection, reflective teaching, reflective tools, teacher diary, peer observation, students' feedback

1- Assistant Professor, Islamic Azad University, Roudehen Branch (Email: hamidfatemipour@yahoo.com)

Introduction

Reflective teaching is an outstanding model in teacher education which has been taken into consideration by educators in foreign and second language teacher education. As it is emphasized in reflective teaching model, to optimize the teaching and learning of a language in EFL contexts, it may be useful to embrace the concept of the teacher as a learner and a professional. “Every teacher has a professional responsibility to be reflective and evaluative about their practice. As a result of this reflection teachers will be able to identify how to improve their professional activity in order to improve the quality of pupils' learning.” Reflection enables teachers to observe what is going on around the class. It persuades teachers to apply their thoughts and “promote changes in pupils' learning behavior.” Reflection is also an important factor in cooperation or collaboration among teachers. “Reflective partnerships between teachers are particularly effective. Peer mentoring partnerships will support individual teachers in reflecting on and describing their practice. As a result of these focused discussions a teacher is able to better understand practice and be able to take steps to improve practice” (Rose, 2007, p. 1).

Graves (2002) is another educator who is in favor of reflective teaching. She believes that “reflection is one of the most powerful tools teachers can use to explore, understand, and redirect their practice. Reflection is about learning to see and to understand what is seen. It is not simply being able to identify problems and frame solutions, although both are crucial.” According to her, there are two potential pitfalls which teachers should be aware of when they observe their own teaching reflectively. The first potential danger is to follow reflective process but not to take any action based on the obtained data “—to hold up a mirror,

acknowledge what is there and how one feels about it, but go no further." The second possible danger is to merely consider reflection as a process through which an observed problem is solved. While it can be a part of reflective process, the main goal is to find the underlying reasons which have caused the observed problem. "When teachers are able to explore the root issues and beliefs, a shift occurs in their understanding and a wider range of effective, intelligent actions becomes possible" (Stanley, cited in Graves, 2002, p. 20). According to Richards & Lockhart (1996), when teachers are involved in the process of teaching some events occur that they can use to have a better understanding of their teaching. Sometimes they take these events for granted and they fail to reflect on them; in fact, the events that occur around the classroom can provide the teachers with "the basis for critical reflection". The authors suggest some procedures that can be used by teachers to investigate classroom teaching. The proposed procedures are as follows: teaching journals, lesson reports, surveys and questionnaires, audio and video recordings, observation, and action research. Some of these procedures are also mentioned by Murphy (2001) and Tice (2002) and they properly label them as tools.

The reflective teaching tools which are usually applied for obtaining data in teaching English as a Foreign Language (EFL) are teacher diary, peer observation, audio recording and students' feedback.

In order to prepare a diary, the teacher writes about what happens in the class after each lesson. He can note his "reactions and feelings and those . . . [he] observes on the part of the students". It can be done by answering some general questions that form a teaching diary.

To benefit from the peer observation tool, the teacher asks a colleague to

attend his class and collect information about the lesson. It can be done through note taking or "a simple observation task."

Audio recording of lessons is considered a suitable tool in obtaining data used for reflective teaching. "You may do things in class you are not aware of or there may be things happening in the class as the teacher you do not normally see." Recording of lessons can be useful in showing the teachers different aspects of their behavior.

Students' feedback is a tool used for finding out the learners' opinions and perceptions about the teaching process, and teachers' efforts that "can add a different and valuable perspective." The data can be obtained through questionnaires (Tice, 2002, pp. 2-3).

The problem under investigation in this study is that while some experts in reflective teaching suggest different tools for obtaining the data that teachers need in order to evaluate their own way of teaching (Richards and Lockhart, 1996; Murphy, 2001; Tice, 2002), they do not fully discuss the effectiveness of the tools regarding the data that teachers need to obtain. For instance it is mentioned that "each procedure has advantages and limitations, and some are more useful for exploring certain aspects of teaching than others. The reader [teacher] will have to decide which procedures are useful and for what purposes" (Richards & Lockhart, 1996, p. 6). This is the problem which needs to be resolved during the process of this research.

Regarding the aforementioned research problem, the purpose of the present study is to evaluate the efficiency of reflective teaching tools in English language teaching. In other words, the researcher wants to determine whether different

reflective teaching tools obtained the same kind of data.

Regarding the aforementioned problem, the following research question would arise:

1. Do all reflective tools provide the reflective teachers with the same kind of data?

The following null hypothesis is proposed on the basis of the above research question.

1. All reflective tools provide the reflective teachers with the same kind of data. No matter which tool is used, the data remain the same.

Methodology

Participants

The population of the study consisted of all the Teachers (i.e., around 60 teachers) and students (i.e., around 2500 students) in the English Department of Islamic Azad University, Roudehen Branch in the academic year of 2006-2007. The method of selecting the research sample was simple random sampling. Based on this method of choosing the research sample, each teacher had an equal and independent chance of being selected as a research subject. The research sample consisted of ten lecturers chosen randomly from the department of English, Islamic Azad University, Roudehen Branch. The teachers had at least one class in the academic year of 2006-2007. Based on the research design, one of the classes of each teacher was selected and the total population of their classes comprised 251 students who had chosen basic or specialized English courses to pass in the same year. The teachers were free to choose the class that they wanted to expose to

observation. The reason behind this was to let the teacher choose a class where s/he felt more comfortable and probably there was less resistance and more cooperation from the students' side.

Instruments

As the purpose of the study was to compare the data derived from the application of different tools (i.e. Teacher Diary, Peer Observation, Students' Feedback and Audio Recording), the researcher had to choose a uniform procedure which let him quantify the obtained data for the specified intention of the comparison and contrast of data. Inspired by Tice's (2002) suggestions for writing a teaching diary which guides the reflective teachers to start by answering some open questions in the form of a questionnaire, the researcher designed a questionnaire.

Therefore, the designed questionnaire was the instrument given to all research subjects (i.e. students, class teachers, colleagues and observers) to fill it out by converting their data from observation reports, recording transcripts, diary writings to one of the options linked to the questionnaire items. Although questionnaires can contain open and/or closed questions as well as statements, the researcher preferred to write the items just in the form of statements. The prepared items were Likert type statements which ask the respondents to indicate the degree of agreement to the posed idea by choosing one of the given five options. Likert type statements have all the advantages of closed questions. In other words, while statements are capable of expressing desired points clearly and extract the respondent's idea, their related options can be readily converted to quantifiable data for further statistical analysis.

As the focus of the study was on the area of "teacher decision making," the

researcher first reviewed the 45 questions which were suggested by Richards and Lockhart (1996) to explore the three components of decision making, i.e. planning decisions, interactive decisions and evaluative decisions. Then, the questions which were more appropriate to be investigated through using a questionnaire and/or were emphasized by other experts (i.e. Hiller, 2005; Murdoch, 1998; Murphy, 2001; Tice, 2002), were chosen and transformed into statements. The change was done based on the Likert's suggestion related to designing the questionnaire items and responses.

In the present study, piloting consisted of randomly choosing two classes which comprised 98 students among the English classes held in Islamic Azad University, Roudehen Branch in the academic year of 2006-2007. In the next step, the questionnaire which had been modified according to experts' recommendations was administered to the subjects. Finally, the researcher utilized the extracted data to calculate the predictive validity and reliability of the designed questionnaire.

Regarding the content validity of the research instruments, a group of experts in the field of applied linguistics and education reviewed the designed questionnaire and provided the researcher with their feedback. Thereafter, the questionnaire was examined and the ambiguous items were modified or totally removed from the body of the questionnaire.

To test the predictive validity of the designed questionnaire, following the Best and Kahn's (2003) suggested procedure, the researcher chose one of the classes which had been selected for pilot study. The class consisted of 52 students. In order to make some external criterion for comparison, the researcher asked the class teacher as another source to fill out the questionnaire. In addition, the researcher as a non-participant observer attended the students' class and gathered data to be

compared with the data obtained from the students. Finally, the data gathered from the administration of the questionnaire was compared with the overall data obtained from the observer's and teacher's questionnaires to determine whether the designed instrument had the acceptable predictive validity. The coefficient of correlation between the two sets of scores was 0.41. The obtained result was significant at the 0.01 level (2-tailed). Therefore, it was concluded that the instrument had the required predictive validity.

To statistically investigate the construct validity of the questionnaires, the researcher computed the internal consistency of the items using Cronbach's Alpha. Garson (2006) states that one of the common methods of estimating the construct validity of a questionnaire is to compute the internal consistency of the items by applying Cronbach's Alpha. According to him, for confirmatory purposes, the consistency of .70 and .80 is considered as "acceptable" and "good" respectively. Since the obtained Alpha of the instrument equaled .9065, it was proved that the instrument had a "good" index of construct validity.

In this research, the internal consistency type of reliability was used to measure the coefficient of the questionnaire reliability. The results of conducting Cronbach's Alpha model showed that the items were highly correlated with the obtained Alpha of 0.91 and none of the items proved to be irrelevant.

Design and Procedure

Regarding research design, this research falls into the category of descriptive (non-experimental and correlational) studies. Descriptive research is a kind of research which refers to "investigation which utilizes already existing data or non-experimental research with a preconceived hypothesis" (Seliger & Shohamy, 1989,

p. 117). Following the descriptive research procedure, the researcher applied the statistical analysis to compare and contrast the data obtained from the administration of each reflective tool with the data provided by the three other individual tools and the obtained mean of all tools. The statistical analysis of data included the application of ANOVA and Pearson Coefficient of Correlation.

In the data gathering procedure, the first step consisted of a very short orientation session for the teacher. The purpose was to make the subject familiar with the purpose and process of doing the research and agree on a teaching session for administering the research tools. Then, on the due date, in a single session, all four tools were administered for each class. It means that while the teacher started teaching, the researcher as a non-participant observer was present in the class, taking notes which he used later to complete the questionnaire. The teaching process was tape recorded in the same session. Later, the recorded tape was reviewed by a colleague and the given questionnaire was filled out accordingly. When the teacher was finished with the lesson, students were asked to give their feedback on the teaching process by completing their questionnaires. The teacher was the last subject who was asked to reflect on his/her teaching by completing the first questionnaire.

Results

Table 1 presents the extracted data resulting from the administration of the research instrument.

Table 1
Descriptive Statistics: The Sample of Students

Number	Valid	251
	Missing	0
Mean		156.19
Std. Error of Mean		1.52
Median		155
Mode		160
Std. Deviation		24.14
Variance		582.91
Skewness		.05
Std. Error of Skewness		.15
Kurtosis		.02
Std. Error of Kurtosis		.31
Range		126
Minimum		89
Maximum		215
Sum		39206

As it is shown in Table 1, the three measures of central tendency (mean, median and mode) nearly overlap one another in a distribution which has a wide range of 126. Furthermore, considering the skewness and kurtosis of the data distribution, Price (2000) explains that in order for a degree of skewness not to be considered as significantly skewed it should be within the range of (-2 Std Error of Skewness) and (+2 Std. Error of Skewness). As it is evident from the table, the index of skewness (.05) is within the acceptable range (-.30 and +.30); so, the distribution is normal in this regard.

The distribution also has an acceptable index of Kurtosis. The range of normality for the index of Kurtosis, as Price (2000) notes, is from -2 Std. Error of Kurtosis to +2 Std. Error of Kurtosis. Since, the obtained index of kurtosis (.02) is within the range of normality (-.62 and +.62), the data are comparable with the

normal curve.

In order to investigate the research question, the first step was to prepare the needed data and choose the appropriate statistical procedure. Since, the purpose was to examine whether there was any difference among the obtained data, and there were more than two groups of subjects, the researcher decided to apply the analysis of variance (ANOVA). To do this, the needed data were the scores obtained from the administration of the four reflective tools to four groups i.e. students, class teachers, observers and colleagues.

Table 2
Scores Obtained from Administration of the Four Reflective Tools

Class No.	1	2	3	4	5	6	7	8	9	10
Students [Class Mean] (Group1)	159.1	160.4	158.5	153.8	157.5	160.3	144.5	162.2	139.6	163.6
Teachers (Group2)	170	150	154	204	164	182	143	180	168	200
Observers (Group3)	166	147	175	178	174	163	146	157	155	155
Colleagues (Group4)	140	145	148	152	149	139	151	143	130	143

Table 2 presents the scores resulting from the administration of the four reflective tools. Because there was one teacher, one observer and one colleague for each class, the mean of scores obtained from the students' questionnaires in each class was calculated and presented in the following table as the class mean.

Table 3
One Way ANOVA: Summary of Four Group Analysis of Variance

	Sum of Squares	df	Mean Square	F Observed	F Critical	Sig.
Between Groups	395.50	3	1317.17	8.12*	2.88	.000
Within Groups	5843.44	36	162.32			
Total	9794.95	39				

***** *The mean difference is significant at the .05 level.*

Table 3 shows that the F ratio resulting from running ANOVA is statistically significant. In other words, the F-observed (8.12) is greater than the F-critical (2.88); so, the means of four samples are too different to attribute to chance or sampling error. This results in rejecting the null hypothesis which states that there is no difference among the kind of data that all the tools obtain. Furthermore, Post Hoc Tests were implemented to compare the means and determine the exact role of each tool in producing variant data.

Table 4
Post Hoc Tests (Tukey HSD): Multiple Comparisons

Groups of Subjects (I)	Groups of Subjects (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-15.60*	5.69	.046	-30.89	-.21
	3	-5.60	5.69	.755	-20.99	9.69
	4	11.94	5.69	.174	-3.39	27.29
2	1	15.60*	5.69	.046	.21	30.90
	3	9.90	5.69	.320	-5.45	25.25
	4	27.50*	5.69	.000	12.15	42.85
3	1	5.60	5.69	.755	-9.69	20.99
	2	-9.90	5.69	.320	-25.25	5.45
	4	17.60*	5.69	.019	2.25	32.95
4	1	-11.94	5.69	.174	-27.29	3.40
	2	-27.50*	5.69	.000	-42.85	-12.15
	3	-17.60*	5.69	.019	-32.95	-2.25

***** *The mean difference is significant at the .05 level.*

Table 4 summarizes the results of implementing the Post Hoc Test. As it is shown, all the mean differences in the second column are within the range of 95% confidence interval presented in the last column. It means that the differences between all pairs of means are considered normal. From the presented comparisons, the following can be concluded:

1. The tools of Teacher Diary (2) and Audio Recording (4) obtain the most various data with 27.50 mean difference which is much greater than the std. error (i.e. 5.69), and rejects the null hypothesis at the .05 level.
2. The tools of Peer Observation (3) and Audio Recording (4) stand in the second place in obtaining the various data with 17.60 mean difference which is much greater than the std. error, and rejects the null hypothesis at the .05 level.
3. The tools of Students' Feedback (1) and Teacher Diary (2) stand in the fourth place in obtaining the various data with 15.60 mean difference which is much greater than the std. error, and rejects the null hypothesis at the .05 level.

The statistical analysis showed that different reflective tools provided the data which were significantly different from each other. Therefore, the null hypothesis which stated, "There is no difference among the kind of data that all the tools obtain" was rejected. The rejection of the null hypothesis justifies the existence of the problem that there is no guideline available to novice teachers providing them with the choice of an appropriate tool based on their needs.

Furthermore, since the findings of the present study emphasize on diversity of the data given by different reflective tools, reflective teachers who generally

reflect on their teaching may be recommended to make use of just not one reflective tool (e.g., Teacher Diary) and try different plausible reflective tools.

The finding was also confirmed by running Post Hoc Tests. The results of examining the obtained data through Post Hoc Tests showed the difference between the means of three group pairs at the .05 level appeared to be significant: Students' Feedback and Teacher Diary; Audio Recording and Teacher Diary; as well as Peer Observation and Audio Recording. The implication is that all four groups have different ideas about the desirability of the teaching/learning process.

Discussion and Conclusions

In order to investigate which tool is more effective in obtaining the data for teachers to reflect on, the researcher calculated the mean of the data obtained from all reflective tools. It was done based on the assumption that it is ideal for a reflective teacher to apply all four tools in his/her class to have an accurate and comprehensive view about what is going on in the class. Since it is time consuming and sometimes not possible, this study was done to show the reflective teachers which tool could obtain the data which were as close as the mean of the four tools, that is, which tool is more effective in obtaining the data close to the mean of all data. Comparing the results obtained from administering each reflective tool and the mean of all data, the researcher concluded the following:

1. Teacher Diary was the most efficient reflective tool. The coefficient of correlation between the obtained data from this tool and the mean of all data appeared to be .84.
2. Peer Observation was more efficient than the other two remaining tools

(Students' Feedback and Audio Recording) in obtaining the data close to the mean of all data (r = 0.71).

3. Students' Feedback was positioned in the third place. The coefficient of correlation between the data obtained by this tool and the mean of all data appeared to be 0.58.
4. Audio Recording was the least efficient tool in obtaining the close to the mean of all data. The coefficient of correlation between the data obtained by this tool and the mean of all data was just 0.31.

The data obtained from this study is in accordance with the Farrell's (2001) findings related to peer observation. While this tool is effective in providing reflective teachers with valuable data, some teachers do not like to be observed by their colleagues; therefore, they do not teach normally in the presence of an outsider. The findings of the study are supported by Bailey (1991) and Tice (2007) who found out that diary writing makes different aspects of teaching known to reflective teachers. Regarding the application of the audio recording tool, the findings were supported by the Tice's (2007) experience of using the tool for reflection. She believes that by recording the teaching session, the teacher can become aware of the things happening in the class. The experience of using audio recording which appeared to be intrusive and affected the behavior of both teachers and students was somehow different from Kember's (2000) who believes that audio recording is the least intrusive method for gathering data for reflection.

Finally, regarding the obtained data, the reflective teachers are recommended to apply more than one tool in order to obtain more reliable data. In case, applying different reflective tools is not possible, Teacher Diary and Peer Observation are the

tools which are strongly recommended to be used for extracting the required data for reflection. An important point is that Students' feedback can provide teachers with the data which are unique and cannot be obtained by the other three reflective tools. Therefore, it is recommended that if teachers need to obtain the data from students, the results had better be checked by the data obtained through administering another tool i.e., peer observation.

References

Bailey, K. M. (1991). Diary studies in teacher education programs. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 215-26). Cambridge: Cambridge University Press.

Best, J. W., & Kahn. J. V. (2003). *Research in Education* (9th ed.) USA: Prentice-Hall, Inc.

Farrell, T. (2001). Tailoring Reflection to Individual Needs: A TESOL Case Study. *Journal of Education for Teaching, 27*(1), 23-38.

Garson, G. D. (2006). *Validity. North Carolina State University (NCSU).* Retrieved on May 20, 2007 from <http:// www2.chass.ncsu.edu/garson/ pa765/ validity.htm>.

Graves, K. (2002). Developing a reflective practice through disciplined collaboration. *The Language Teacher, 26*(7), 19-21.

Hillier, Y. (2005). *Reflective teaching in future and adult education*. (2nd ed.) UK: Continuum.

Kember, D. (2000). *Action learning and action research: Improving the quality of teaching and learning*. London: Routledge.

Murdoch, G. (1998). A progressive teacher evaluation system. *Forum, 36*(3), Retrieved on May 22, 2006 from http:// exchanges.state.gov/ Forum/ Vols/ Vol36/ No3/ p2.html

Murphy, J. M. (2001). Reflective teaching in ELT. In M. Celce-Murcia. (Ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 499-514). Boston: Heinle and Heinle.

Price, I. (2000). Determining if skewness and kurtosis are significantly non-normal. *School of Psychology, University of New England (UNE).* Retrieved on May 19, 2007 from http:// www.une.edu.au/ WebStat/ unit_materials/ c4_descriptive_statistics/ determine_skew_kurt.html

Richards, J. C., & Lockhart, C. (1996). *Reflective teaching in second language classrooms*. Cambridge: Cambridge University Press.

Rose, M. (2007). The reflective practitioner. *European Center for Modern Languages*. Retrieved on May 15, 2007 from <http:// www.ecml.at/ html/ quality/ english/ continuum/ self_assessement/ teachers/ MR_reflective%20practitioner. html

Seliger, H. W., & Shohamy, E. (1989). *Second language research methods.* Oxford: Oxford University Press.

Tice, J. (2002). Reflective teaching: Exploring our own classroom practice. *BBC/ British Council Teaching English.* Retrieved on May 15, 2007 from http://teachingenglish.org.uk/ think/ methodology/ reflection/ html